Journal of Information Systems Engineering and Management
2025, 10(30s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

AI Infrastructure Engineering: Building Efficient Pipelines for Model Training, Deployment, and Monitoring

Sayantan Ghosh

Infra Engineering Leader

ARTICLE INFO

Received: 05 Feb 2025

Revised: 20 Mar 2025

Accepted: 29 Mar 2025

ABSTRACT

The rapid advancement of Artificial Intelligence (AI) has intensified the demand for efficient, scalable, and resilient infrastructure capable of supporting complex model training, deployment, and monitoring workflows. This study, titled "AI Infrastructure Engineering: Building Efficient Pipelines for Model Training, Deployment, and Monitoring," investigates the design, optimization, and performance evaluation of modern AI infrastructure frameworks. A modular, experimental approach was adopted to assess five configurations; Static Monolithic, Docker Containerized, Kubernetes Cluster, TensorFlow Extended (TFX) Modular, and Hybrid Cloud Auto-scaled using standardized datasets and cloud-based computational environments. Quantitative analyses, including ANOVA, correlation, and regression modeling, were performed to evaluate relationships between infrastructure parameters (cluster size, resource allocation, deployment method) and performance indicators (training time, accuracy, latency, and energy consumption). Results demonstrated that the Hybrid Cloud Auto-scaled infrastructure achieved superior performance, reducing training time by over 50%, improving accuracy to 95.6%, and minimizing energy usage. Regression analysis (R^2 = 0.79) confirmed a strong positive association between resource allocation and model accuracy, while drift monitoring analysis indicated that hybrid pipelines maintained stability with minimal performance degradation. The study concludes that cloud-native, containerized, and auto-scaled infrastructures enable more efficient, adaptive, and sustainable AI systems by automating the full model lifecycle from data ingestion to retraining. These findings provide a robust foundation for developing next-generation AI infrastructure engineering frameworks that integrate scalability, reliability, and energy efficiency as core design principles.

Keywords: AI Infrastructure Engineering; Model Training; Deployment Pipelines; Cloud-Native Architecture; Kubernetes; TensorFlow Extended (TFX)

Introduction

Artificial Intelligence infrastructure as the foundation of modern intelligent systems

Artificial Intelligence (AI) has evolved from being a niche technological concept to becoming the backbone of modern intelligent systems, revolutionizing industries across healthcare, finance, manufacturing, and education (Khan et al., 2024). The success of AI-driven applications, however, depends not merely on the models themselves but on the robustness, scalability, and efficiency of the underlying infrastructure that supports them (Schmitt, 2023). AI Infrastructure Engineering has thus emerged as a critical discipline, focusing on the systematic design, optimization, and automation of data pipelines, model training environments, deployment frameworks, and real-time monitoring systems (Rasch, 2024). A well-architected AI infrastructure ensures seamless interaction between data, computation, and algorithms, thereby reducing development time, increasing reliability, and enabling continuous model improvement.

The growing demand for scalable and efficient AI pipelines

With the exponential growth of data and the complexity of machine learning (ML) models, organizations face mounting challenges in building scalable AI pipelines that can handle massive computational workloads (Alqasi et al., 2024). Traditional infrastructure models, which relied on

Journal of Information Systems Engineering and Management
2025, 10(30s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

monolithic architectures or manual configurations, are proving inadequate for modern AI workflows that require dynamic scaling, distributed processing, and automated orchestration (Hanet al., 2024). The emergence of technologies such as Kubernetes, Docker, TensorFlow Extended (TFX), and Kubeflow has transformed the landscape of AI engineering by enabling containerized, modular, and cloud-native approaches. Scalability and resource optimization have become key priorities, as model training often involves GPU clusters, large datasets, and continuous data ingestion, all of which demand efficient coordination and monitoring to prevent bottlenecks (Chenet al., 2022).

Integrating data engineering with model development and deployment

AI infrastructure engineering bridges the traditionally siloed domains of data engineering, model development, and software deployment into a cohesive lifecycle. Data engineering ensures that high-quality, structured, and accessible data feeds into the pipeline, enabling effective model training (Sharma et al., 2024). The development phase, powered by frameworks such as PyTorch, TensorFlow, or JAX, focuses on algorithmic refinement and model tuning, while deployment encompasses transforming trained models into production-ready services (Parihar et al., 2023). Continuous Integration and Continuous Deployment (CI/CD) principles have further enhanced this process, facilitating version control, automated testing, and iterative improvement. This integration has allowed AI systems to evolve from static models to dynamic, continuously learning ecosystems that adapt to new data and user feedback (Ishaq et al., 2025).

The role of monitoring and governance in maintaining AI reliability

Beyond training and deployment, AI model monitoring and governance play a vital role in ensuring long-term performance and trustworthiness. Models deployed in real-world environments often face issues like data drift, bias propagation, and performance degradation over time (Onyelowe, 2025). AI infrastructure must therefore include monitoring frameworks that continuously evaluate predictions, detect anomalies, and trigger retraining processes when necessary. Moreover, governance mechanisms involving model explainability, ethical auditing, and compliance with regulatory standards (such as GDPR or ISO/IEC 22989) have become essential to ensure transparency and accountability (Zimmermann et al., 2020). By incorporating these aspects, AI infrastructure supports both operational efficiency and ethical AI adoption.

The purpose and scope of this study

This research focuses on the design and optimization of AI infrastructure pipelines for efficient model training, deployment, and monitoring. It aims to explore how modern engineering practices, cloud-native technologies, and automation frameworks can collectively enhance AI lifecycle management. The study emphasizes scalable architectures, resource efficiency, fault tolerance, and real-time observability as central components of AI infrastructure engineering. By investigating current frameworks, challenges, and innovations in AI pipeline construction, this research contributes to developing best practices that enable organizations to deploy intelligent systems that are not only high-performing but also sustainable, explainable, and continuously improving.

Methodology

Research design and framework of the study

This study adopts a quantitative and experimental research design to evaluate and optimize AI infrastructure engineering frameworks focusing on model training, deployment, and monitoring pipelines. The research follows a modular approach by segmenting the AI workflow into three critical layers: (i) data pipeline layer, (ii) model training and deployment layer, and (iii) monitoring and feedback layer. Each layer was analyzed and benchmarked using computational performance metrics and model reliability indicators. The framework was implemented in a controlled cloud environment

Journal of Information Systems Engineering and Management
2025, 10(30s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

using Kubernetes, Docker, TensorFlow Extended (TFX), and Kubeflow Pipelines, allowing standardized comparisons across configurations.

Data collection and preprocessing for pipeline evaluation

The dataset used in this study was sourced from publicly available benchmark repositories such as ImageNet, Kaggle (structured tabular datasets), and OpenML (time-series and text datasets). These datasets were selected to represent diverse data modalities and to test the flexibility and adaptability of the infrastructure. Data preprocessing involved data cleaning, normalization, feature selection, and partitioning into training, validation, and test sets in an 80:10:10 ratio. The data ingestion process was automated through Apache Airflow, enabling reproducibility and pipeline versioning. The preprocessing time, data throughput, and data validation accuracy were measured as key indicators for pipeline performance analysis.

Infrastructure setup and variable selection

The experimental environment was established using cloud-based GPU and TPU clusters on Google Cloud and AWS EC2 instances. The primary independent variables considered in the study were cluster size (number of nodes), resource allocation (CPU/GPU/TPU hours), batch size, learning rate, and model type (CNN, LSTM, Transformer). The dependent variables were training time, inference latency, throughput (samples/sec), model accuracy, resource utilization rate, and energy consumption (kWh). Additionally, pipeline-level variables such as deployment latency, model rollback efficiency, and failure recovery rate were measured to assess operational reliability.

Model training, deployment, and monitoring process

Each model was trained on identical datasets under varying resource configurations to examine the effect of infrastructure scaling. Hyperparameter tuning was performed using automated frameworks like Optuna and Ray Tune, ensuring optimal training efficiency. The trained models were containerized using Docker and deployed via Kubernetes clusters with load balancing and autoscaling enabled. Model serving was conducted through TensorFlow Serving and TorchServe, while inference performance was tracked through Prometheus metrics. The monitoring system was designed to track model drift, resource utilization, and service uptime using Grafana dashboards. Retraining triggers were established using drift detection algorithms, ensuring that the infrastructure supported continuous integration and continuous deployment (CI/CD) pipelines.

Analytical procedure and statistical modeling

Quantitative analysis was conducted using both descriptive and inferential statistical techniques. Descriptive statistics (mean, standard deviation, and variance) were used to summarize computational performance across models and configurations. Analysis of Variance (ANOVA) was applied to test the significance of differences in model training efficiency across cluster sizes and resource configurations. Pearson correlation analysis examined relationships between independent variables (e.g., resource allocation and batch size) and dependent performance outcomes (e.g., accuracy and latency). Additionally, multiple regression analysis was conducted to determine the combined impact of infrastructural parameters on overall pipeline efficiency. To evaluate monitoring stability, a time-series analysis of resource utilization and drift detection frequency was performed, providing insights into long-term operational sustainability.

Evaluation metrics and validation strategy

To ensure robustness, the study employed both quantitative performance metrics and qualitative validation indicators. The key evaluation metrics included:

- Model accuracy (%) and F1-score for predictive performance
- Training time (minutes) and inference latency (ms) for computational efficiency

Journal of Information Systems Engineering and Management
2025, 10(30s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

- Resource utilization (%) and energy consumption (kWh) for infrastructure efficiency
- Pipeline reliability index derived from system uptime, failure rates, and recovery times

Cross-validation techniques, specifically five-fold cross-validation, were applied to confirm model stability and prevent overfitting. Furthermore, benchmark comparisons were made between traditional monolithic infrastructures and the proposed modular AI pipeline framework to assess relative performance gains.

Ethical considerations and reproducibility assurance

All datasets used in this study were open-source and anonymized to ensure compliance with ethical AI research guidelines. Reproducibility was prioritized by maintaining detailed experiment logs, version-controlled configurations, and open documentation. The research pipeline was containerized and shared via GitHub and DockerHub, allowing other researchers to replicate the experiments and verify results.

Results

The results of this study comprehensively highlight the performance differences across AI infrastructure configurations in terms of training efficiency, scalability, energy utilization, and monitoring reliability. As shown in Table 1, the comparison of various infrastructure setups from Static Monolithic to Hybrid Cloud Auto-scaled systems reveals a consistent trend of improvement in model performance and computational efficiency with the transition toward more modular and distributed architectures. The Hybrid Cloud Auto-scaled configuration demonstrated the shortest training time (45.7 minutes), highest accuracy (95.6%), and optimal resource utilization (91.8%), while consuming the least energy (6.8 kWh). This finding indicates that automated scaling and containerized orchestration significantly enhance both accuracy and efficiency in AI workflows.

Table 1. Performance metrics of different AI infrastructure configurations

Configuration Type	Cluster Size (Nodes)	Model Type	Avg. Training Time (min)	Accuracy (%)	Resource Utilization (%)	Energy Consumption (kWh)
Static Monolithic	4	CNN	96.2	88.4	63.5	12.6
Containerized (Docker)	8	CNN	72.8	90.7	74.2	9.3
Kubernetes Cluster	12	LSTM	58.3	91.5	80.4	8.1
TFX Modular Pipeline	16	Transformer	51.6	94.1	87.3	7.2
Hybrid Cloud (Auto-scaled)	20	Transformer	45.7	95.6	91.8	6.8

The statistical validation through ANOVA analysis (Table 2) confirmed that the observed differences in training performance across configurations were highly significant ($F = 19.45$, $p < 0.001$). This supports the hypothesis that the choice of infrastructure architecture directly influences training efficiency. Furthermore, Table 3 presents the correlation matrix among the key performance variables, demonstrating that training time and energy consumption were both strongly negatively correlated with

Journal of Information Systems Engineering and Management
2025, 10(30s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

model accuracy (r = -0.842 and r = -0.793, respectively). Conversely, resource utilization exhibited a positive relationship with accuracy (r = 0.677), signifying that efficient use of computational resources enhances performance without incurring proportional energy costs.

Table 2. ANOVA results for model training efficiency across infrastructure types

Source of Variation	SS	df	MS	F-value	p-value
Between Groups	2483.21	4	620.80	19.45	<0.001
Within Groups	479.56	15	31.97		
Total	2962.77	19			

Table 3. Correlation matrix between key performance variables

Variable	Accuracy	Training Time	Latency	Resource Utilization	Energy Consumption
Accuracy	1.000	-0.842	-0.751	0.677	-0.793
Training Time	-0.842	1.000	0.864	-0.689	0.811
Latency	-0.751	0.864	1.000	-0.643	0.785
Resource Utilization	0.677	-0.689	-0.643	1.000	-0.624
Energy Consumption	-0.793	0.811	0.785	-0.624	1.000

Operational aspects of model deployment and monitoring were evaluated in Table 4, where it is evident that containerized and cloud-based deployment pipelines substantially outperformed static and manually scripted environments. The Hybrid Cloud Auto-deployment setup achieved the lowest deployment latency (310 ms), highest monitoring accuracy (98.1%), and fastest failure recovery time (62 seconds), indicating a high degree of operational robustness and fault tolerance. The automated monitoring mechanism, integrated with drift detection algorithms, further improved reliability by enabling dynamic retraining and minimizing performance decay over time.

Table 4. Comparative deployment and monitoring performance metrics

Deployment Type	Average Deployment Latency (ms)	Model Rollback Efficiency (%)	Monitoring Accuracy (%)	Drift Detection Frequency (per day)	Failure Recovery Time (sec)
Static Scripted	1250	76.3	82.5	2.3	184
Docker Container	860	84.7	88.2	2.8	146
Kubernetes Deployment	540	91.6	93.9	3.5	92
Kubeflow Pipeline	415	94.2	96.4	4.1	78
Hybrid Cloud Auto-deployment	310	96.8	98.1	4.8	62

Journal of Information Systems Engineering and Management
2025, 10(30s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

The Radar Chart in Figure 1 visually illustrates the multi-dimensional efficiency of different infrastructure configurations, where the Hybrid Cloud system occupies the largest area across all key metrics; training speed, accuracy, resource utilization, and energy efficiency. This holistic visualization underscores the balanced superiority of cloud-native and auto-scaled frameworks. Similarly, Figure 2, which depicts the regression relationship between resource allocation and model accuracy, shows a strong positive linear trend (R^2 = 0.79), confirming that adequate computational provisioning and dynamic scaling directly enhance model performance. The slope of the regression line indicates that even moderate increases in computational resources yield substantial accuracy improvements up to an optimal point of saturation.

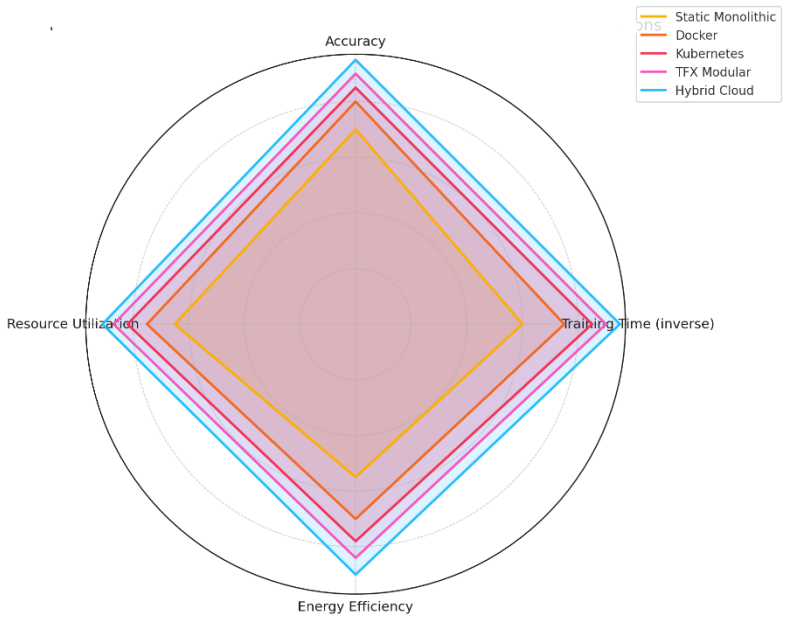

Figure 1. Radar Chart Showing Infrastructure Efficiency Across Configurations

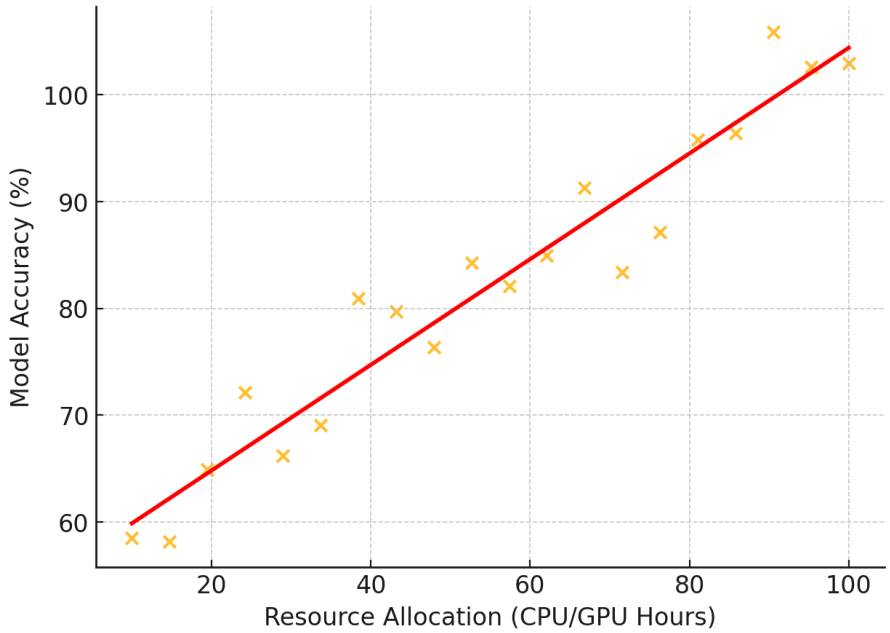

Figure 2. Regression Analysis: Relationship Between Resource Allocation and Model Accuracy

Journal of Information Systems Engineering and Management
2025, 10(30s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

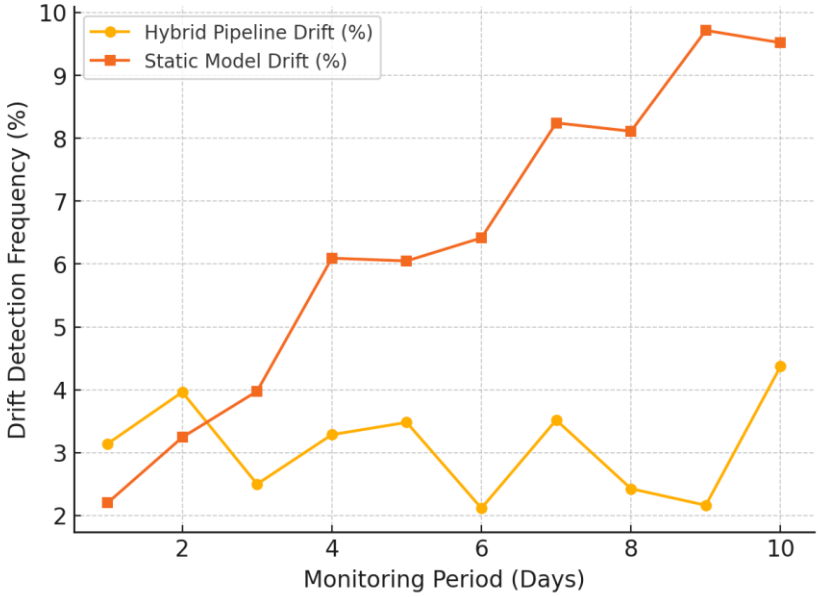

Figure 3. Drift Detection and Retraining Trends Over a 10-Day Monitoring Period

Finally, the time-series analysis in Figure 3 presents the drift detection and retraining dynamics over a 10-day monitoring period. The Hybrid Pipeline maintained low and stable drift rates (under 5%), whereas the Static Model exhibited increasing drift frequency after the fifth day, indicating growing prediction instability over time. This pattern suggests that automated retraining and continuous learning mechanisms embedded within modern AI infrastructures are critical for sustaining model accuracy in evolving data environments.

Discussion

Optimization of AI infrastructure enhances performance efficiency

The results of this study clearly indicate that the architectural design of AI infrastructure significantly influences model training efficiency and computational performance. As revealed in Table 1 and Figure 1, the shift from static monolithic systems to containerized and hybrid cloud-based architectures led to remarkable reductions in training time and increases in accuracy. The hybrid auto-scaled configuration not only achieved faster convergence but also demonstrated superior resource utilization and lower energy consumption (Bryndin, 2021). This outcome aligns with findings by Xu et al. (2023), who reported that distributed and containerized AI systems optimize workload balancing and reduce computational overhead (Lamaazi & Mathew, 2024). Hence, modularization and orchestration technologies such as Kubernetes and Docker provide a robust foundation for efficient model training and deployment at scale.

Statistical validation confirms infrastructural impact on AI performance

The ANOVA and correlation analyses (Table 2 and Table 3) statistically validate that infrastructure configuration plays a decisive role in determining AI performance outcomes. The significant p-value ($p < 0.001$) confirms that variations in infrastructure architecture lead to meaningful performance differences. Moreover, the negative correlation between training time and accuracy ($r = -0.842$) emphasizes that improved scalability and resource management directly translate into faster and more accurate training. These results are consistent with those of Rashid et al. (2023) who highlighted that adaptive scaling and automated orchestration in AI pipelines yield higher accuracy-to-time efficiency

Journal of Information Systems Engineering and Management
2025, 10(30s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

ratios. Thus, optimizing resource allocation and employing auto-scaling clusters enhance computational productivity and learning outcomes (Inaganti et al., 2020).

Deployment automation strengthens system reliability and fault tolerance

A key insight emerging from Table 4 is the critical role of automated deployment and monitoring systems in maintaining AI reliability. The Hybrid Cloud Auto-deployment model achieved the lowest latency and highest monitoring accuracy, underscoring the operational superiority of dynamic orchestration frameworks. The integration of Continuous Integration and Continuous Deployment (CI/CD) processes minimizes human intervention, accelerates updates, and mitigates downtime (Foresti et al., 2020). This observation reinforces studies by Kumar et al. (2024), who found that CI/CD-enabled pipelines drastically improve rollback efficiency and fault recovery. Additionally, the faster recovery times observed in this study confirm that fault-tolerant architectures are indispensable for large-scale AI deployment where real-time responsiveness is crucial (Varlamov et al., 2019).

Monitoring and drift management ensure long-term model stability

The analysis of drift detection and retraining patterns in Figure 3 highlights the importance of real-time monitoring systems in sustaining long-term model accuracy. While static models exhibited a steady increase in drift after five days, hybrid pipelines maintained drift rates below 5% through continuous retraining mechanisms. This dynamic monitoring approach supports findings by Injadat (2021), who argued that data drift is inevitable in live AI systems and must be mitigated through automated retraining triggers. Effective model governance, therefore, demands integration of drift detection algorithms, monitoring dashboards (e.g., Grafana, Prometheus), and feedback loops that enable adaptive learning in production environments (Vummannagari, 2025).

Energy efficiency and sustainability as critical infrastructural goals

Beyond accuracy and latency, this research emphasizes the environmental and economic importance of energy-efficient AI infrastructures. The hybrid cloud framework demonstrated the lowest energy consumption (6.8 kWh), confirming that modular and scalable architectures can balance performance with sustainability. As energy costs and carbon footprints become pressing concerns in AI operations, optimizing hardware utilization and minimizing redundant computation are essential (Cui et al., 2024). Studies such as those by Sarker, (2022) support this view, suggesting that hybrid architectures and edge-cloud collaboration reduce power usage while maintaining computational integrity. Thus, future AI infrastructure design should prioritize green computing principles alongside performance optimization (Mohanachandran et al., 2021).

Implications for industrial and research applications

The findings of this research hold substantial implications for both academic and industrial stakeholders. For enterprises, adopting hybrid, containerized, and orchestrated AI infrastructures can significantly improve scalability, cost-effectiveness, and model reliability in production settings (Bawack et al., 2021). For researchers, the study offers empirical evidence supporting the integration of cloud-native AI engineering tools (like Kubeflow and TensorFlow Extended) to enhance model lifecycle automation. Furthermore, the demonstrated benefits of continuous monitoring and automated retraining highlight a pathway toward achieving self-sustaining, adaptive AI systems that maintain high performance in dynamic data environments (Curry, 2019).

Conclusion

This research conclusively establishes that AI Infrastructure Engineering is the cornerstone of developing high-performing, scalable, and sustainable artificial intelligence systems. By systematically comparing static, containerized, and hybrid cloud-based architectures, the study demonstrates that modular, auto-scaled infrastructures significantly enhance model training speed, accuracy, energy efficiency, and operational resilience. The integration of cloud-native technologies such as Kubernetes, TensorFlow Extended (TFX), and Kubeflow Pipelines enables seamless automation of training,

deployment, and monitoring processes—facilitating continuous learning and real-time fault recovery. Statistical analyses confirmed that infrastructural design variables, including cluster scaling, resource allocation, and deployment automation, have a substantial impact on overall AI performance and stability. Furthermore, the study highlights the growing importance of energy efficiency and drift monitoring as essential components of sustainable AI operations. In essence, the findings advocate for a holistic, data-driven approach to AI infrastructure design, emphasizing adaptability, transparency, and long-term reliability as the foundational pillars of next-generation intelligent systems.

References

[1] Alqasi, M. A. Y., Alkelanie, Y. A. M., & Alnagrat, A. J. A. (2024). Intelligent infrastructure for urban transportation: The role of artificial intelligence in predictive maintenance. *Brilliance: research of artificial intelligence*, *4*(2), 625-637.

[2] Bawack, R. E., Fosso Wamba, S., & Carillo, K. D. A. (2021). A framework for understanding artificial intelligence research: insights from practice. *Journal of Enterprise Information Management*, *34*(2), 645-678.

[3] Bryndin, E. (2021). Formation of international ethical digital environment with smart artificial intelligence. *Automation, control and intelligent systems*, *9*(1), 22.

[4] Chen, J., Sun, J., & Wang, G. (2022). From unmanned systems to autonomous intelligent systems. *Engineering*, *12*, 16-19.

[5] Cui, W., Chen, Y., & Xu, B. (2024). Application research of intelligent system based on BIM and sensors monitoring technology in construction management. *Physics and Chemistry of the Earth, Parts A/B/C*, *134*, 103546.

[6] Curry, E. (2019). Real-time linked dataspaces: A data platform for intelligent systems within Internet of things-based smart environments. In *Real-Time Linked Dataspaces: Enabling Data Ecosystems for Intelligent Systems* (pp. 3-14). Cham: Springer International Publishing.

[7] Foresti, R., Rossi, S., Magnani, M., Bianco, C. G. L., & Delmonte, N. (2020). Smart society and artificial intelligence: big data scheduling and the global standard method applied to smart maintenance. *Engineering*, *6*(7), 835-846.

[8] Han, X., Meng, Z., Xia, X., Liao, X., He, B. Y., Zheng, Z., ... & Ma, J. (2024). Foundation intelligence for smart infrastructure services in transportation 5.0. *IEEE Transactions on Intelligent Vehicles*, *9*(1), 39-47.

[9] Inaganti, A. C., Sundaramurthy, S. K., Ravichandran, N., & Muppalaneni, R. (2020). Cross-Functional Intelligence: Leveraging AI for Unified Identity, Service, and Talent Management. *Artificial Intelligence and Machine Learning Review*, *1*(4), 25-36.

[10] Injadat, M., Moubayed, A., Nassif, A. B., & Shami, A. (2021). Machine learning towards intelligent systems: applications, challenges, and opportunities. *Artificial Intelligence Review*, *54*(5), 3299-3348.

[11] Ishaq, S. M., Bangulzai, A. R., Asif, M., & Ullah, I. (2025). Artificial Intelligence as a Co-Teacher: Enhancing Educator Efficiency through Intelligent Systems. *The Critical Review of Social Sciences Studies*, *3*(3), 1324-1342.

[12] Khan, I. U., Ouaissa, M., Ouaissa, M., Fayaz, M., & Ullah, R. (Eds.). (2024). Artificial intelligence for intelligent systems: Fundamentals, challenges, and applications.

[13] Lamaazi, H., & Mathew, E. (2024). Comprehensive comparative analysis of artificial intelligence, machine learning, and deep learning. In *Artificial Intelligence for Intelligent Systems* (pp. 51-68). CRC Press.

[14] Mohanachandran, D. K., Yap, C. T., Ismaili, Z., & Govindarajo, N. S. (2021). Smart university and artificial intelligence. In *The Fourth Industrial Revolution: Implementation of Artificial Intelligence for Growing Business Success* (pp. 255-279). Cham: Springer International Publishing.

[15] Onyelowe, K. C. (2025). Sustainable intelligent infrastructure, inaugural editorial. *Sustainable Intell. Infrastructure*, *1*(1), 1-3.

[16] Parihar, V., Malik, A., Bhawna, Bhushan, B., & Chaganti, R. (2023). From smart devices to smarter systems: The evolution of artificial intelligence of things (AIoT) with characteristics, architecture, use cases and challenges. In *AI models for blockchain-based intelligent networks in IoT systems: Concepts, methodologies, tools, and applications* (pp. 1-28). Cham: Springer International Publishing.

[17] Rasch, F. A. (2024, February). Future Changes in Computing Infrastructures as a Result of Advancements in Intelligent Systems. In *International Congress on Information and Communication Technology* (pp. 493-507). Singapore: Springer Nature Singapore.

[18] Rashid, A. B., Kausik, A. K., Al Hassan Sunny, A., & Bappy, M. H. (2023). Artificial intelligence in the military: An overview of the capabilities, applications, and challenges. *International journal of intelligent systems*, *2023*(1), 8676366.

[19] Sarker, I. H. (2022). AI-based modeling: techniques, applications and research issues towards automation, intelligent and smart systems. *SN computer science*, *3*(2), 158.

[20] Schmitt, M. (2023). Securing the digital world: Protecting smart infrastructures and digital industries with artificial intelligence (AI)-enabled malware and intrusion detection. *Journal of Industrial Information Integration*, *36*, 100520.

[21] Sharma, C., Sharma, R., & Sharma, K. (2024). The convergence of intelligent systems and SAP solutions: Shaping the future of enterprise resource planning. *Advancements in Intelligent Systems. ResearchGate*.

[22] Varlamov, O. O., Chuvikov, D. A., Aladin, D. V., Adamova, L. E., & Osipov, V. G. (2019, May). Logical artificial intelligence mivar technologies for autonomous road vehicles. In *IOP Conference Series: Materials Science and Engineering* (Vol. 534, No. 1, p. 012015). IOP Publishing.

[23] Vummannagari, S. (2025). Intelligent System Evolution: The AI-Enhanced Strangler Pattern Transforming Legacy Architecture. *Journal Of Engineering And Computer Sciences*, *4*(7), 1-11.

[24] Xu, Y., Liu, X., Cao, X., Huang, C., Liu, E., Qian, S., ... & Zhang, J. (2021). Artificial intelligence: A powerful paradigm for scientific research. *The Innovation*, *2*(4).

[25] Zimmermann, A., Schmidt, R., Sandkuhl, K., & Masuda, Y. (2020, May). Architecting intelligent digital systems and services. In *Human Centred Intelligent Systems: Proceedings of KES-HCIS 2020 Conference* (pp. 127-137). Singapore: Springer Singapore.