 ISSN: 3139-1141

Emerging Trends and Challenges in Cloud Computing: A Study of Security, Scalability, and Future Technologies

Omkar Prabhakar Panchal[#1], Monika Madhavrao Pande[#2], Aaditi Vijay Patil[#3]

[#]*Zeal Institute of Business Administration, Computer Application and Research, Savitribai Phule Pune University, Pune, India*

Under the guidance of Prof. Vishwas Kenchi, Academic Year 2026-2027

[1]omkarpanchal280@gmail.com (corresponding author), [2]monikasunilpatil@gmail.com, [3]paaditi722@gmail.com

ORCID: 0009-0007-4891-4046 (Panchal); 0009-0002-3373-3721 (Pande); 0009-0003-9588-2244 (Patil)

***Abstract*—**

Cloud computing is the underlying technology of modern digital services. However, it has evolved well beyond renting virtual machines. Cloud providers are now offering machine learning platforms and foundation models; organizations are running workloads in a hybrid, multi-cloud environment; workloads are being run at the network edge; code is being executed as event-driven functions; and energy consumption of data centres is becoming a design consideration. While all of these trends offer the benefits of scalability and flexibility, they also introduce new risks such as data exposure due to misconfiguration, vendor lock-in, management complexity, and increased compute and energy consumption. This paper is a secondary research study. It provides a literature review of academic papers, standards, and industry publications in three focus areas: cloud security and privacy, integration of artificial intelligence (AI) with cloud, and hybrid and multi-cloud computing. Related technologies such as edge computing, serverless computing, and green cloud computing are also covered. A Cloud Trends and Challenges Evaluation Framework is presented along with two procedures: source screening algorithm and weighted scoring algorithm for comparing traditional (single public cloud), hybrid, multi-cloud computing approaches and major cloud providers (Amazon Web Services (AWS), Microsoft Azure, Google Cloud Platform (GCP)). The literature review shows that the major causes of security incidents in cloud computing are configuration and identity weaknesses rather than failures in provider infrastructure. Multi-cloud reduces vendor lock-in but increases operational complexity. Cost and energy consumption are visible design constraints for AI workloads. An illustrative weighted comparison shows that the ranking of deployment models can be different depending on the weight of vendor lock-in. Thus, the optimal approach depends on the priorities of an organization. This paper also highlights the limitation of our approach: no primary measurements are taken and the comparison is qualitative. Thus, the comparison needs to be validated with benchmarks and case studies.

***Keywords*—Cloud Computing; Cloud Security; Data Privacy; Artificial Intelligence; Machine Learning; Hybrid Cloud; Multi-Cloud; Edge Computing; Serverless Computing; Green Cloud Computing; Vendor Lock-in; Scalability.**

I. INTRODUCTION

Cloud computing is the on-demand delivery of computing, storage and application resources over a network. Cloud computing is defined by the National Institute of Standards and Technology (NIST) as a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources that can be rapidly provisioned and released with minimal management effort [1]. Cloud computing is characterized by five essential characteristics: on-demand self-service, broad network access, resource pooling, rapid elasticity, and measured service. It is provided in three service models: Infrastructure as a Service (IaaS), Platform as a Service (PaaS) and Software as a Service (SaaS) [1].

Historically, organizations purchased and managed their own servers. The capacity needs to be predicted months in advance. Underestimation of demand leads to downtime. Overestimation of demand leads to idle expensive hardware. Armbrust et al. explain the attraction of cloud as the illusion of infinite resources on demand. Cloud eliminates up-front commitment and provides pay-per-use on a short term basis [2]. Cloud computing today is much richer than what they described.

With the richer cloud, comes the need to evaluate success differently. Simply adopting a cloud service does not mean that the service is secure, efficient or cost effective. For example, a publicly accessible storage bucket may lead to a data breach. An unoptimized auto-scaling policy may lead to unexpected cost. A model training job may consume more energy than expected. Thus, a good study on cloud computing needs to consider both opportunities and limitations. It needs to evaluate each technology on the basis of its impact on security, scalability, cost, portability and energy consumption.

A. Problem Statement

Cloud computing landscape is more powerful but also more complex. Cloud enables data availability. However, misconfiguration or poor access control can increase risk of data breach. Multiple cloud providers can be used to reduce vendor lock-in. However, multiple providers may lead to interoperability issues and more management overhead. AI enabled services can provide automation and analytics. However, they require large computing resources. There are privacy and energy consumption concerns.

Information on emerging trends is available in the form of academic papers, industry reports and vendor documentation. Often, the information is disjointed and contains varying or biased perspectives. This makes it difficult for students and practitioners to have an integrated, balanced and up-to-date view. There is a need for a structured comparison of major

emerging trends and deployment approaches and providers. This needs to be done on the basis of common criteria.

B. Aim

The aim of this study is to analyze emerging trends and challenges in cloud computing with focus on security and privacy, AI integration and hybrid and multi-cloud computing. This is done by reviewing existing research and comparing major deployment models and cloud platforms. This is to identify opportunities, limitations and future directions.

C. Objectives

1. Study latest trends in cloud computing.
2. Identify major security and privacy challenges in cloud environment.
3. Understand the role of AI and machine learning in cloud platforms.
4. Compare traditional cloud, hybrid cloud and multi-cloud approaches on common criteria.
5. Compare services and features of AWS, Azure and Google Cloud.
6. Explore related trends such as edge computing, serverless computing and green cloud computing.
7. Explore future opportunities and challenges.
8. Provide recommendations for secure, efficient and sustainable cloud adoption.

D. Research Questions

1. What are the emerging trends in cloud computing? What are the benefits of each emerging trend in terms of scalability, flexibility and cost?
2. What are the common security and privacy risks as reported in literature? What is the proportion of risks that are due to customer side configuration and identity and not due to provider infrastructure?
3. How do single public cloud, hybrid cloud and multi-cloud approaches compare on security, cost, flexibility, vendor lock-in and interoperability?
4. Does the choice among AWS, Azure and Google Cloud depend on technical capabilities or organization's context?
5. What are the evidence standards that a cloud comparison needs to meet so that the comparison is considered balanced? Much of the available material is from vendors.

II. Research Gap, Scope and Contributions

A. Research Gap

Cloud computing is an active area of research. Individual topics such as security, edge computing, serverless computing etc. are well researched [3], [4], [11]-[14]. The research gaps addressed in this paper are as follows:

1. *Disjointed coverage:* Most studies focus on a single technology. Security, AI, multi-cloud, edge, serverless and green computing are rarely studied under a single framework. Thus, interactions between these technologies are not captured.
2. *Lack of balance:* Industry reports often emphasize benefits. Security studies often emphasize risks.
3. *Lack of comparison between deployment approaches:* Traditional, hybrid and multi-cloud approaches are studied separately. There is limited direct comparison on security, cost, flexibility, lock-in and interoperability [9], [10].
4. *Lack of independent comparison of providers:* Much of the information on platforms is provided by vendors.
5. *Increasing impact of AI but understudied:* Generative AI workloads are changing the architecture, cost and energy consumption of cloud. This is changing faster than literature can keep up [7], [8].
6. *Energy and sustainability separated from performance:* Energy and carbon are often studied separately from scalability and cost. However, they are related [15].

Another methodological weakness is the approach to comparison. Often, different criteria are used for different technologies. Or, options are compared without specifying how criteria are weighted. This makes it difficult to reproduce or challenge results. We address this by providing a framework, specifying weights and a documented screening procedure.

B. Research Focus

We focus on three areas: cloud security and privacy, AI and cloud computing, and hybrid and multi-cloud computing. We provide an overview on edge computing, serverless computing and green cloud computing. This is to show how these are related to focus areas. We compare deployment approaches and major providers on eight common criteria. These are security features, scalability, cost model, AI and ML services, hybrid and multi-cloud support, interoperability, vendor lock-in risk, sustainability initiatives.

C. Scope

The study is limited to secondary research. No survey, interview, experiment or benchmark was conducted. No primary data was collected. Only the documentation from providers is used to determine what services are available. We do not accept superiority claims in documentation without independent support. Since cloud products are changing rapidly, product names and features are representative of the period of writing. This paper does not certify any provider as secure. This paper does not provide legal advice on compliance.

D. Expected Contributions

- *Integrated framework:* We provide a single framework for evaluating security, AI, multi-cloud, edge, serverless and green computing.
- *Transparent approach to comparison:* We provide a weighted scoring approach. We specify the weights and provide a sensitivity check. This shows how the ranking is dependent on the priorities.
- *Documented screening:* We provide a source screening algorithm. This makes the selection of literature repeatable.

- *Separation of benefit and risk:* We provide an analysis where we present both the advantages and limitations of each technology.
- *Practical recommendations:* We provide guidelines for secure, efficient and sustainable adoption.

III. Background and Related Work

A. Cloud Architecture and Shared Responsibility Model

A cloud environment is a layered architecture. Each layer hides the complexity of the lower layers. At the physical layer, we have data centres, servers, storage and networks. Virtualisation or containers are used to slice the hardware for multiple tenants. There is a resource management layer for provisioning, scheduling, billing and monitoring. Customers interact with top layers as IaaS, PaaS or SaaS.

The service models vary in terms of the level of control that the customer has. In IaaS, the customer controls the operating system, middleware and application. In PaaS, only the application and its data. In SaaS, the provider controls almost everything. This is the basis of the shared responsibility model. In this model, the provider secures the underlying infrastructure. The customer secures the things that are on top of it [1], [4]. In all the models, the customer is responsible for the data. The customer is responsible for identities that have access to the data. The customer is responsible for proper configuration. This is why many incidents occur due to misconfiguration.

NIST defines four deployment models. These are private, community, public and hybrid [1]. In this paper, we use the term multi-cloud for using services from more than one public provider. This can be with or without private infrastructure. Hybrid cloud is used when sensitive data is kept on private systems. However, public cloud is used for elasticity. Multi-cloud is used to reduce vendor lock-in. It is used to leverage best service for each task. It is used to meet resilience and regulatory requirements [9].

B. Security and Privacy

Subashini and Kavitha showed that security risks vary depending on the service model. This is because the level of exposure of customer is higher at lower layers. The level of dependence on the security of the provider's application is higher at higher layers [4]. Ristenpart et al. showed that virtual machines of different tenants running on the same hardware can leak information through side channels. This showed that the multi-tenancy feature, which makes cloud economical, also introduces risk [5].

Configurations and identities are nowadays more in the focus of attacks than exotic attacks. The security incident at Capital One (2019) with a misconfigured web application firewall, allowing an attacker to grab credentials and read data in the cloud storage, is a good example. The platform itself worked as expected. But the setup created the exposure. The Cloud Security Alliance has identified misconfiguration, weak identity management, insecure interfaces as the top threats for cloud computing in multiple iterations.

Regulation is an important driver for privacy. General Data Protection Regulation (GDPR) in the European Union sets the rules for the processing and transfer of personal data. Digital Personal Data Protection Act, 2023 in India establishes similar obligations for entities processing personal data of individuals in India. These regulations pose questions about data residency, cross-border transfer, and the shared responsibility with the cloud provider for global cloud providers users. The zero trust model in NIST SP 800-207, which doesn't trust any request based on its network location, is suitable for cloud environments where the traditional perimeter is difficult to define [6].

C. AI and Cloud Computing

The cloud computing and AI have a bidirectional relationship. Cloud computing provides the compute, storage, and tooling to train and serve AI models. On the other hand, AI is used in the cloud to predict the load, place the workload, detect anomalies, automate operations, etc. [3].

Strubell et al. and Patterson et al. shifted the focus on the cost of the first direction. The training of large AI models can be energy intensive. The carbon footprint strongly depends on the hardware, the model design, the data centre location [7], [8].

AI for cloud management has the advantages of reduced manual effort, faster reaction. However, it is limited by the dependence on good training data, the bad decision making when the conditions are different from the training period, and the difficulty to explain the automated decisions [3].

D. Hybrid and Multi-Cloud Computing

Petcu listed the reasons for multi-cloud as avoiding lock-in, better availability, cost optimisation, regulation, etc. But there are challenges like incompatible interfaces, need for abstraction layers, etc. [9]. Opara-Martins et al. identified that lock-in can happen at multiple levels: application, platform, infrastructure, data. It is better to view it as a spectrum rather than as a yes/no property [10]. Kubernetes is the common technical layer that improves the portability, with the container orchestration. But application portability is not the same as data, identity, monitoring service portability.

E. Edge, Serverless and Green Cloud Computing

Satyanarayanan argued that edge computing is a thing because some applications need low latency, low bandwidth usage, operation even with poor connectivity, or raw data to remain local for privacy [11]. Shi et al. defined the vision, the challenges of edge computing like programmability, service management, security. They presented edge as a complement to cloud computing [12]. Jonas et al. defined serverless computing as a model where developers provide functions, and the provider takes care of the servers, scaling, etc. They predicted the growth of serverless computing but listed the limitations like stateful, long-running tasks, etc. [13]. Hellerstein et al. were more critical of the first generation of serverless platforms for data-intensive tasks [14]. On the energy front, Masanet et al. showed that the data-centre computing output increased very rapidly between 2010 and 2018, but the energy usage increased modestly due to efficiency gains, moving to large, efficient data centres. But they warned that the gains cannot be guaranteed [15].

F. Positioning in the Literature

There is a strong coverage of individual topics in the literature, but the integration is weak. Security studies rarely cover AI workloads in depth. AI studies rarely cover multi-cloud portability. Only occasionally energy studies are related

 ISSN: 3139-1141

to scalability decisions. This paper contributes by fixing a set of common criteria. We state how we weight them. We keep both benefits, limitations in the picture.

IV. Literature Corpus and Data Description

A. Identification of Sources

We use only secondary data. We selected candidate sources from IEEE Xplore, Google Scholar, ACM Digital Library. We used NIST publications and standards [1], [6]. We used industry reports like Gartner's cloud trend publications [17]. We used official documentation of AWS, Azure, Google Cloud. Table I provides a summary of the corpus.

TABLE I. Summary of the Corpus and Types of Sources

Attribute	Specification
Study type	Secondary research; systematic literature review in the spirit of Kitchenham and Charters [16]
Academic databases	IEEE Xplore; Google Scholar; ACM Digital Library
Standards, guidelines	NIST SP 800-145 [1]; NIST SP 800-207 [6]
Industry reports	Gartner cloud computing trends, strategic technology reports [17]
Vendor documentation	AWS, Microsoft Azure, Google Cloud (service information only; vendor bias noted)
Topic areas	Security and privacy; AI and cloud; hybrid and multi-cloud; edge; green; serverless
Candidate sources identified	[Count: to be filled from search log]
Sources retained after screening	[Count: to be filled after Algorithm 1]
Primary data collected	None

B. Topic Areas and Key Concepts

We assign the retained sources to one or more topic areas. They are listed in Table II. We use key concepts to code them.

TABLE II. Topic Areas and Key Concepts

No.	Area	Key concepts
1	Cloud security and privacy	Data breaches, data privacy, cyberattacks, identity and access management
2	AI and cloud computing	AI/ML integration, generative AI infrastructure, AI-powered cloud management
3	Hybrid and multi-cloud computing	AWS, Azure, Google Cloud, vendor lock-in, interoperability
4	Edge computing	Faster data processing, IoT integration, reduced latency
5	Green cloud computing	Energy consumption, carbon footprint, sustainable data centres
6	Serverless computing	Function as a Service (FaaS), automatic scaling, cost optimisation

C. Source Bias and Imbalance

There are two types of imbalance in this corpus. First, vendor documentation paints a favourable picture of their products. It is available in much larger volume than the independent evaluation. So, if we just count the sources, we would have an overrepresentation of the vendors' perspective. Second, the security-oriented studies tend to report incidents. But industry reports tend to report adoption benefits. If we weight the sources by number, then the conclusions would depend on which type of source is more numerous, not on the quality of evidence. So, we tag the vendor sources as "service information only". We give the analytic weight to independent peer-reviewed, standards body sources. We separate benefits, risks in the coding sheet columns.

D. Data Quality Assurance

- *Source verification.* We verified author, date, venue, traceable origin of each source. We discarded unverifiable material.
- *Duplicate control.* We removed the duplicates across databases based on the title, digital object identifier.
- *Currency check.* We recorded the product names, feature claims from vendor documentation with access date. It changes often.
- *Claim separation.* We classified claims into three types: definition/standard, empirical finding, vendor statement. We used only the first two types for comparative judgments.
- *Fixed criteria.* We fixed the criteria, weights before scoring the comparison. So, the scoring cannot be manipulated to get a desired result.

V. Methodology

A. Study Workflow

The workflow follows the steps of Cloud Trends and Challenges Evaluation Framework: literature collection, source screening, topic categorisation, trend study, challenge identification, comparative analysis, gap identification, future scope, recommendations. We use the same eight criteria with the same scoring scale for each deployment model, provider to keep the comparison fair. The steps are as follows.

1. *Literature collection.* Search the databases in Table I using keyword strings with the keywords of this paper. Log each query and the result counts.
2. *Source screening.* Use inclusion, exclusion criteria with Algorithm 1.
3. *Topic categorisation.* Assign each retained source to one or more of the six topic areas in Table II.
4. *Trend study.* Record the developments, forces driving them in each area.
5. *Challenge identification.* Record the risks, limitations separate from benefits. So, neither dominates.
6. *Comparative analysis.* Score the deployment models, providers using Algorithm 2.
7. *Gap identification.* Note the weak topics covered in the retained literature.
8. *Future scope.* Outline the directions based on the gaps.
9. *Recommendations.* Turn the findings into practical guidance.

B. Inclusion, Exclusion Criteria

A source is included if it is peer-reviewed or from a reputable body like a standards organisation, recognised industry analyst. It is relevant to at least one topic area. Preferably, it was published recently. We keep older works if they are foundational. For example, the NIST definition [1], the

Berkeley view of cloud computing [2]. A source is excluded if it is a duplicate, irrelevant, unverifiable.

C. Source Screening and Categorisation

Algorithm 1: Source Screening and Categorisation

Input: C = set of candidate sources; T = {Security, AI, HybridMulti, Edge, Green, Serverless}; Y_{min} = earliest preferred publication year
Output: S = retained sources with topics tagged

```
S ← empty set
for each source c in C do
    if c duplicates a source in S already then discard c; continue
    if c is not peer-reviewed AND not from a reputable body then
    discard c; continue
    if c is unverifiable (no author, date, origin) then discard c; continue
    topics ← { t in T : c addresses t in title, abstract, keywords }
    if topics is empty then discard c; continue
    if c is vendor documentation then tag c "service information only"
    if year(c) < Y_min then tag c "foundational"
    add (c, topics) to S
end for
return S
```

The retention rate of screening can be reported as

$$R = |S| / |C| \tag{1}$$

where |C| is the number of candidate sources, |S| is the number of retained sources. The procedure runs in linear time with respect to |C|, except the duplicate check which is efficient by indexing the sources based on the title, digital object identifier.

D. Comparison Criteria and Scoring

We score each option (deployment model, provider) on the eight criteria from 1 (weak) to 5 (strong). We invert the scale for risk type criteria like vendor lock-in. So, 5 is the most favourable outcome always.

$$s' = 6 - s \tag{2}$$

where s is the raw risk rating, s' is the favourability score. We have a weight w_j for each criterion j. It is set by the decision-maker. The weights sum to one. The total score for option o_i is the weighted sum

$$S(o_i) = \Sigma_j \, w_j \, s_{ij} \tag{3}$$

with $\Sigma_j \, w_j = 1$. We can check the stability of the ranking by increasing the weight of one criterion k by d. We rescale other weights to keep them sum to one.

$$w_k' = w_k + d,$$

$$w_j' = w_j \times (1 - w_k') / (1 - w_k), \tag{4}$$

for $j \neq k$. We consider the ranking fragile if it changes with a small shift.

E. Weighted Scoring with Sensitivity Check

Algorithm 2: Weighted Scoring of Cloud Options

Input: O = {o_1, ..., o_m} candidate options; K = {k_1, ..., k_n} criteria; w[j], j = 1..n weights, w[j] ≥ 0, sum(w) = 1; s[i][j] score of option i on criterion j (1..5), inverted for risk criteria using (2)
Output: ranked options with totals, sensitivity notes

```
for i = 1 to m do
    total[i] ← 0
    for j = 1 to n do
        total[i] ← total[i] + w[j] × s[i][j]   // (3)
    end for
end for
rank ← sort O by total in descending order
for each criterion k in K do
    w′ ← weights shifted by d for k using (4)
    recompute totals with w′
    if the order of options is different from rank then note k as a
    sensitive criterion
end for
return rank, totals, sensitive criteria
```

The running time is proportional to $m \times n$ for the scoring, $m \times n^2$ for the full sensitivity check. It is trivial for the small numbers we have.

F. Supporting Relations Used in the Analysis

Three standard relations are used in the analysis of the different deployment models to ground the discussion of scalability, resilience and sustainability.

Auto-scaling. A common horizontal scaling rule, used for example by the Kubernetes horizontal pod autoscaler, sets the desired number of replicas as

$$N_{desired} = \lceil N_{current} \times M_{current} / M_{target} \rceil \tag{5}$$

where M is a chosen metric such as CPU utilisation. If 4 replicas run at 80% utilisation against a 50% target, $N_{desired}$ = ceil(4 × 1.6) = 7.

Availability of redundant providers. If a service can run on either of two providers whose failures are independent, with availabilities A_1 and A_2, the combined availability is

$$A = 1 - (1 - A_1)(1 - A_2) \tag{6}$$

For two providers at 99.9% each, $A = 1 - 0.001 \times 0.001 = 99.9999\%$. This is an upper bound: shared dependencies such as a common application bug, a single DNS service or a common identity provider reduce it, which is why multi-cloud resilience must be designed and not assumed.

Data-centre efficiency. Power Usage Effectiveness (PUE) is the ratio of total facility energy to the energy used by IT equipment:

$$PUE = E_{facility} / E_{IT} \tag{7}$$

A PUE of 1.0 would mean that all energy reaches the IT equipment. PUE does not capture the carbon intensity of the energy source, so it must be read together with measures of that intensity [8], [15].

VI. Evaluation Metrics

A. Criteria and Scoring Rubric

Because no benchmark data were collected, quality is judged through a rubric that fixes what each score means. Table III gives the rubric for the eight criteria.

TABLE III. Comparison Criteria and Scoring Rubric

Criterion	Question asked	Score 1	Score 5
Security features	How much protection is available by default and by configuration?	Few controls, mostly manual	Rich controls, strong defaults, good monitoring
Scalability	How easily can capacity change with demand?	Manual, slow scaling	Automatic, fast scaling
Cost model	How predictable and efficient is cost?	Opaque or high overhead	Transparent, low overhead
AI and ML services	How broad and mature is the AI offering?	Minimal	Broad managed platform and model access
Hybrid and multi-cloud support	How well can workloads span environments?	None	Native, well supported
Interoperability	How easily do components work with others?	Proprietary only	Open standards and portable interfaces
Vendor lock-in risk (inverted)	How hard is it to leave?	Very hard	Easy
Sustainability	How visible and strong are efficiency and carbon measures?	None reported	Detailed reporting and targets

B. *Reliability of the Coding*

If two reviewers score the same sources independently, their agreement beyond chance can be reported with Cohen's kappa:

$$\kappa = (p_o - p_e) / (1 - p_e) \quad (8)$$

where p_o is the observed proportion of agreement and p_e the proportion expected by chance. Values above about 0.6 are usually read as substantial agreement. Disagreements are resolved by discussion and the resolution rule is recorded.

C. *Evidence Strength*

Each comparative judgment is also tagged with the strength of its support: strong when backed by a standard or by several independent peer-reviewed studies, moderate when backed by one study or by consistent industry commentary, and weak when it rests mainly on vendor documentation. Weak judgments are reported but not used to rank options.

VII. Study Design and Reproducibility

A. *Protocol*

- *Uniform criteria.* All options are judged against the same eight criteria and the same rubric.
- *Fixed weights before scoring.* Weights are set before any option is scored and are then changed only in the sensitivity step.
- *Separation of claim types.* Definitions, empirical findings and vendor statements are recorded separately.
- *Search log.* Every query, database, date and result count is saved so that the search can be repeated.
- *Archived scoring sheet.* Raw scores, weights and totals are kept in a spreadsheet so that Algorithm 2 can be rerun independently.

B. *Reproducibility Checklist*

TABLE IV. Reproducibility Checklist

Parameter	Specification
Study type	Secondary research; systematic literature review [16]
Databases	IEEE Xplore; Google Scholar; ACM Digital Library
Search date	[to be added]
Search strings	[to be added from the search log]
Inclusion criteria	Peer-reviewed or reputable; relevant; preferably recent; foundational works kept
Exclusion criteria	Duplicate; irrelevant; unverifiable
Screening procedure	Algorithm 1
Scoring procedure	Algorithm 2, scale 1 to 5, risk criteria inverted by (2)
Criteria and weights	Eight criteria; illustrative weights in Table VII
Reference manager	Zotero
Analysis tools	Microsoft Excel for comparison tables; Microsoft Word for documentation
Inter-rater agreement	[to be computed with (8) if two reviewers score]
Primary data	None

C. *Threats to Validity*

Selection bias arises if the search strings favour certain topics or vendors. Reviewer bias arises because scores are judgments. Currency bias arises because product features change faster than publications. These threats are reduced, but not removed, by the search log, the rubric, the evidence-strength tags and the sensitivity check.

VIII. Findings and Discussion

A. *Summary of Trends*

Table V summarises the findings for the six topic areas, giving the principal benefit and the principal limitation of each.

TABLE V. Trends, Benefits and Limitations

Trend	Principal benefits	Principal limitations	Evidence strength
AI and ML in the cloud	Automation, advanced analytics, low entry barrier	High compute demand, data privacy, explainability	Moderate
Hybrid and multi-cloud	Flexibility, resilience, reduced single-vendor dependence	Complexity, skill needs, data-transfer cost	Moderate
Edge computing	Low latency, reduced bandwidth, local privacy	Device management, larger attack surface	Strong
Serverless computing	No server management, automatic scaling, pay per use	Cold starts, state handling, lock-in	Strong
Green cloud computing	Lower energy and carbon, cost savings	Rising AI demand, measurement difficulty	Moderate
Cloud security and privacy	Strong provider controls, shared responsibility	Misconfiguration, identity weakness, compliance	Strong

B. Security and Privacy Findings

The literature supports a consistent message: the most common routes to exposure are customer-side configuration errors, over-privileged identities and insecure interfaces, and not a failure of the provider's infrastructure [4], [6]. Multi-tenant isolation remains a research concern [5], and compliance duties under laws such as the GDPR and India's Digital Personal Data Protection Act, 2023 add a regulatory dimension. Table VI pairs the main challenges with typical causes and mitigations.

TABLE VI. PRINCIPAL CHALLENGES AND MITIGATION APPROACHES

Challenge	Typical cause	Common mitigation
Data breach	Misconfiguration, stolen credentials	Configuration monitoring, encryption, least privilege
Weak access control	Over-privileged roles, no MFA	Strong IAM, MFA, regular access review
Compliance risk	Cross-border data, sub-processors	Data classification, regional controls, contracts
Vendor lock-in	Proprietary services, data gravity	Open standards, containers, portable data formats
Multi-cloud complexity	Different tools and policies	Common tooling, infrastructure as code, central governance
Cost overruns	Idle or oversized resources	Monitoring, right-sizing, budgets and alerts
AI privacy and cost	Sensitive data, heavy compute	Data governance, model selection, usage limits
Energy use	Large data centres, AI workloads	Efficient hardware, renewable energy, workload scheduling

C. AI and Cloud Findings

Managed ML platforms and foundation-model services lower the barrier to using AI, with representative offerings from all three providers: Amazon SageMaker and Bedrock, Azure Machine Learning and Azure OpenAI Service, and Google's Vertex AI. The same literature shows that training and serving large models are energy intensive and that the footprint depends strongly on hardware, model design and location [7], [8]. The practical finding is that AI adoption should be governed like any other data-processing service, with attention to what data is sent, where it is processed, what the outputs are checked against and what the usage costs.

D. Deployment Model Comparison

TABLE VII. ILLUSTRATIVE WEIGHTS AND SCORES (5 = MOST FAVOURABLE)

Criterion	Weight	Single public cloud	Hybrid cloud	Multi-cloud
Security control	0.25	3	4	3
Cost	0.20	5	3	3
Flexibility	0.15	3	4	5
Lock-in risk (inverted)	0.15	2	3	5
Interoperability	0.10	4	3	2
Operational simplicity	0.15	5	3	2
Weighted total (3)	**1.00**	**3.65**	**3.40**	**3.35**

The weights and scores in Table VII are assumptions made for illustration for a hypothetical mid-sized organization handling some regulated data, derived from the qualitative assessments in the literature. They are not measurements. For the single public cloud, 0.25(3) + 0.20(5) + 0.15(3) + 0.15(2) + 0.10(4) + 0.15(5) = 3.65. For hybrid cloud, 0.25(4) + 0.20(3) + 0.15(4) + 0.15(3) + 0.10(3) + 0.15(3) = 3.40. For multi-cloud, 0.25(3) + 0.20(3) + 0.15(5) + 0.15(5) + 0.10(2) + 0.15(2) = 3.35.

The sensitivity step of Algorithm 2 shows why the ranking is not a verdict. If the lock-in weight is raised from 0.15 to 0.30 and the others are rescaled by (4), the totals become approximately 3.64 for multi-cloud, 3.36 for the single public cloud and 3.33 for hybrid cloud, so multi-cloud moves to first place. If the security weight is raised from 0.25 to 0.35, the single public cloud (about 3.56) still leads hybrid cloud (about 3.48), but the gap narrows.

Lock-in is therefore the sensitive criterion in this example. The general lesson is that deployment choice follows from an organization's priorities, and a method that records the weights makes that dependence visible.

The redundancy relation of (6) supports the resilience argument for multi-cloud but only under independence. An organization that runs two providers behind one DNS service or one identity provider has not obtained the full benefit.

E. Provider Comparison

TABLE VIII. REPRESENTATIVE SERVICES ACROSS PROVIDERS

Criterion	AWS	Microsoft Azure	Google Cloud
Identity and security	IAM, KMS, GuardDuty	Microsoft Entra ID, Key Vault, Defender for Cloud	Cloud IAM, Cloud KMS, Security Command Center
Compute and containers	EC2, ECS, EKS	Virtual Machines, AKS	Compute Engine, GKE
Serverless	Lambda	Azure Functions	Cloud Functions, Cloud Run
AI and ML platform	SageMaker, Bedrock	Azure Machine Learning, Azure OpenAI Service	Vertex AI, TPUs
Hybrid and multi-cloud	Outposts, EKS Anywhere	Azure Arc, Azure Stack	GKE Enterprise (formerly Anthos)
Data analytics	Redshift, Athena	Synapse Analytics	BigQuery
Sustainability tools	Customer carbon footprint tool	Emissions Impact Dashboard	Carbon Footprint reporting

Table VIII is based on public documentation and shows that every criterion can be met by all three providers, with differences of emphasis. AWS has the longest history and the widest catalogue. Azure integrates closely with Microsoft's enterprise software, identity and licensing. Google Cloud has a strong heritage in data analytics, container orchestration and machine learning. These are descriptions of ecosystems, not rankings, and since they rest partly on vendor documentation their evidence strength is weak to moderate. Product names and features should be verified before use.

F. Edge, Serverless and Green Computing

Edge computing complements the cloud by handling time-sensitive and local processing, at the cost of managing many distributed devices and a larger attack surface [11], [12]. Serverless computing removes server management and scales automatically, but cold starts, limits on state and execution time, and provider-specific event models limit its fit for some workloads [13], [14]. For green computing, efficiency gains were large between 2010 and 2018 [15], but AI demand makes continued gains uncertain, and measures such as PUE (7) need to be read together with the carbon intensity of the energy used. Many sustainability measures, such as right-sizing and auto-scaling (5), also reduce cost, so sustainability and cost optimisation often point the same way.

G. Discussion

Three conclusions follow. First, the choice among deployment models is a trade-off between control and flexibility on one side and simplicity and cost on the other, and the ranking depends on weights. Second, AWS, Azure and Google Cloud are comparable in breadth, so the choice among them is driven more often by existing skills, software relationships, specific services and pricing than by a decisive technical gap. Third, the criteria interact: reducing lock-in through multi-cloud .