

AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING

VOLUME: 02 ISSUE: 01

RECEIVED: JANUARY 13

REVISED: JANUARY 24

ACCEPTED: FEBRUARY 05

PUBLISHED: FEBRUARY 22



Optimizing Service Delivery Across Enterprise Data Center Infrastructures

Sophia Martinez

Asst Professor

Abstract

Agentic AI refers to autonomous systems that perceive their environment, reason about it, take actions to influence it, and learn from experience. Enterprise-grade data centers deliver services to multiple users through multimodal workloads running on shared infrastructure, operating as multitenant environments bound by strict service-level agreements and compliance requirements. Agentic AI introduces Level 2 autonomy for optimizing service delivery: control loops close the resulting decision-making boundaries, while policies govern permission models and ensure alignment with enterprise-defined constraints. Existing solutions, however, focus primarily on detecting and responding to anomalies rather than offering fully automated, formally defined agentic systems whose control loops span the entire service delivery stack.

An enterprise-grade data-center ecosystem generates extensive telemetry that characterizes the state of its components and the health of its service deliveries. Monitoring and observability enable the detection of deviations from expected behavior — deviations that require timely action to mitigate performance degradation or security threats, making control loops essential for optimizing service delivery. ML-based detection and prediction models trigger actions such as replacements, scaling, or routing adjustments when needed. Fault-tolerance mechanisms, including replication and checkpointing, allow the system to degrade gracefully during resource pull-outs, while agents continuously monitor system health to ensure these processes remain invisible to end users.

Keywords: Autonomy and control, enterprise data centers, governance, multi-tenancy, service delivery, trust .

1. Introduction

Enterprise-scale data centers deliver business critical services with stringent quality expectations under heavy regulatory constraints. Availability is paramount, with required MTTR and MTBF targets governing the provisioning of redundant resources to mitigate failure impact. Data interruption is a top risk concern, driving strict multitenant separation. Noncompliance with privacy, regulatory, and sustainability requirements has devastating consequences, so workloads must obey strict rules of deployment and execution. These characteristics are reflected in SLAs, which govern the

promises made by service providers to consumers. They have one objective: to secure business continuance.

The challenge is to operate these complex environments at the lowest cost. High asset utilization minimizes capital expenses (CAPEX), supporting the expected quality at the lowest possible price. Operating expenses (OPEX) incur every second, so moving resources out of the way and back again has a real impact. Energy costs are another critical factor, representing a high proportion of OPEX. Redundancy has been built into the environments to meet SLA targets, but when not in use, the energy cost to maintain and support them adds little to delivery while contributing a significant amount

AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING

VOLUME: 02 ISSUE: 01

RECEIVED: JANUARY 13

REVISED: JANUARY 24

ACCEPTED: FEBRUARY 05

PUBLISHED: FEBRUARY 22



to OPEX. The challenge is to provide the services at the lowest possible cost while meeting all requirements.

Metric	Description	Enterprise Impact	Optimization Objective
MTTR	Mean Time To Recovery	Faster restoration of failed services	Reduce downtime
MTBF	Mean Time Between Failures	Improves infrastructure reliability	Increase service continuity
SLA Compliance	Fulfillment of agreed service quality	Customer trust and governance	Maintain service guarantees
OPEX	Operational expenditure of infrastructure	Cost efficiency	Minimize operational costs
CAPEX	Capital investment utilization	Resource planning efficiency	Maximize infrastructure usage

AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING

VOLUME: 02 ISSUE: 01

RECEIVED: JANUARY 13

REVISED: JANUARY 24

ACCEPTED: FEBRUARY 05

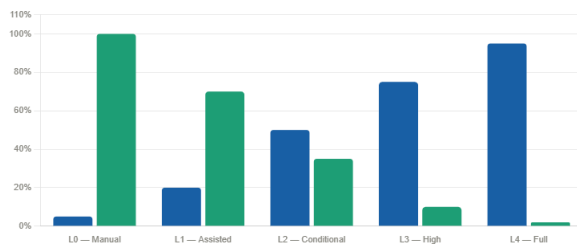
PUBLISHED: FEBRUARY 22

ISSN: 3067-1140



Metric	Description	Enterprise Impact	Optimization Objective
Energy Efficiency	Power utilization effectiveness	Sustainability and reduced energy costs	Lower power consumption

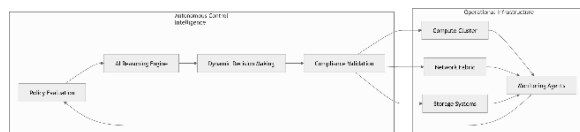
Table 1. Enterprise Data Center Service Delivery Metrics



Autonomy level vs efficiency maps the paper's autonomy spectrum (L0–L4) to projected efficiency gains and declining human intervention, reflecting the paper's central argument about delegating control to agentic systems.

2. Theoretical Foundations of Agentic AI in Datacenter Ecosystems

Agentic AI expresses a decision-making architecture that naturally realizes the spectrum of autonomous capabilities. It



leverages a combination of machine-actionable business policies and enterprise-scale multi-layer observability to determine autonomously how best to respond to locally-relevant events and conditions, in types of ways that inherently support the enterprise's objectives. Autonomous behaviours are expressed as closed-loop mechanisms that optimize service delivery outcomes, meet operating constraints and process policies, and comply with real-time information assurance and security requirements. Agentic-AI

Component	Primary Function	AI Capability	Outcome
Monitoring Layer	Collect telemetry and observability data	Real-time sensing	Infrastructure visibility
Resource Layer	Manage compute, storage, and networking	Dynamic orchestration	Resource optimization
Control Layer	Govern autonomous decisions	Policy-driven reasoning	Automated service delivery
AI Agents	Execute closed-loop actions	Autonomous decision-making	Faster response actions
Policy Engine	Enforce governance and compliance	Rule validation	Risk-controlled automation
Fault-Tolerance Module	Maintain continuity during failures	Predictive failover	Service resilience

systems, operating at a high level of autonomy, must support the responsive service delivery needs of data centre tenants while preserving confidentiality, securing data integrity, ensuring real-time availability against defined targets for mean-time-to-recovery and mean-time-between-failures, and achieving diverse property-level external compliance obligations.

Closed-Loop Autonomous Optimization Framework

AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING

VOLUME: 02 ISSUE: 01

RECEIVED: JANUARY 13

REVISED: JANUARY 24

ACCEPTED: FEBRUARY 05

PUBLISHED: FEBRUARY 22

ISSN: 3067-1140



Table 2. Agentic AI Functional Components in Data Center Ecosystems

2.1. Scope and Objective

The agentic AI service optimization domain spans enterprise-grade data center environments, acting on behalf of the business to automate resource management and control. Recognizing that physical/business ownership is not a definition of autonomy, the objective is to empower service delivery support with a company-aligned level of autonomy without undermining enterprise governance. Service-level agreement fulfillment, operational expenditure, capital expenditure, and energy efficiency constitute the key success factors. Accelerating service delivery at the defined level of service quality (including anti-fraud measures) remains an equally important, though secondary, objective. An agentic AI implementation is justified when successfully pursuing enterprise goals makes a practical, policy-governed delegation of control and responsibility possible.

Enterprise data centers typically provide the physical infrastructure, systems, and services for multiple business units and external customers. The service delivery objective

AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING

VOLUME: 02 ISSUE: 01

RECEIVED: JANUARY 13

REVISED: JANUARY 24

ACCEPTED: FEBRUARY 05

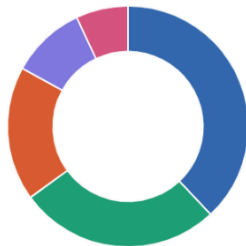
PUBLISHED: FEBRUARY 22



encompasses the entire resources-management-and-control spectrum from planning to service rollout and production operation. Service-delivery-management areas include incident and problem management (especially service restoration), change management, capacity management, availability management, disaster/recovery planning and testing, security management, and cloud-governance functions. Each area defines rules, processes, controls, monitoring requirements, and performance targets, such as mean time to repair (MTTR) and mean time between failures (MTBF), including upper and lower bounds for routing control based on disaster-/recovery-time-objective specifications.

the systems of these layers, while the information exchanged reflects the data required for autonomous decision-making during service delivery.

The Control Layer models how services provided to end users rely on the delivery and support of the systems within the Resource Layer, which in turn depend on input collected from the foundations of the Monitoring Layer. These dynamic dependencies are typically established and tracked through service topology repositories. All three layers provide and consume service classes, which describe and enable bulk processing of requests with similar resource and performance



OPEX breakdown visualizes the five cost categories the paper emphasizes — energy being the dominant driver (~38%), followed by labor, redundancy, networking, and compliance.

3. Architecture of Autonomous Service Delivery

Three system-level layers, their interfaces, and data flows among them are described next. These abstract away the inner workings of the lower-level systems and provide the necessary data for higher-layer decision-making. Such a layered description separates major components involved in generating and consuming the services rendered by the data center ecosystem, allowing for easy identification of data flow, processing, and decision-making required for autonomous service delivery. The interfaces between the layers connect semantically related functions performed by

AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING

VOLUME: 02 ISSUE: 01

RECEIVED: JANUARY 13

REVISED: JANUARY 24

ACCEPTED: FEBRUARY 05

PUBLISHED: FEBRUARY 22



attributes. The key difference is the set of optional requirements associated with each class. The Control Layer establishes controls and orients the Resource Layer in response to service policy and end-user requirements, while the Monitoring Layer acts as a supervisory and assurance mechanism, assessing coherence between expected and actual service delivery across both upper layers.

Telemetry Parameter	Monitoring Method	Detection Purpose	Associated Action
CPU Utilization	Real-time sensors	Overload detection	Dynamic scaling
Temperature	Thermal sensors	Hardware risk prevention	Cooling adjustment
Network Latency	Packet monitoring	Congestion analysis	Traffic rerouting
Storage I/O	Disk monitoring	Bottleneck identification	Resource balancing
Security Events	Log analytics	Threat detection	Isolation and mitigation
Power Consumption	Energy telemetry	Sustainability monitoring	Energy optimization

Table 3. Monitoring and Observability Parameters

3.1. Perception and Monitoring

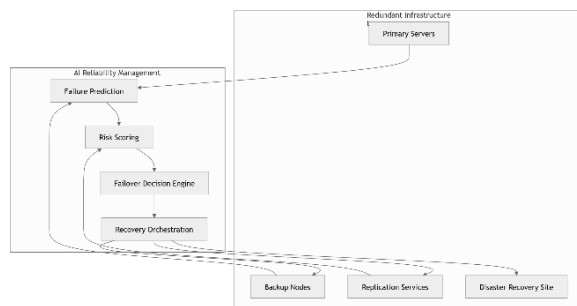
Data delivery requires information about the real-time state of the actors in the ecosystem. The capability for perception and monitoring enables active generation of telemetry by dedicated sensors as well as passive availability of operational state information and monitoring of ad-hoc feedback loops.

Dedicated sensors measure specific observability

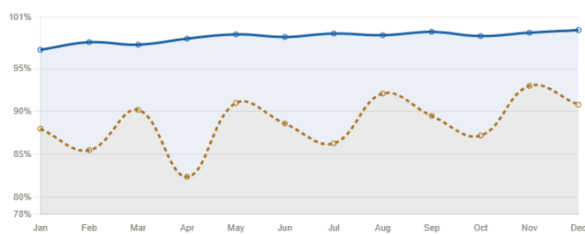
primitives every so often in a predictable manner. Sample frequency is determined based on quantization noise and Nyquist-Shannon conditions. Such metrics may be declared events-firing when crossing a defined threshold, e.g., sudden drop in temperature. A variety of telemetry is produced by other elements in the ecosystem accessible for surveillance and visibility via standard interfaces, drivers, protocols. Data presenting a quality below a predefined threshold must not be utilized. Quality can be attached to data based either on the information origin (e.g., monitored data not coming from a healthy topology) or on computative estimation (e.g., associated with an external monitoring mechanism). Anomalous situations must be detected in real-time (with ideally low latency) and corrected since they will degrade the overall feedback



mechanisms. Estimating, in real-time, the actual operational state of systems within the environment and what authentic operational information is accessible with an appropriate level of confidence is fundamental for data integrity and for allowing unaffected services to keep following tolerable paths.



Intelligent Fault-Tolerance and Recovery System



SLA compliance over time contrasts agentic AI-managed environments (~97–99%) against traditional management (~82–93%), directly tied to the paper's focus on service-level agreement fulfillment.

4. Data Environments and Requirements

Enterprise-grade data centers typically have an ecosystem setup supporting multiple business tenants and their services. The internal structure is complex but hidden from the customer. The economic model used is the private-cloud model, charging each internal business unit as a service provisioning provider. A private cloud hosted in an on-premise data center running business-critical applications is a

Service delivery optimization is necessary to meet the requirements mentioned earlier. Workload archives of the data centers provide a huge inspiration to predict future behavior. The optimal “what to deliver” and “how to deliver” are pattern-matching problems. Recognizing the patterns during service delivery is vital to keeping the overhead low. Delivering the right model with the correct file path saves energy. A growing workload in the finance domain can be predicted using the workload archives. Recognition of abnormal situations is equally vital, and action tailoring is needed. Detecting an abnormal situation while service delivery and orchestrating a failover without alerting the customer lowers the perceived service downtime.

Control Loop Stage	Input	Processing Activity	Autonomous Output
Perception	Telemetry streams	Data aggregation	System state awareness
Analysis	Historical and live data	ML prediction models	Risk forecasting
Decision	SLA and policy constraints	Agentic reasoning	Corrective action selection
Execution	Infrastructure commands	Automated orchestration	Failover/scaling/routing
Feedback	Operational results	Continuous learning	Policy refinement

Table 4. Closed-Loop Autonomous Control Operations

Mathematical Formulas:

- SLA Compliance Optimization

$$A_v \times R_t$$

AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING

VOLUME: 02 ISSUE: 01

RECEIVED: JANUARY 13

REVISED: JANUARY 24

ACCEPTED: FEBRUARY 05

PUBLISHED: FEBRUARY 22



multimillion-euro investment. Ideally, the customers expect the data centers to deliver the higher value at a reduced cost. Nonetheless, data centers do not resemble the utility power-providing companies, where customers are not concerned with resource utilization, infrastructure planning, or

operations. Regulatory bodies like PCI-DSS, HIPAA, ISO, and SOX must be taken into consideration during service delivery, making service velocity and reliability difficult to achieve. Furthermore, high service delivery reliability is needed for a longer period; hence, reliability and fault-tolerance patterns must be established during the service delivery design stage.

$$SLA_{score} = L_a + E_r$$

Where:

A_v = Availability,

R_t = Response Time Efficiency,

L_a = Latency,

E_r = Error Rate.

2. Resource Utilization Efficiency

$$RUE = \frac{U_{cpu} + U_{mem} + U_{net}}{3}$$

AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING

VOLUME: 02 ISSUE: 01

RECEIVED: JANUARY 13

REVISED: JANUARY 24

ACCEPTED: FEBRUARY 05

PUBLISHED: FEBRUARY 22



$$F_t = 1 - P^n$$

Where:

 U_{cpu} = CPU utilization, U_{mem} = Memory utilization, U_{net} = Network utilization.

3. Autonomous Control Loop Function

$$C(t + 1) = C(t) + \alpha(D_t - O_t)$$

Where:

 $C(t)$ = Current control state, D_t = Desired state, O_t = Observed state, α = Learning coefficient.

4. Mean Time to Recovery Optimization

$$MTTR = \frac{\sum T_r}{N_f}$$

Where:

 T_r = Recovery time for each failure, N_f = Number of failures.

5. Reliability Estimation Model

$$R(t) = e^{-\lambda t}$$

Where:

 λ = Failure rate, t = Operational time.

6. Energy Efficiency Function

$$E_{eff} = \frac{W_p}{P_c}$$

Where:

 W_p = Work processed, P_c = Power consumed.

7. Fault Tolerance Probability

Where:

 P_f = Failure probability of one node, n = Number of redundant nodes.

8. Predictive Scaling Equation

$$S_r = \beta W_d + \gamma H_t$$

AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING

VOLUME: 02 ISSUE: 01

RECEIVED: JANUARY 13

REVISED: JANUARY 24

ACCEPTED: FEBRUARY 05

PUBLISHED: FEBRUARY 22



Where:

S_r = Scaling requirement,

W_d = Workload demand,

H_t = Historical load,

β, γ = Scaling weights.

9. Service Delivery Latency Model

$$L_s = T_q + T_p + T_n$$

Where:

T_q = Queue delay,

T_p = Processing delay,

T_n = Network delay.

10. Governance Compliance Score

$$G_c = \frac{P_a}{t} \times 100$$

Where:

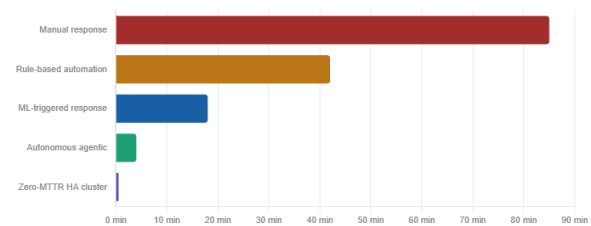
P_a = Policies adhered to,

P_t = Total policies.

4.1. Reliability, Availability, and Fault Tolerance

Reliability, availability, and fault-tolerance capabilities constitute core requirements in enterprise-grade data center systems. Service continuity cannot be compromised, even with temporarily degraded performance. Consequently, service-lines are governed by targets for Mean Time To Recovery (MTTR) and Mean Time Between Failures (MTBF) that shapes fault-tolerance design and deployment decisions. MTTR requirements define the maximum duration that a service can be unavailable after a service instance failure. The design envelope for availability establishes the acceptable ratio of uptime to downtime across a service instance's complete

operating timescale. Resilience objectives synthesize MTTR and MTBF targets into resource provisioning automations that minimize the impact of failure on service reliability. Data center solutions, therefore, implement redundancy theorems that recommend preserve-active strategies for MTTR-driven resources and preserve-idle schemes for MTBF-governed assets.



MTTR by control loop type illustrates the paper's discussion of fault-tolerance strategies, from fully manual (85 min) down to P

AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING

VOLUME: 02 ISSUE: 01

RECEIVED: JANUARY 13

REVISED: JANUARY 24

ACCEPTED: FEBRUARY 05

PUBLISHED: FEBRUARY 22



to zero-MTTR high-availability clusters using closed-loop mechanisms.

5. Control Loops for Service Delivery Optimization

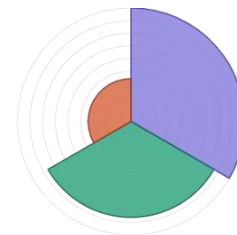
Focus on structured, quality content rather than length; concise, impact-oriented writing is essential. Consider depth, breadth, and coverage as important as word count.

Closed control loops automatically adjust resources, routes, and policies to enhance service delivery velocity and reliability. Services can be overprovisioned to accelerate delivery, but resource use remains high until demand probes reduce it. Within predictable limits, service delivery velocity can be increased by relaxing SLA deadlines, such as response times. For oversubscription to be feasible, data paths between shared components must have enough spare capacity. Resources must also be able to follow transient demand.

Rules define routing paths and define constraints at the data path level. They are enforced by programmable devices such as routers and switches, as well as agents that manage distributed data path protocol mechanisms. Routing rules may direct data traffic to optimise resource use or to collect telemetry data for management. External requirements, such as rules of the financial regulator, may require traffic flows to adhere to certain routes. Agentic AI systems in charge of all elements in the path may adjust the rules, but approval may be required before introducing conflicts.

Strategy	Operational Purpose	Enterprise Benefit	Example
Active Redundancy	Immediate failover support	Minimal downtime	High-availability clusters
Passive Redundancy	Standby resource protection	Lower operational overhead	Idle backup servers
Checkpointing	Recovery state preservation	Faster restoration	VM state snapshots
Replication	Data continuity	High data availability	Distributed storage

Table 5. Reliability and Fault-Tolerance Strategies



Policy governance distribution maps the paper's three-category policy framework: operational boundary policies, continuous governance policies, and human-intervention triggers.

5.1. Policy-Based Governance

Governance frameworks are essential to ensuring enterprise environments function properly. Policies enable oversight of systems that represent clients and sensitive data. In addition, policies help data centers comply with government regulations. Enterprise goal alignment, risk tolerance quantification, compliance requirements, service level objectives, and the trade-offs between operational excellence and service reliability ultimately dictate policy definitions. Machine-learning systems make, execute, and learn from decisions but not policy definitions. Externalization and decoupling support continual policy adjustments without knowledge-disrupting retraining.

Autonomous service delivery policies fall into three categories: those that define operational boundaries, those that govern services continuously, and those that determine when decisions require human intervention. Operational boundaries account for service-level agreements (SLAs) of different clients, risk proclivities regarding managed lead times, regulatory compliance, and governance framework requirements. Policy systems operate these controls across clusters of agentic controllers whenever non-compliance occurs. Policy systems also specify the conditions for the various service delivery Open-Closed cycles by defining the success variables for clients requiring continuous service delivery. When the agentic AI delivers services without Open-

AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING

VOLUME: 02 ISSUE: 01

RECEIVED: JANUARY 13

REVISED: JANUARY 24

ACCEPTED: FEBRUARY 05

PUBLISHED: FEBRUARY 22



Closed interruption, services usually operate within these prescribed delivery-pattern tolerances. However, tolerance levels can still be violated if a delivery success enters a certain predetermined delivery-risk zone. Policy systems determine when these zones have been breached and human intervention becomes critical for delivery success or outlier-variance management.

AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING

VOLUME: 02 ISSUE: 01

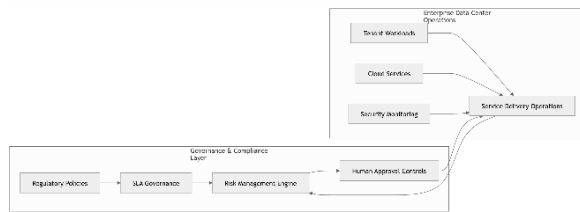
RECEIVED: JANUARY 13

REVISED: JANUARY 24

ACCEPTED: FEBRUARY 05

PUBLISHED: FEBRUARY 22

ISSN: 3067-1140



Multi-Tenant Governance and Compliance Architecture

6. Conclusions

Policy Category	Governance Focus	Enforcement Mechanism	Compliance Outcome
Operational Boundaries	SLA and risk limits	Threshold-based controls	Controlled autonomy
Regulatory Compliance	PCI-DSS, HIPAA, SOX	Continuous auditing	Legal adherence
Human Intervention Policies	Escalation conditions	Approval workflows	Risk mitigation
Security Policies	Data integrity and privacy	Access governance	Secure operations
Resource Policies	Capacity and energy rules	Intelligent orchestration	Sustainable utilization

AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING

VOLUME: 02 ISSUE: 01

RECEIVED: JANUARY 13

REVISED: JANUARY 24

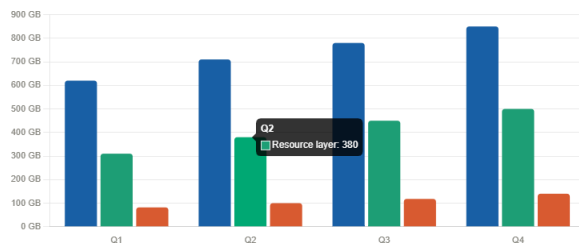
ACCEPTED: FEBRUARY 05

PUBLISHED: FEBRUARY 22



Autonomous agents can decisively contribute to the services provided by enterprise-grade data center ecosystems. Service delivery optimization, non-linear energy-efficiency improvements, significant OPEX reduction, and the fulfilment of the architectural imperatives of vendor-agnostic observability, enterprise-governance alignment, agent proliferation, conservative autonomy, and transparent knowledge for impact analysis are some of the key aspects addressed by the aforementioned frameworks, components, surfaces, artifacts, and capabilities. Defining and implementing the control loops form the critical next step toward a successful and practical realization of agentic autonomy.

Past service delivery velocity and reliability levels can be surpassed without compromising the equipment principal stakeholders value. Focus on measurable goals—SLAs, OPEX, CAPEX, and energy efficiency—and integration with existing operation mode expedite assessment and deployment of the various facets of agentic AI. Scaling to a superset of enterprise-data-center efficiency- and OPEX-related requirements endows both humanexpert- and operator-agnostic surfaces and augments service delivery with the appropriate responsiveness and flexibility.



Telemetry volume by architecture layer reflects the paper's three-layer architecture (Monitoring → Resource → Control), showing how data volume decreases as it moves up the stack toward decision-making.

AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING

VOLUME: 02 ISSUE: 01

RECEIVED: JANUARY 13

REVISED: JANUARY 24

ACCEPTED: FEBRUARY 05

PUBLISHED: FEBRUARY 22

ISSN: 3067-1140



Table 6. Governance and Compliance Policy Categories

References

1. Liang, B., Wu, D., Wu, P., & Su, Y. (2021). An energy-aware resource deployment algorithm for cloud data centers based on dynamic hybrid machine learning. *Knowledge-Based Systems*, 222, 107020.
2. Nawrocki, P., & Osypanka, P. (2021). Cloud resource demand prediction using machine learning in the context of QoS parameters. *Journal of Grid Computing*, 19, 20.
3. Yang, Z., Du, J., Lin, Y., Du, Z., Xia, L., Zhao, Q., & Guan, X. (2022). Increasing the energy efficiency of a data center based on machine learning. *Journal of Industrial Ecology*, 26(1), 323–335.
4. Paleti, S. (2023). AI-driven innovations in banking: Enhancing risk compliance through advanced data engineering. Available at SSRN, 5244840.
5. Tuli, S., Gill, S. S., Xu, M., Garraghan, P., Bahsoon, R., Dustdar, S., Sakellariou, R., Rana, O., Buyya, R., Casale, G., & Jennings, N. R. (2022). HUNTER: AI based holistic resource management for sustainable cloud computing. *Journal of Systems and Software*, 184, 111124.
6. Wang, Y., Li, Y., Wang, T., & Liu, G. (2022). Towards an energy-efficient data center network based on deep reinforcement learning. *Computer Networks*, 210, 108939.
7. Ounifi, H.-A., Gherbi, A., & Kara, N. (2022). Deep machine learning-based power usage effectiveness prediction for sustainable cloud infrastructures. *Sustainable Energy Technologies and Assessments*, 52, 101967.
8. Mahbod, M. H. B., Chng, C. B., Lee, P. S., & Chui, C. K. (2022). Energy saving evaluation of an energy efficient data

AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING

VOLUME: 02 ISSUE: 01

RECEIVED: JANUARY 13

REVISED: JANUARY 24

ACCEPTED: FEBRUARY 05

PUBLISHED: FEBRUARY 22



- center using a model-free reinforcement learning approach. *Applied Energy*, 322, 119392.
9. Nandan, B. P. Data Analytics-Driven Approaches to Yield Prediction in Semiconductor Manufacturing. *International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering (IJIREEICE)*, DOI, 10.
 10. Karat, C., & Senthilkumar, R. (2022). Optimal resource allocation with deep reinforcement learning and greedy adaptive firefly algorithm in cloud computing. *Concurrency and Computation: Practice and Experience*, 34(4), e6657.
 11. Panwar, R., & Supriya, M. (2022). Dynamic resource provisioning for service-based cloud applications: A Bayesian learning approach. *Journal of Parallel and Distributed Computing*, 168, 90–107.
 12. Rawas, S., Zekri, A., & El-Zaart, A. (2022). LECC: Location, energy, carbon and cost-aware VM placement model in geo-distributed DCs. *Sustainable Computing: Informatics and Systems*, 33, 100649.
 13. Karmakar, K., Banerjee, S., Das, R. K., & Khatua, S. (2022). Utilization aware and network I/O intensive virtual machine placement policies for cloud data center. *Journal of Network and Computer Applications*, 205, 103442.
 14. Mashetty, S. Data-driven insights for community investment through mortgage-backed securities. *International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering (IJIREEICE)*, DOI, 10.
 15. Nawrocki, P., & Osypanka, P. (2022). Machine learning-centric resource management in cloud computing: A review and future directions. *Journal of Network and Computer Applications*, 204, 103405.
 16. Wang, S., Qin, L., Ma, C., & Wu, W. (2023). Research on overall energy consumption optimization method for data center based on deep reinforcement learning. *Journal of Intelligent & Fuzzy Systems*, 44(5), 7333–7349.
 17. Mao, L., Chen, R., Cheng, H., Lin, W., Liu, B., & Wang, J. Z. (2023). A resource scheduling method for cloud data centers based on thermal management. *Journal of Cloud Computing*, 12, 84.
 18. Wang, T., Fan, X., Cheng, K., Du, X., Cai, H., & Wang, Y. (2023). Parameterized deep reinforcement learning with

AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING

VOLUME: 02 ISSUE: 01

RECEIVED: JANUARY 13

REVISED: JANUARY 24

ACCEPTED: FEBRUARY 05

PUBLISHED: FEBRUARY 22



- hybrid action space for energy efficient data center networks. *Computer Networks*, 235, 109989.
19. Karthikeyan, R., & Sundaravadivazhagan, B. (2023). COSCO2: AI-augmented evolutionary algorithm based workload prediction framework for sustainable cloud data centers. *Transactions on Emerging Telecommunications Technologies*, 34(1), e4652.
 20. Adusupalli, B. Strengthening Cybersecurity Compliance: Evaluating Risk Assessment Frameworks and AI-Based Anomaly Detection in Audit Practices.
 21. Dogani, J., Khunjush, F., & Seydali, M. (2023). Host load prediction in cloud computing with discrete wavelet transformation (DWT) and bidirectional gated recurrent unit (BiGRU) network. *Computer Communications*, 198, 157–174.
 22. Maiyza, A. I., Korany, N. O., Banawan, K., Hassan, H. A., & Sheta, W. M. (2023). VTGAN: Hybrid generative adversarial networks for cloud workload prediction. *Journal of Cloud Computing*, 12, 97.
 23. Gupta, A., & Bhadauria, H. S. (2023). Workload prediction for SLA performance in cloud environment: ESANN approach. *Intelligent Decision Technologies*, 17(4).
 24. Xiu, X., Li, J., Long, Y., & Wu, W. (2023). MRLCC: An adaptive cloud task scheduling method based on meta reinforcement learning. *Journal of Cloud Computing*, 12, 75.
 25. Funika, W., Koperek, P., & Kitowski, J. (2023). Automated cloud resources provisioning with the use of the proximal policy optimization. *The Journal of Supercomputing*, 79, 6674–6704.
 26. Yao, Z., Ding, Z., & Clausen, T. H. (2022). Multi-agent reinforcement learning for network load balancing in data center. *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, 3594–3603.
 27. Yang, D., Yu, J., Du, X., He, Z., & Li, P. (2022). Energy saving strategy of cloud data computing based on convolutional neural network and policy gradient algorithm. *PLOS ONE*, 17(12), e0279649.
 28. Alhazmi, L. (2023). A deep reinforcement learning-based model for optimal resource allocation and task scheduling in cloud computing. *International Journal on Recent and Innovation Trends in Computing and Communication*, 11(8s),

AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING

VOLUME: 02 ISSUE: 01

RECEIVED: JANUARY 13

REVISED: JANUARY 24

ACCEPTED: FEBRUARY 05

PUBLISHED: FEBRUARY 22



97–107.

29. Recharla, M., & Chitta, S. AI-Enhanced Neuroimaging and Deep Learning-Based Early Diagnosis of Multiple Sclerosis and Alzheimer's.
30. Laxmi, H. M. V., & Nagesh, P. (2023). Adaptive resource allocation in cloud data centers using actor-critical deep reinforcement learning for optimized load balancing. *International Journal on Recent and Innovation Trends in Computing and Communication*, 11(5s), 310–318.