Systemic Risk in UK Financial Services

AI: An Amplifier of the Average

Amplification of the Average

Automation Complacency

Personal Accountability

Your new AI based system just approved 10,000 transactions. How confident are you that none will result in a regulatory investigation?

THE HIDDEN RISKS OF AI

AI is transforming business, delivering significant competitive advantage. Recent research shows 92% of business leaders expect measurable ROI from these AI investments within 2 years. However, this transformation to autonomous AI demands rigorous governance to avoid the system risk of AI solutions.

EXECUTIVE OVERVIEW

AI Amplifies the Average, Hiding Outlier Risk: AI systems optimise for what's typical and overlook what's exceptional, excelling at common cases but systematically failing on rare, critical outliers that define risk, liability and regulatory breach in financial services. The good news: With the right assessment framework these risks are identifiable, quantifiable and manageable.

Mismatched Confidence is a Core Technical Failure: AI systems can produce factually incorrect outputs with the same confidence as correct ones, especially when extrapolating beyond validated data - leading to undetected, plausible errors ("*hallucinations*").

Systemic Risk is a Governance Issue: The main risk is not "*rogue algorithms*" but governance that overestimates AI's capability and calibration. Unquantified AI models are unassessed liabilities.

Personal Accountability Under SM&CR: Senior leaders are personally responsible for AI-related risks under the Senior Managers and Certification Regime (SM&CR). Regulatory failure of AI will be interpreted as a failure of governance and operational resilience.

Real-World Incidents Highlight the Stakes: Documented failures include gender bias in credit models, massive trading losses and regulatory settlements - demonstrating the material impact of errant AI.

Mitigation Demands Integrated Technical and Procedural Controls: Effective risk management controls spanning technology functions, corporate governance and operational procedures.

Priority Recommendations for Executives: A strategy for managing this new risk within your organisation.

Independent AI Risk Assessments are Essential: Regular, independent assessments provide clarity on current capabilities, identify gaps and offer a prioritised roadmap for remediation - supporting regulatory defensibility and operational resilience.

ADVANTAGE AI

The FCA and PRA expect assurance. ADVANTAGE AI delivers it.

Systemic Risk in UK Financial Services

AI: An Amplifier of the Average

THE STRUCTURAL VULNERABILITY

Many AI models are mathematically designed to maximise statistical probability - they amplify the average. This creates an invisible but systematic bias: **high reliability on common cases masks catastrophic failure on rare, critical outliers** that define risk, liability and regulatory breach in financial services.

The core technical failure is **mismatched confidence**: AI systems produce factually incorrect outputs with the same authoritative presentation as correct ones. When extrapolating beyond validated data patterns, they can become "*overconfident interpolators*" - generating plausible errors without warning signals (*hallucinations*). The human operator receives no reliable indication the system is operating outside its competence.

The emerging **systemic risk** is not rogue algorithms - it is governance that assumes AI is more capable, cautious and calibrated than it is.

In financial services, this ***Amplifier of the Average*** problem translates directly to core business areas:

- **Credit Models:** Excel at typical borrower profiles but fail to correctly assess novel business models or atypical income patterns (e.g. gig economy workers).
- **Risk Models:** Trained on historical normality, they fail to recognize novel correlations or regime changes during periods of market stress.
- **Compliance:** Models optimised for general compliance may miss subtle, rare but crucial deviations indicative of fraud or market abuse.

"AI doesn't just fail because it's wrong. It also fails because we think it's right."

FROM SYSTEMIC RISK TO PERSONAL LIABILITY

Under the Senior Managers and Certification Regime (SM&CR), **senior leaders hold personal responsibility** for managing material risks. If your AI's confidence isn't measured, your risk isn't managed. The failure of AI will be interpreted by regulators as a failure of governance assurance and operational resilience on the part of the SMF holder.

AI INNOVATIONS CREATE ORGANISATIONAL VULNERABILITIES

Organisations continue to follow the same "deploy first, govern later" playbook - even when regulations (with significant penalties) already exist and face the loss of customer confidence and trust combined with regulator significant penalties.

Apple Card (2019)	**Allianz Global** (2024)	**Knight Capital** (2012)
A regulatory investigation and severe reputational damage following an AI-driven creditworthiness model was accused of gender bias discriminating against women.	A hedge-fund settled with the U.S. SEC for $170M after the firm replaced human traders with an under-performing AI algorithm.	£440M loss in 45 minutes from algorithmic failure. Flash Crash demonstrated systemic amplification when outlier events trigger correlated AI responses.

AI: An Amplifier of the Average

TECHNICAL AND PROCEDURAL MITIGATION STRATEGIES

Mitigating against these risks is neither a static nor single activity. To be effective it requires a comprehensive, cohesive and constant approach spanning technology functions, corporate governance and operational procedures to minimise these systemic risks.

Technical Mitigation: Beyond Accuracy

1. **Uncertainty Quantification (UQ) and Calibration:** High-stakes systems must quantify the confidence of their predictions, not just state a result. Prioritising calibration ensures the alignment between predicted confidence and actual correctness, thereby reliably flagging low-confidence, high-risk results for mandatory human intervention.
2. **Explainability (Chain-of-Thought):** Implementing Chain-of-Thought explanations significantly improves human diagnostic accuracy. This technical solution allows practitioners to audit the AI's step-by-step logic, transforming the AI from an oracle into a transparent assistant.
3. **RAG Optimisation for Outliers:** RAG systems must be explicitly enhanced with outlier detection rules and distance metrics to ensure that semantically distant, but critically relevant "*deep tail*" data points are retrieved, actively preventing the algorithm from overlooking rare facts.
4. **Data Integrity and Curation:** To prevent model collapse, continuous monitoring and rigorous curation are necessary to maintain a sustained balance between synthetic (AI-generated) data and diverse, real-world human data.

Board Questions

Have we reviewed AI model performance in the last 12 months?

Are we confident our AI systems can flag outlier risks before regulators do?

Are our frontline staff identifying and challenging AI-generated outputs that may be incorrect, biased or poorly calibrated?

If our AI discriminates against an edge case tomorrow, who is personally accountable under SM&CR?

Governance and Procedural Safeguards

5. **Adversarial Testing:** Validation must move beyond superficial benchmarking on average test sets. Systems must be subjected to rigorous stress tests against known adversarial edge cases, rare events and data from vulnerable populations to proactively probe for failure modes on tail distributions.
6. **Built-in challenge mechanisms:** Drawing on principles of democratic governance, safety frameworks must anticipate failure and bias. This involves embedding mechanisms of distrust - such as mandatory contestability and redress mechanisms - to ensure public trust is earned through verifiable fallibility rather than assumed infallibility.
7. **Mandatory Human-in-the-Loop (HILT):** For high-stakes decisions, the human professional must be the ultimate arbiter, requiring independent expert validation for critical outlier cases. This requires continuous AI literacy training to counter automation complacency.
8. **Audit and Monitoring:** Third-party auditing and continuous monitoring of real-world outcomes are essential to check for disparate impact across demographic subgroups, ensuring that the AI is not accurate on average but biased against an outlier population.

AI: An Amplifier of the Average

CONCLUSION

The concept of **AI: An Amplifier of the Average** defines a fundamental tension where the pursuit of algorithmic efficiency structurally increases exposure to low-probability, high-impact failures. The technology's bias toward the mean risks systematically erasing the critical outliers necessary for safety, equity and innovation. The most critical dynamic is the self-reinforcing governance blind spots: technical flaws in the algorithm are compounded by human cognitive limitations and weak institutional accountability, leading to systemic, long-term degradation of professional expertise and the data ecosystem itself.

PRIORITY RECOMMENDATIONS FOR EXECUTIVES AND SENIOR LEADERS

- **Mandate Uncertainty Quantification (UQ) as a Core Design Requirement:** systems must reliably quantify the confidence of their predictions, ensuring that high-risk extrapolations are automatically flagged.

- **Establish Clear, Unambiguous Accountability via SM&CR/Governance:** Executives should leverage the Senior Managers and Certification Regime (SM&CR) to explicitly map AI oversight and validation responsibilities, ensuring individual accountability for managing outlier risks within the organisation.

- **Invest in Human Oversight Capability (Augmentation over Replacement):** Implement mandatory professional training focused on auditing AI outputs, recognising hallucination patterns and interpreting calibration scores. The strategic objective is the augmentation of human judgment, not replacement, thereby mitigating skill atrophy and the vicious circle of erosion.

- **Adopt the EU AI Act's High-Risk Standards Proactively:** Regardless of UK policy divergence, treating all high-stakes AI use as compliant with the data quality, documentation and human oversight requirements of the EU AI Act is the most effective measure to ensure global business resilience and minimise the risk of catastrophic outlier failure and subsequent legal liability.

"With the FCA, PRA, and Bank of England signalling increased scrutiny of AI oversight in 2026, the accountability expectations for boards are rising fast."