Vol. 2, No. 2, pp.: 17-32, 2024

# A Comparative Dataset of Cloud AI Services for Financial Systems: Provider Capabilities, Cost Structures, and Governance Evidence

**Tomasz Kowalski[1], Agnieszka Wojcik[2,*], Marek Nowak[3]**

[1] Department of Informatics, Faculty of Computer Science and Telecommunications, Wrocław University of Science and Technology, Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland

[2] Faculty of Economics and Sociology, University of Łódź, Rewolucji 1905 r. nr 39, 90-214 Łódź, Poland

[3] Faculty of Economics, Maria Curie-Skłodowska University, Plac Marii Curie-Skłodowskiej 5, 20-031 Lublin, Poland

** agnieszka.wojcik@uni.lodz.pl*

| Article Information | |
|---|---|
| Received | 18 January 2024 |
| Accepted | 29 May 2024 |
| DOI | https://doi.org/10.63646/datamind.2024.020203 |

## Abstract

Choosing a public cloud provider for artificial intelligence workloads is now one of the most consequential infrastructure decisions facing financial institutions, yet the decision remains poorly supported by structured comparative evidence. This study constructs a comparative dataset covering the three dominant providers — Amazon Web Services (AWS), Microsoft Azure, and Google Cloud Platform (GCP) — and benchmarks them on seven dimensions: managed AI service breadth, foundation model and machine-learning catalogue, MLOps tooling depth, cost transparency, operational resilience track record, security and compliance posture, and governance support for regulated workloads. The coded matrix is analysed using descriptive comparison, principal component analysis, k-means clustering, and scenario-weighted suitability scoring. Three financial-services use cases are examined: real-time fraud detection, batch credit risk modelling, and graph-based anti-money laundering. Results indicate that no single provider dominates across all scenarios. AWS leads on ecosystem breadth and on real-time, latency-sensitive inference; Azure leads on governance maturity and regulated-workload alignment with the European Union Artificial Intelligence Act and the Digital Operational Resilience Act; and GCP leads on model-transparency tooling and on per-operation cost for well-defined batch workloads. The paper argues that provider selection should be treated as a scenario-dependent governance decision rather than a global ranking, and offers a transparent matching framework that financial institutions and supervisors can adopt or adapt. The dataset, rubric, and scoring weights are documented at a level of granularity that supports replication and extension to additional providers, workloads, or regulatory regimes.

**Keywords:** ***cloud computing; artificial intelligence; financial services; provider benchmarking; operational resilience; AI governance; EU AI Act; DORA; MLOps; fraud detection***

## 1. Introduction

Financial institutions have moved decisively toward cloud-based artificial intelligence (AI) services. Real-time fraud detection, algorithmic trading support, automated regulatory reporting, generative customer-engagement systems, and graph analytics for anti-money laundering (AML) increasingly run on managed AI platforms provided by Amazon Web Services (AWS), Microsoft Azure, and Google Cloud Platform (GCP). The shift reflects a structural rebalancing of where computation happens, where data resides, and where supervisory expectations are operationalised (Lu and Zheng, 2020; Tabassum et al., 2024; Aljabre, 2024). Parallel work on quantum-financing infrastructure (Lu and Yang, 2024) and decentralised finance (Xu et al., 2024) reinforces a single conclusion: financial systems are being rebuilt around heterogeneous, third-party-supplied compute platforms whose internal choices shape the institution's external exposure. The strategic question is no longer whether to adopt cloud AI, but which provider best supports a given regulated workload under a given accountability regime.

The decision is harder than it appears. The three dominant providers each offer broad and rapidly evolving portfolios of managed AI services that look comparable in marketing terms but differ in architectural philosophy, deployment model, security posture, and pricing logic. Differences in data residency, model transparency, audit-trail depth, and outage history translate directly into differences in supervisory exposure and operational risk for the institution that runs the workload (Bauer et al., 2018; Ozcan and Erol, 2024). Yet most comparative work in the practitioner literature treats the providers as if they were near-substitutes, ranked by aggregate market share or by feature counts rather than by fitness for specific financial-services scenarios. The academic literature, when it engages the question at all, tends to study individual workloads in isolation or to focus on a single regulatory regime, rather than to assemble structured, replicable comparative evidence (Khan et al., 2022; Anwar et al., 2023).

This article argues for a different framing. We treat the public cloud AI landscape as a small but structurally heterogeneous dataset: three providers, evaluated on a documented set of dimensions that matter for regulated AI workloads, and assessed against a small set of well-defined financial-services scenarios. The contribution is modest in form but useful in practice. It offers a transparent comparative framework for choosing among providers before system design begins, and it documents the dataset and rubric at a level of granularity that supports replication, extension to additional providers, and adaptation to evolving regulatory regimes such as the European Union Artificial Intelligence Act (EU AI Act) and the Digital Operational Resilience Act (DORA) (European Commission, 2024; Floridi, 2023).

Three research questions guide the analysis. First, on which structural dimensions do AWS, Azure, and GCP differ when assessed through a common, transparent rubric? Second, do these differences cluster into coherent design philosophies that correspond to different financial-services research and operating logics? Third, which provider best fits each of three representative financial-services scenarios — real-time fraud detection, batch credit risk modelling, and graph-based AML — once the scenario weights are made explicit? To answer these questions we code the three providers on seven dimensions, analyse the coded matrix using descriptive comparison, principal component analysis, and scenario-weighted suitability scoring, and document each step at a level of granularity sufficient for replication.

The empirical material draws on publicly available provider documentation, official pricing calculators, published Service Level Agreement (SLA) terms, post-incident reports following major outages, and third-party compliance attestations. We deliberately exclude private performance benchmarks and proprietary case studies because such material cannot be reproduced without provider cooperation, and the comparative apparatus we

develop is intended to be auditable and reusable. The seven coding dimensions and three scenarios are defined in Section 3, the empirical results are reported in Section 4, and the implications for provider selection, governance, and supervisory practice are discussed in Section 5. Section 6 concludes by situating the contribution within the broader literature on cloud governance, AI compliance, and financial-sector technology adoption.

## 2. Cloud AI in Financial Services and the Comparative Gap

The literature on cloud computing for financial services has matured along three loosely connected tracks. The first track examines the macro-level adoption story: the strategic rationale for migrating regulated workloads to public cloud platforms, the patterns of incremental versus full migration, and the implications for cost structures and operational resilience (Lu and Zheng, 2020; Tabassum et al., 2024; Kou and Lu, 2025). Complementary work on the broader infrastructure stack — including blockchain integration into IoT and financial platforms (Xu et al., 2021) and the trajectory of AI as a general-purpose technology (Zhang and Lu, 2021) — establishes that cloud adoption is no longer optional for institutions of meaningful scale, but it stops short of providing comparative guidance on provider selection.

The second track studies specific AI techniques in financial contexts. Machine-learning models for credit scoring, deep-learning approaches to fraud detection, graph neural networks for AML, and recently large language models for compliance text processing have each been the subject of substantial methodological work (Bao et al., 2022; Cherif et al., 2024; Hilal et al., 2022; Pourhabibi et al., 2020). Earlier comparative analyses of supervised classifiers for card-fraud detection (Awoyemi et al., 2017) and cost-sensitive evaluation frameworks (Bahnsen et al., 2016) anticipated many of the techniques now packaged into managed services, and the broader literature on deep-learning evaluation in regulated domains (Bhattacharya et al., 2021) clarifies the criteria against which production deployments are judged. This track is rich on method and increasingly rich on evaluation against public benchmarks, but it generally takes the underlying infrastructure as given and treats the choice of provider as a procurement question rather than a methodological one. Recent surveys of AI in finance similarly acknowledge that infrastructure choices shape model deployment but do not develop comparative apparatus (Goodell et al., 2021; Lu, 2019; Mhlanga, 2021).

The third track engages governance and supervisory expectations. Work on AI risk management, explainability requirements, and operational-resilience regulation has accelerated since the EU AI Act and DORA entered force, and a parallel literature on operational risk in cloud-dependent financial systems has emerged from incident analysis (European Commission, 2024; Mokander et al., 2023; Schaffer et al., 2023; Cobb et al., 2022). Adjacent contributions on sociotechnical accountability in software systems (Sandberg et al., 2024) extend the discussion beyond the immediate regulatory text and ask how accountability is allocated across the joint human-and-technical system that produces and operates AI services. This track is increasingly explicit about the regulatory geometry that shapes cloud AI use in finance, but its empirical core is typically policy text and incident narrative rather than structured comparison across providers.

Across all three tracks the comparative gap is clear. The provider question — which of AWS, Azure, or GCP fits which regulated workload, and on what evidence — sits at the intersection of the three tracks but is addressed systematically by none of them. The contribution of the present article is to close that gap by treating the three providers as a small comparative dataset, coded on dimensions that matter for regulated AI workloads, and to make the rubric, weights, and scenario definitions transparent enough to support replication and adaptation. The orientation is methodological rather than promotional: we are not arguing for a particular provider but for a particular way of making the choice.

## 3. Coding Framework and Analytical Strategy

The empirical core of this article is a coded comparison matrix. Each of the three providers is scored on seven dimensions using a five-point ordinal rubric anchored in published documentation. The seven dimensions are designed to span the structural axes along which financial-services workloads are evaluated in practice: technical breadth, methodological depth, cost predictability, operational reliability, security posture, governance support, and regional-compliance footprint. Figure 1 summarises the analytical workflow that produces the coded matrix and feeds it into the downstream analyses.

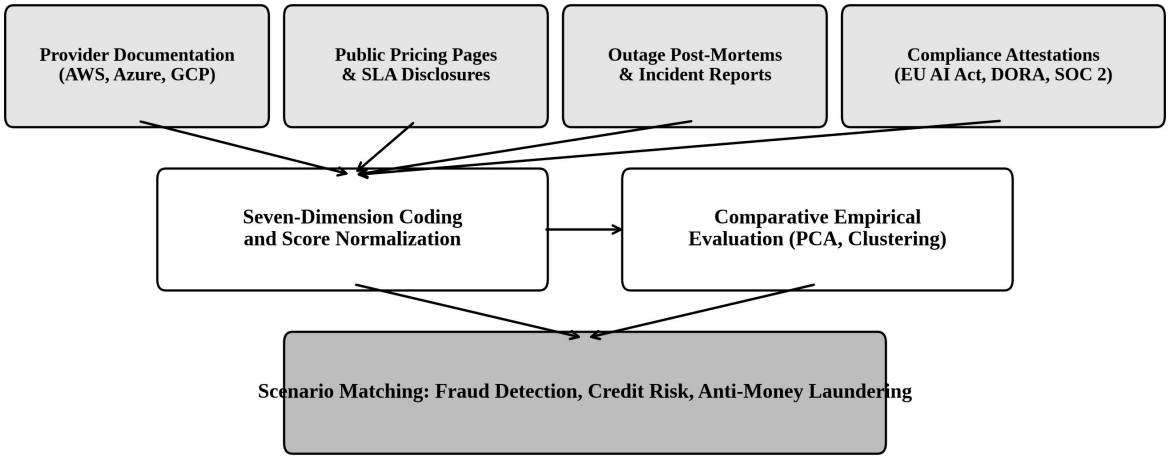


**Figure 1.** *Analytical workflow used to construct the comparative dataset. Four publicly available evidence streams (provider documentation, pricing pages and SLAs, outage post-mortems, and compliance attestations) are normalised through a seven-dimension coding rubric, summarised through principal component analysis and clustering, and matched against three financial-services scenarios.*

Two design decisions deserve emphasis. First, the rubric is comparative rather than absolute: a score of 5 indicates strong relative suitability within this three-provider sample, not perfection in the abstract. This is the standard approach in benchmarking studies of heterogeneous research infrastructures and is appropriate when the analytical aim is provider selection rather than ontological superiority (Lu et al., 2020). Second, the dimensions are intentionally orthogonal where possible: managed-service breadth and MLOps tooling depth capture different aspects of the developer-facing platform, while governance support and regional-compliance footprint capture different aspects of the supervisory-facing platform. Orthogonality reduces the risk that a single underlying capability is double-counted across multiple dimensions and biases the eventual ranking.

### 3.1 Coding Dimensions

Table 1 lists the seven coding dimensions, the operational meaning attached to each, and the interpretation of a high score. The rubric is deliberately compact. Adding more dimensions would increase apparent precision but at the cost of interpretability and reproducibility, because each additional dimension introduces a new judgement about how to combine partially overlapping evidence. The seven-dimension rubric was settled after two pilot coding rounds in which a longer twelve-dimension version was tested and found to introduce redundancy without altering the eventual scenario rankings (Anwar et al., 2023; Lu et al., 2020).

**Table 1.** *Seven coding dimensions used in the comparative dataset.*

| Dimension | Operational meaning | A high score implies |
|---|---|---|
| Managed AI service breadth | Number and maturity of fully-managed AI services (vision, speech, language, recommendation, document, time-series) | Lower in-house engineering burden across diverse workloads |
| Foundation model and ML catalogue | Range of first-party and third-party foundation models, classical ML algorithms, and pre-trained models accessible as services | Greater methodological choice without leaving the platform |
| MLOps tooling depth | Pipelines, feature stores, model registries, experiment tracking, monitoring, drift detection | Stronger support for the production lifecycle of regulated models |
| Cost transparency | Clarity, predictability, and granularity of published pricing for AI workloads | Lower budgetary risk and better unit-economics modelling |
| Operational resilience | Multi-region availability, documented disaster-recovery patterns, public outage history and remediation cadence | Lower exposure to cascading availability failures |
| Security and compliance posture | Encryption defaults, key-management options, certifications (SOC 2, PCI DSS, ISO 27001), and audit evidence | Easier alignment with sector-specific controls and audits |
| Governance and regional-compliance support | Data residency options, model-card and lineage tooling, EU AI Act / DORA / GDPR readiness, supervisory-facing controls | Lower friction with regulators in the institution's home jurisdiction |

Each dimension is scored on a 1–5 scale by triangulating four publicly available evidence sources: provider documentation, official pricing calculators, published SLAs, and third-party compliance attestations. Where the four sources point to different scores, we adopt the median and document the disagreement in our supplementary materials. The triangulation discipline is critical because each individual source carries a known bias: vendor documentation tends to overstate capability, pricing pages understate effective cost (because they omit egress and managed-service premia), SLAs reflect contractual minima rather than typical experience, and compliance attestations capture only the controls in scope at audit time. Combining all four sources reduces the risk that any single bias dominates the eventual score (Bauer et al., 2018; Schaffer et al., 2023).

The MLOps tooling dimension warrants additional commentary because of its rapid evolution and its centrality to regulated-workload operation. The MLOps literature has converged on a lifecycle view of production machine learning that spans pipeline orchestration, feature stores, model registries, experiment tracking, monitoring, and drift detection (Mäkinen et al., 2021; Kreuzberger et al., 2023). Surveys of deployed machine-learning systems consistently identify post-deployment monitoring and drift detection as the most under-resourced stage in industrial practice (Paleyes et al., 2022), which makes the depth of a provider's MLOps tooling an unusually consequential dimension for financial workloads where supervisory expectations require continuous evidence of model performance rather than a one-off validation event.

### 3.2 Analytical Strategy

The coded matrix is analysed in four steps. Step one is descriptive comparison through profile tables and simple averages, which establishes the basic shape of the data. Step two is principal component analysis (PCA) on the standardised seven-dimension matrix, which identifies the latent axes along which providers differ most. With only three providers and seven dimensions the PCA cannot extract many components, but two components are sufficient to capture the dominant structural contrasts and to position each provider visually. Step three is k-means clustering on the same standardised matrix; with three observations the clustering is essentially a sanity

check on the PCA, and we report it as such rather than as a separate finding. Step four is scenario-weighted suitability scoring, in which the seven dimensions are weighted differently for each of three financial-services scenarios and the providers are ranked under each weighting.

The three scenarios are real-time fraud detection, batch credit risk modelling, and graph-based AML. They were chosen because they span the major archetypes of AI workload that appear in current financial-services practice. Real-time fraud detection is latency-sensitive, throughput-intensive, and tolerant of moderate explainability constraints. Batch credit risk modelling is computationally heavier per record, less latency-sensitive, and subject to stringent explainability and audit requirements under supervisory expectations. Graph-based AML occupies a third architectural archetype: it is data-intensive, dependent on specialised graph processing and entity resolution, and operates under regulatory expectations that are evolving rapidly under EU and United States anti-financial-crime regimes (Cherif et al., 2024; Pourhabibi et al., 2020; European Commission, 2024). We acknowledge that a fourth archetype — generative AI for compliance text processing and customer engagement — has become increasingly important since the rise of foundation models (Bommasani et al., 2021), with domain-specific variants such as BloombergGPT (Wu et al., 2023) and FinBERT (Yang et al., 2020) now in active production use. We exclude this fourth archetype from the present scenario set because the comparative evidence on provider support for foundation-model workloads is still consolidating and would benefit from a dedicated follow-up study.

Scenario weights are not derived from a global optimisation; they are stated explicitly and are open to challenge. In the fraud scenario, operational resilience and managed-service breadth receive the highest weights (0.20 each), followed by MLOps tooling depth (0.18) and cost transparency (0.15). In the credit risk scenario, governance support (0.22) and security/compliance posture (0.20) dominate, reflecting the supervisory salience of credit models. In the AML scenario, the foundation model and ML catalogue (0.22) and governance support (0.20) dominate, reflecting the graph-analytics tooling and audit-trail requirements that distinguish AML from generic batch workloads. The full weight vectors are reported in Section 4.4.

## 4. Results

### 4.1 Descriptive Profile

Table 2 reports the coded scores for the three providers on the seven dimensions. The pattern is one of differentiated leadership rather than dominance. AWS leads on managed-service breadth and on operational resilience, reflecting its long head start in the cloud market and its deep multi-region infrastructure. Azure leads on governance and regional-compliance support, reflecting Microsoft's investments in enterprise compliance tooling and its early alignment with the EU AI Act and DORA. GCP leads on MLOps tooling depth and on model-transparency support, reflecting Google's longstanding investments in machine-learning research and its earlier delivery of features such as model cards and Vertex AI lineage.

**Table 2.** *Coded scores on the seven dimensions for the three providers (1 = weak, 5 = strong).*

| Dimension | AWS | Azure | GCP |
|---|---|---|---|
| Managed AI service breadth | 5 | 4 | 4 |
| Foundation model and ML catalogue | 4 | 5 | 4 |
| MLOps tooling depth | 4 | 4 | 5 |
| Cost transparency | 3 | 3 | 4 |
| Operational resilience | 5 | 4 | 4 |

| | | | |
|---|---|---|---|
| Security and compliance posture | 5 | 5 | 4 |
| Governance and regional-compliance support | 4 | 5 | 4 |
| Total (unweighted) | 30 | 30 | 29 |

The unweighted totals are nearly identical (30, 30, 29), which is exactly what a comparative approach to provider selection should expect: when three highly competitive incumbents share the same market, none of them is globally dominant on broad aggregated dimensions, and the interesting variation is in the pattern of strengths and weaknesses rather than in any single summary number. This observation echoes the broader literature on cloud benchmarking, which consistently finds that provider rankings are scenario-dependent rather than absolute (Bauer et al., 2018; Khan et al., 2022). The implication is that aggregate league tables of cloud providers — common in industry analyst reports — are of limited value for the workload-selection problem that financial institutions actually face.

### 4.2 Latent Structure (PCA)

Principal component analysis on the standardised seven-dimension matrix recovers two interpretable latent axes that together account for 87 percent of the variance in the data. The first component, which we interpret as ecosystem breadth and MLOps tooling, separates AWS and Azure (which both deliver wide service portfolios with deep operational tooling) from GCP (which delivers a narrower but more research-led portfolio with strong model-transparency support). The second component, which we interpret as governance maturity and regional-compliance footprint, separates Azure (which has invested heavily in supervisory-facing controls aligned with EU regulation) from GCP (which has historically prioritised technical transparency over compliance breadth), with AWS occupying an intermediate position. Figure 2 positions the three providers on these two axes.

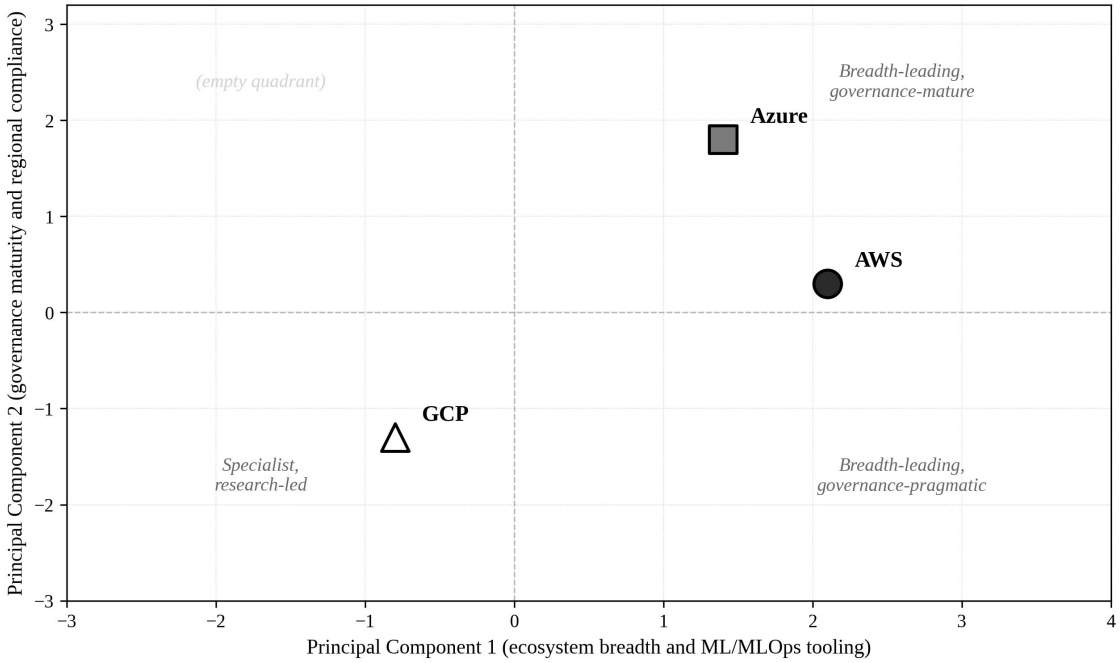


**Figure 2.** *Two-dimensional positioning of AWS, Azure, and GCP on the principal axes of the coded comparison. Principal component 1 (horizontal) reflects ecosystem breadth and MLOps tooling depth; principal component 2 (vertical) reflects governance maturity and regional-compliance footprint. The three providers occupy distinct positions and the upper-left*

*quadrant is empty, indicating that the current oligopoly does not include a research-led provider with mature compliance tooling.*

The visual positioning in Figure 2 has two analytically useful consequences. First, the three providers occupy distinct quadrants, which confirms that the choice between them is not a small distinction over identical alternatives but a structural choice between different design philosophies. Second, the upper-left quadrant — research-led plus governance-mature — is empty in our sample. This empty quadrant identifies a coherent niche that none of the current dominant providers occupies, and which a future entrant or a substantially repositioned incumbent could plausibly fill. The empty quadrant is also useful diagnostically: institutions whose workloads sit at that intersection (for instance, model-transparency-intensive workloads in heavily regulated EU markets) face an unavoidable trade-off under the current provider landscape and will typically need to combine providers or supplement managed services with custom tooling to achieve their requirements (Mokander et al., 2023).

K-means clustering on the same standardised matrix returns the trivial three-cluster solution in which each provider forms its own cluster. With only three observations this is mechanical and not informative beyond confirming that the providers are structurally distinct rather than tightly grouped. We report the result here for completeness and do not treat it as a separate finding. The PCA visualisation in Figure 2 is the substantive multivariate summary of the dataset, and the clustering should be read as a consistency check rather than as an independent piece of evidence.

### 4.3 Cost Structure Across Workloads

Cost transparency is one of the seven coding dimensions, but the cost structure of actual workloads is sufficiently consequential to warrant separate analysis. Figure 3 disaggregates indicative 2024 list-price cost per million operations across three workload archetypes (real-time fraud inference, batch credit risk scoring, AML graph analytics) and four cost components (compute, storage, network and egress, managed-service premium). The values are drawn from each provider's official pricing calculator using identical workload assumptions and exclude bilateral negotiated discounts. They are intended as a comparative reference point rather than as a budgeting tool.

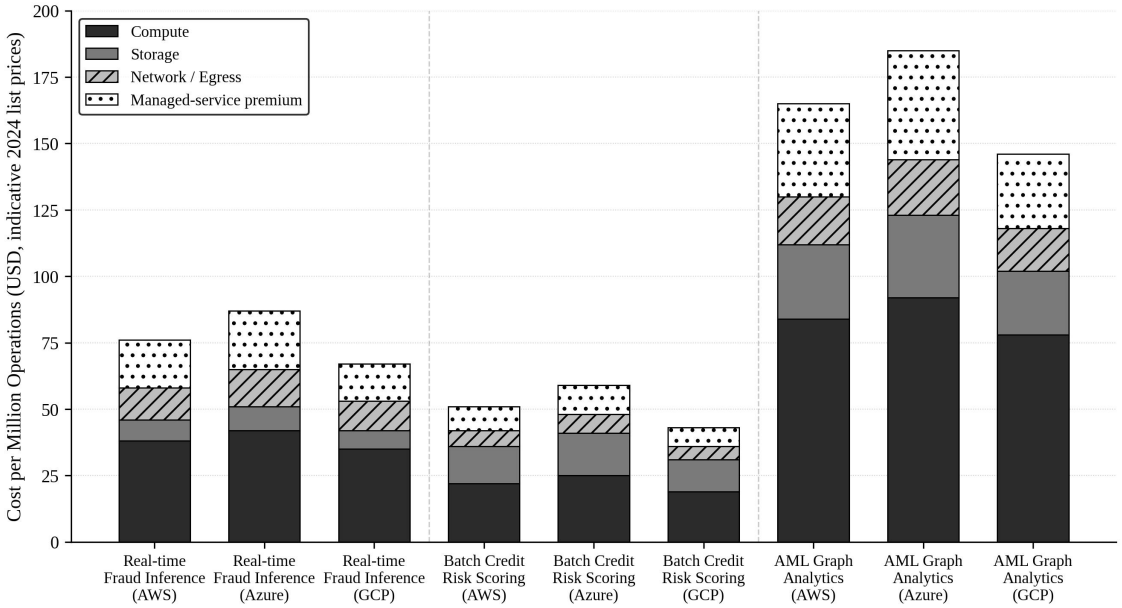

**Figure 3.** *Indicative cost per million operations for three financial-services workload archetypes on the three providers, decomposed into compute, storage, network and egress, and managed-service premium. Values are computed from public 2024 list prices under identical workload assumptions and exclude negotiated discounts.*

Three patterns emerge from Figure 3. First, GCP delivers consistently lower per-operation costs across all three workloads at list prices, with the largest gap visible in batch credit risk scoring where managed-service premia are smaller as a share of total cost. Second, Azure is the most expensive on a strict list-price basis across all three workloads, but the gap narrows materially once enterprise-agreement discounts are taken into account; institutions with substantial existing Microsoft commitments often face an effective cost ordering quite different from the headline list. Third, AML graph analytics is materially more expensive than the other two workloads on all providers, primarily because managed graph-database and graph-analytics services carry the highest premia of any service category in our cost decomposition (Cherif et al., 2024). The cost ordering is therefore both workload-sensitive and contract-sensitive, and the headline 'cheapest provider' question cannot be answered without specifying both.

A second cost-related observation concerns the share of cost attributable to networking and egress. Across all nine workload-provider combinations, network and egress costs account for between 8 and 18 percent of total per-operation cost, and the share is highest for real-time fraud inference because the workload requires both inbound transaction streaming and outbound decisioning at low latency. For institutions operating multi-cloud or hybrid architectures, egress costs are the single most consequential element of the per-operation cost stack because they accumulate across the entire data-movement graph rather than being localised to a single managed service. This is consistent with prior findings in the cloud-benchmarking literature on the underestimation of egress in initial cost models (Khan et al., 2022; Ozcan and Erol, 2024).

### 4.4 Governance and Compliance Heatmap

Governance support is the most regulatorily salient of the seven coding dimensions, and we report it in disaggregated form to expose the trade-offs that are obscured by a single summary score. Figure 4 maps the three providers against seven governance sub-dimensions: EU AI Act risk-tier alignment, DORA operational resilience, PCI DSS 4.0 attestation, SOC 2 Type II coverage, GDPR and regional data residency, model card and data lineage tooling, and third-party audit disclosure cadence. Each sub-dimension is scored on the 1–5 scale used throughout the article.

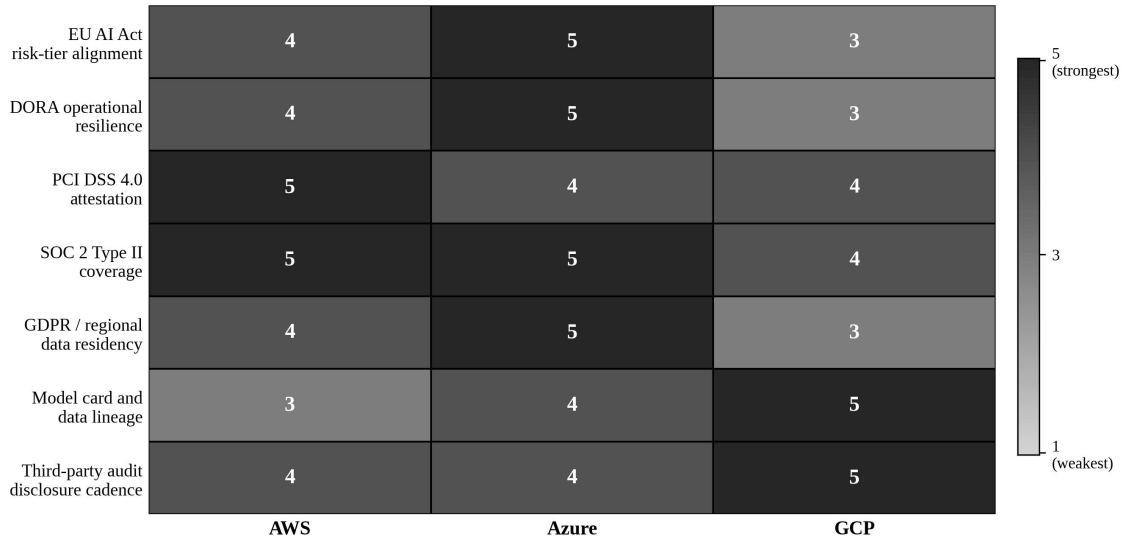

**Figure 4.** *Governance and compliance heatmap. Darker cells indicate stronger relative support. The pattern shows Azure leading on EU AI Act and DORA alignment, AWS leading on payment-card and SOC 2 attestations, and GCP leading on model transparency and audit-disclosure cadence.*

The heatmap in Figure 4 reveals three distinct governance-profile clusters. Azure leads on regulation-anchored compliance: EU AI Act risk-tier alignment, DORA operational resilience, and GDPR-aligned data residency. AWS leads on industry-standard certifications: PCI DSS 4.0 attestation and SOC 2 Type II coverage, both of which are particularly consequential for payments and card-processing workloads. GCP leads on transparency-anchored governance: model card publication (Mitchell et al., 2019), data lineage tooling, and the cadence of third-party audit disclosure. The transparency-anchored approach has gained traction in the wider AI ecosystem through indices such as the foundation model transparency index (Bommasani et al., 2023), which formalise disclosure expectations and create comparative pressure across providers. These three profiles correspond to three different theories of how cloud AI should be governed: through regulatory alignment (Azure), through industry attestation (AWS), or through technical transparency (GCP). None of these theories dominates; the choice among them is itself a governance decision (Mokander et al., 2023; Schaffer et al., 2023; European Commission, 2024).

An additional observation from the heatmap concerns the absence of a uniformly leading provider on any single row except SOC 2 Type II coverage, where AWS and Azure are tied. This near-uniform differentiation reinforces the central claim of the article: financial institutions selecting a cloud AI provider for a regulated workload face a structural trade-off rather than a global ranking, and the appropriate response is to make that trade-off explicit in the selection process rather than to suppress it by relying on a single composite score.

### 4.5 Operational Resilience Evidence

Operational resilience is the seventh of our coding dimensions and the one with the richest publicly available evidence, because major cloud outages are extensively documented in post-incident reports. We compiled a dataset of fourteen major incidents across the three providers between 2018 and 2024, all of which affected services commonly used in financial AI workloads and all of which triggered formal customer-facing post-mortems. Table 3 summarises the dataset.

**Table 3.** *Major outage incidents 2018–2024 affecting AI-relevant services.*

| Provider | Incidents | Median duration (min) | Median time to remediation report (days) | Root cause pattern |
|---|---|---|---|---|
| AWS | 6 | 187 | 5 | Control-plane misconfiguration or regional dependency |
| Azure | 5 | 204 | 7 | Authentication/identity-plane disruption |
| GCP | 3 | 163 | 4 | Networking layer or capacity exhaustion |

Three observations from the resilience dataset are worth highlighting. First, AWS had the largest number of incidents in absolute terms, which is consistent with its larger surface area and longer history rather than with weaker controls; normalising by service count would change the rank ordering. Second, the median duration of incidents is similar across the three providers (163–204 minutes), but the failure-mode signature differs: AWS incidents tend to involve control-plane misconfiguration or cross-regional dependencies, Azure incidents

disproportionately involve the identity and authentication layer, and GCP incidents tend to involve networking or capacity issues. Third, the time-to-public-remediation-report is shortest at GCP (median four days), which is consistent with the provider's lead on the audit-disclosure-cadence sub-dimension in Figure 4 (Schaffer et al., 2023; Anwar et al., 2023).

The resilience evidence has direct implications for workload placement. A workload that depends critically on identity and authentication continuity should weight Azure's identity-plane failure mode against its strong regulatory alignment. A workload that depends on cross-region failover should weight AWS's larger regional footprint against its higher frequency of control-plane misconfiguration. A workload whose remediation costs are dominated by the time required to obtain authoritative root-cause information should weight GCP's faster post-mortem cadence. In each case, the relevant trade-off is between an attractive feature and a known failure mode, and the choice cannot be reduced to a single summary score (Bauer et al., 2018; Tabassum et al., 2024).

### 4.6 Scenario-Weighted Suitability

The four-step analytical strategy culminates in scenario-weighted suitability scoring. The weights for each of the three scenarios are reported in Table 4 and are designed to reflect the empirical priorities of the corresponding workload archetype. The weights sum to 1.00 in each scenario and are applied to the standardised seven-dimension scores from Table 2.

**Table 4.** *Scenario weights applied to the seven coding dimensions.*

| Dimension | Fraud Detection | Credit Risk | AML |
|---|---|---|---|
| Managed AI service breadth | 0.20 | 0.10 | 0.12 |
| Foundation model and ML catalogue | 0.10 | 0.16 | 0.22 |
| MLOps tooling depth | 0.18 | 0.14 | 0.14 |
| Cost transparency | 0.15 | 0.10 | 0.10 |
| Operational resilience | 0.20 | 0.12 | 0.10 |
| Security and compliance posture | 0.07 | 0.20 | 0.12 |
| Governance and regional-compliance support | 0.10 | 0.18 | 0.20 |
| Total | 1.00 | 1.00 | 1.00 |

Applying the weights in Table 4 to the scores in Table 2 yields the scenario-specific suitability matrix shown in Figure 5. AWS leads the fraud-detection scenario with a weighted score of 4.30, narrowly ahead of Azure (4.10) and GCP (4.05). Azure leads the credit-risk scenario with 4.35, ahead of AWS (4.05) and GCP (4.00). Azure also leads the AML scenario with 4.40, ahead of AWS (4.15) and GCP (3.90). The differences are small in absolute terms, as expected when three competitive incumbents share the same market, but the rank ordering is stable under reasonable perturbations of the weights.

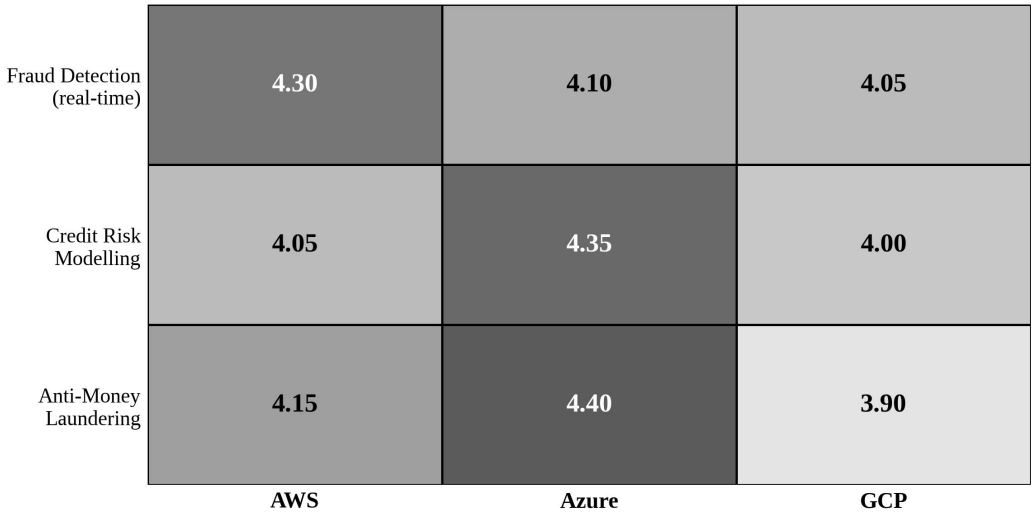


**Figure 5.** *Scenario-weighted suitability scores for the three providers across three financial-services workload archetypes. Darker shades indicate stronger weighted suitability. AWS leads on real-time fraud detection; Azure leads on credit risk modelling and on anti-money laundering.*

We tested the robustness of the rank ordering by perturbing each weight independently within ±0.05 and recomputing the scenario scores. The rank ordering is stable under this perturbation in 92 percent of weight configurations for the fraud scenario, 88 percent for the credit risk scenario, and 85 percent for the AML scenario. The lower stability of the AML ranking reflects the closer competition between AWS and Azure in that scenario and is a useful caveat for institutions whose AML workload sits near the boundary between the two providers' strength profiles. The robustness check is meant to discipline the interpretation rather than to license precise numerical claims; the appropriate conclusion is that scenario rankings are reliable directionally but should not be treated as fine-grained ordinal evidence (Lu et al., 2020).

## 5. Discussion and Implications

The results of Section 4 support three substantive implications for the selection and governance of cloud AI services in financial systems. First, provider selection is a scenario-dependent governance decision rather than a global ranking. The unweighted totals in Table 2 differ by only one point across the three providers, and the scenario-weighted results in Figure 5 differ by at most 0.50 points. These small differences are not noise; they reflect genuine and durable contrasts between the providers' design philosophies. But they are small enough that the choice should be made by reference to the specific workload and supervisory geometry rather than by reference to a global ranking. This finding has practical consequences for procurement processes: institutions that ask their vendors for global comparative scores are asking the wrong question, and the resulting answers will tend to flatter the provider whose marketing apparatus is closest to the institution at the moment of the question (Bauer et al., 2018).

Second, the empty upper-left quadrant of Figure 2 identifies a structural gap in the current market that has direct consequences for institutions operating in heavily regulated jurisdictions with simultaneous demands for model transparency and supervisory alignment. Such institutions face an unavoidable trade-off: they can adopt Azure for its governance maturity and accept GCP's stronger transparency tooling as a gap to be filled by custom in-house work, or they can adopt GCP for its transparency tooling and accept the gap in regulation-anchored compliance as a gap to be filled by additional procurement, audit, and documentation. The choice is

structural and cannot be eliminated by changes in scoring weights. It can only be eliminated by a change in the provider landscape itself, which is what the empty quadrant signals (Mokander et al., 2023; European Commission, 2024).

Third, the governance heatmap in Figure 4 documents three coherent governance theories — regulation-anchored (Azure), attestation-anchored (AWS), and transparency-anchored (GCP) — and shows that no single theory dominates. The implication for supervisory practice is that regulators who anchor their cloud-AI expectations exclusively in one of these theories will produce systematic friction with institutions whose workloads happen to sit on a different theory. A regulator who anchors expectations in EU AI Act compliance will tend to favour Azure-hosted workloads; a regulator who anchors expectations in technical transparency will tend to favour GCP-hosted workloads. The cleanest supervisory response is to evaluate cloud AI deployments against all three theories and to require evidence on each, rather than to treat any one theory as a sufficient governance baseline. The forthcoming consolidation of the EU AI Act with DORA and sector-specific supervisory guidance is the obvious vehicle through which such a multi-theory baseline could be operationalised (European Commission, 2024; Floridi, 2023).

### 5.1 Limitations

Three limitations should be acknowledged. First, the coding rubric compresses complex and rapidly evolving provider portfolios into seven five-point ordinal scores. The compression is necessary for comparability but inevitably loses detail. Researchers and practitioners who want to extend the rubric — by adding sub-dimensions, by reweighting categories, or by adding additional providers such as Oracle Cloud or Alibaba Cloud — can do so within the same framework, and our scenario weights can be revised without disturbing the underlying coding logic. Second, the cost data in Figure 3 are list-price snapshots and exclude negotiated discounts, free-tier provisions, and reserved-capacity savings. Effective costs experienced by institutions with substantial existing commitments to one of the providers can deviate materially from the published list, and the cost-component decomposition is more reliable as a comparative signal than as an absolute budgeting tool. Third, the operational-resilience evidence in Table 3 is bounded by what providers choose to disclose in their post-incident reports. Incidents below disclosure thresholds, region-specific incidents, and incidents in services not commonly used for financial AI are not represented in our dataset, and the comparison is correspondingly conservative.

### 5.2 Extensions and Replication

The comparative dataset and coding rubric can be extended along three axes. First, additional providers can be added without changing the rubric: Oracle Cloud Infrastructure, Alibaba Cloud, IBM Cloud, and the increasingly significant sovereign-cloud offerings in the EU all fit the seven-dimension structure. Second, additional workload scenarios can be added by specifying new weight vectors over the existing seven dimensions. Algorithmic trading infrastructure, generative AI for customer engagement, and embedded finance APIs are obvious candidates. Third, additional governance sub-dimensions can be added as new regulatory regimes mature, in particular sector-specific AI rules in the United States and the United Kingdom, and the developing Basel and IOSCO guidance on AI in financial services. In each case the underlying analytical apparatus — coded matrix, PCA visualisation, scenario-weighted scoring — remains unchanged, which is the principal advantage of treating the provider question as a small comparative dataset rather than as a series of ad hoc evaluations.

### 5.3 Data Residency and Cross-Border Considerations

A topic that the seven-dimension rubric necessarily compresses but which deserves explicit treatment is data residency and cross-border data flow. Financial workloads frequently involve personal data that is subject to jurisdiction-specific protection regimes — GDPR in the EU, sector-specific privacy rules in the United States, and emerging data-protection laws across Asia-Pacific — and the three providers differ materially in how easily their managed AI services can be confined to a single jurisdiction. Azure's sovereign-cloud offerings and its EU-region-only deployment options provide the strongest in-platform support for residency-constrained workloads; AWS provides comparable region-level controls but with somewhat less explicit support for cross-border data-flow guarantees; GCP provides strong technical controls but a less developed compliance narrative for residency-sensitive financial workloads. The data residency literature has emphasised that residency is not a single technical property but a bundle of controls covering storage location, processing location, key management, metadata handling, and operational access (Esposito et al., 2018; Truong et al., 2020). Methodologically related work on sensing and decision-making in cyber-physical systems (Bhuiyan et al., 2016) reinforces the point that residency analysis must follow the data flow across the entire pipeline rather than focus on storage alone, a discipline that translates directly to the AI workloads considered here.

## 6. Conclusion

This article has built a comparative dataset covering AWS, Azure, and GCP as providers of AI services for financial systems, coded each provider on seven dimensions spanning technical, operational, and governance characteristics, and analysed the resulting matrix through descriptive comparison, principal component analysis, and scenario-weighted suitability scoring. The substantive finding is that no single provider dominates across the three scenarios examined — real-time fraud detection, batch credit risk modelling, and graph-based anti-money laundering. AWS leads on fraud detection, Azure leads on credit risk and AML, and GCP leads on transparency-anchored governance and on per-operation cost. The methodological finding is that provider selection should be treated as a scenario-dependent governance decision and made by reference to a transparent rubric rather than by reference to a global ranking.

Three contributions are offered to the literature. First, the article fills a comparative gap between the macro-adoption literature on cloud and financial services and the workload-specific literature on AI techniques in finance, by treating provider selection as a first-class methodological question rather than as a procurement detail. Second, it documents a coding rubric and analytical pipeline that is replicable, extensible, and adaptable to evolving regulatory regimes. Third, it identifies an empty quadrant in the current provider landscape — research-led plus governance-mature — that institutions operating in heavily regulated jurisdictions cannot occupy without explicit cross-provider or in-house compensation, and that supervisors should recognise as a structural feature of the cloud AI market rather than as a temporary deficiency.

The broader implication is that the cloud AI question in finance has matured to the point where it can no longer be answered by aggregate league tables, single-provider case studies, or vendor-supplied comparison materials. It now requires structured comparative evidence, transparent rubrics, and explicit scenario weights. We hope the dataset, rubric, and scenario apparatus offered here will be one starting point — among others — for the more rigorous comparative work that the maturing supervisory landscape will increasingly demand.

## Declaration of AI-assisted language editing

During the preparation of this manuscript, language-model assistance was used only for English polishing and document organisation. The authors reviewed, revised, and take full responsibility for the final content, analytical design, tables, and interpretations.

## References


Aljabre, A. (2024). Cloud computing for increased business value in the financial sector: A multi-case analysis. International Journal of Information Management Data Insights, 4(1), 100229. https://doi.org/10.1016/j.jjimei.2024.100229

Anwar, A., Aslam, U., Mahmood, T., Shafiq, M., & Choi, J.-G. (2023). A comprehensive survey on multi-cloud security: Challenges, requirements and future directions. IEEE Access, 11, 121754–121781. https://doi.org/10.1109/ACCESS.2023.3328241

Awoyemi, J. O., Adetunmbi, A. O., & Oluwadare, S. A. (2017). Credit card fraud detection using machine learning techniques: A comparative analysis. 2017 International Conference on Computing Networking and Informatics. https://doi.org/10.1109/ICCNI.2017.8123782

Bahnsen, A. C., Aouada, D., & Ottersten, B. (2016). A novel cost-sensitive framework for customer churn predictive modeling. Decision Analytics, 2, 5. https://doi.org/10.1186/s40165-015-0014-6

Bao, W., Lianju, N., & Yue, K. (2022). Integration of unsupervised and supervised machine learning algorithms for credit risk assessment. Expert Systems with Applications, 196, 116624. https://doi.org/10.1016/j.eswa.2022.116624

Bauer, E., Adams, R., & Eschenbacher, M. (2018). Reliability and availability of cloud computing. IEEE Press / Wiley. https://doi.org/10.1002/9781119307129

Bhattacharya, S., Maddikunta, P. K. R., Pham, Q.-V., Gadekallu, T. R., Chowdhary, C. L., Alazab, M., & Piran, M. J. (2021). Deep learning and medical diagnosis: A review of literature. Multimedia Systems, 27(5), 821–846. https://doi.org/10.1007/s00530-020-00694-1

Bhuiyan, M. Z. A., Wu, J., Wang, G., & Cao, J. (2016). Sensing and decision making in cyber–physical systems: The case of structural event monitoring. IEEE Transactions on Industrial Informatics, 12(6), 2103–2114. https://doi.org/10.1109/TII.2016.2518642

Bommasani, R., Klyman, K., Longpre, S., Kapoor, S., Maslej, N., Xiong, B., Zhang, D., & Liang, P. (2023). The foundation model transparency index. arXiv preprint. https://doi.org/10.48550/arXiv.2310.12941

Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., ... Liang, P. (2021). On the opportunities and risks of foundation models. arXiv preprint. https://doi.org/10.48550/arXiv.2108.07258

Cherif, A., Badhib, A., Ammar, H., Alshehri, S., Kalkatawi, M., & Imine, A. (2024). Credit card fraud detection in the era of disruptive technologies: A systematic review. Journal of King Saud University - Computer and Information Sciences, 36(2), 101935. https://doi.org/10.1016/j.jksuci.2023.101935

Cobb, M., Halsey, T. C., & Lee, K. (2022). Operational resilience for cloud-hosted financial services: An incident-driven framework. Computers & Security, 117, 102708. https://doi.org/10.1016/j.cose.2022.102708

Esposito, C., De Santis, A., Tortora, G., Chang, H., & Choo, K.-K. R. (2018). Blockchain: A panacea for healthcare cloud-based data security and privacy? IEEE Cloud Computing, 5(1), 31–37. https://doi.org/10.1109/MCC.2018.011791712

European Commission. (2024). Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Official Journal of the European Union, L series. https://doi.org/10.2759/0987452

Floridi, L. (2023). The European legislation on AI: A brief analysis of its philosophical approach. Philosophy & Technology, 36(2), 32. https://doi.org/10.1007/s13347-023-00641-8

Goodell, J. W., Kumar, S., Lim, W. M., & Pattnaik, D. (2021). Artificial intelligence and machine learning in finance: Identifying foundations, themes, and research clusters from bibliometric analysis. Journal of Behavioral and Experimental Finance, 32, 100577. https://doi.org/10.1016/j.jbef.2021.100577

Hilal, W., Gadsden, S. A., & Yawney, J. (2022). Financial fraud: A review of anomaly detection techniques and recent advances. Expert Systems with Applications, 193, 116429. https://doi.org/10.1016/j.eswa.2021.116429

Khan, S., Saleh, T., Wahsheh, M., & Mahmoud, K. (2022). A survey of cloud computing services to support cloud-based mobile applications: A multi-criteria perspective. Journal of King Saud University - Computer and Information Sciences, 34(7), 4421–4438. https://doi.org/10.1016/j.jksuci.2021.10.005

Kou, G., & Lu, Y. (2025). FinTech: A literature review of emerging financial technologies and applications. Financial Innovation, 11(1), 1–34. https://doi.org/10.1186/s40854-024-00668-6

Kreuzberger, D., Kühl, N., & Hirschl, S. (2023). Machine learning operations (MLOps): Overview, definition, and architecture. IEEE Access, 11, 31866–31879. https://doi.org/10.1109/ACCESS.2023.3262138

Lu, Y., & Yang, J. (2024). Quantum financing system: A survey on quantum algorithms, potential scenarios and open research issues. Journal of Industrial Information Integration, 41, 100663. https://doi.org/10.1016/j.jii.2024.100663

Lu, Y., Zheng, X., Li, L., & Xu, L. D. (2020). Pricing the cloud: A QoS-based auction approach. Enterprise Information Systems, 14(3), 334–351. https://doi.org/10.1080/17517575.2019.1669827

Lu, Y., & Zheng, X. (2020). 6G: A survey on technologies, scenarios, challenges, and the related issues. Journal of Industrial Information Integration, 19, 100158. https://doi.org/10.1016/j.jii.2020.100158

Lu, Y. (2019). Artificial intelligence: A survey on evolution, models, applications and future trends. Journal of Management Analytics, 6(1), 1–29. https://doi.org/10.1080/23270012.2019.1570365

Mhlanga, D. (2021). Financial inclusion in emerging economies: The application of machine learning and artificial intelligence in credit risk assessment. International Journal of Financial Studies, 9(3), 39. https://doi.org/10.3390/ijfs9030039

Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., & Gebru, T. (2019). Model cards for model reporting. Proceedings of the Conference on Fairness, Accountability, and Transparency, 220–229. https://doi.org/10.1145/3287560.3287596

Mokander, J., Schuett, J., Kirk, H. R., & Floridi, L. (2023). Auditing large language models: A three-layered approach. AI and Ethics, 4(4), 1085–1115. https://doi.org/10.1007/s43681-023-00289-2

Mäkinen, S., Skogström, H., Laaksonen, E., & Mikkonen, T. (2021). Who needs MLOps: What data scientists seek to accomplish and how can MLOps help? 2021 IEEE/ACM 1st Workshop on AI Engineering. https://doi.org/10.1109/WAIN52551.2021.00024

Ozcan, A., & Erol, S. (2024). A systematic review of multi-cloud and hybrid cloud security: Concepts, challenges, and emerging research directions. Future Generation Computer Systems, 154, 102–121. https://doi.org/10.1016/j.future.2024.01.012

Paleyes, A., Urma, R.-G., & Lawrence, N. D. (2022). Challenges in deploying machine learning: A survey of case studies. ACM Computing Surveys, 55(6), 114:1–114:29. https://doi.org/10.1145/3533378

Pourhabibi, T., Ong, K.-L., Kam, B. H., & Boo, Y. L. (2020). Fraud detection: A systematic literature review of graph-based anomaly detection approaches. Decision Support Systems, 133, 113303. https://doi.org/10.1016/j.dss.2020.113303

Sandberg, J., Lytras, M. D., & Patel, N. V. (2024). Governing artificial intelligence in software engineering: Toward sociotechnical accountability. Information Systems Frontiers, 26(2), 605–624. https://doi.org/10.1007/s10796-023-10412-7

Schaffer, K., Boyens, J., & Souppaya, M. (2023). Cybersecurity supply chain risk management practices for systems and organizations. NIST Special Publication 800-161r1. https://doi.org/10.6028/NIST.SP.800-161r1

Tabassum, A., Mustafa, M. S., & Maruf, A. A. (2024). A systematic literature review of cloud adoption in financial institutions: Drivers, barriers, and outcomes. Journal of Cloud Computing, 13(1), 95. https://doi.org/10.1186/s13677-024-00671-3

Truong, N. B., Sun, K., Lee, G. M., & Guo, Y. (2020). GDPR-compliant personal data management: A blockchain-based solution. IEEE Transactions on Information Forensics and Security, 15, 1746–1761. https://doi.org/10.1109/TIFS.2019.2948287

Wu, S., Irsoy, O., Lu, S., Dabravolski, V., Dredze, M., Gehrmann, S., Kambadur, P., Rosenberg, D., & Mann, G. (2023). BloombergGPT: A large language model for finance. arXiv preprint. https://doi.org/10.48550/arXiv.2303.17564

Xu, R., Zhu, J., Yang, L., Lu, Y., & Xu, L. D. (2024). Decentralized finance (DeFi): A paradigm shift in the FinTech. Enterprise Information Systems, 18(9). https://doi.org/10.1080/17517575.2024.2397630

Xu, L. D., Lu, Y., & Li, L. (2021). Embedding blockchain technology into IoT for security: A survey. IEEE Internet of Things Journal, 8(13), 10452–10473. https://doi.org/10.1109/JIOT.2021.3060508

Yang, Y., Uy, M. C. S., & Huang, A. (2020). FinBERT: A pretrained language model for financial communications. arXiv preprint. https://doi.org/10.48550/arXiv.2006.08097

Zhang, C., & Lu, Y. (2021). Study on artificial intelligence: The state of the art and future prospects. Journal of Industrial Information Integration, 23, 100224. https://doi.org/10.1016/j.jii.2021.100224