An AI-Driven Approach to Efficient Sensing and Autonomous Decision-Making with Adaptive Learning in Large-Scale Environments

Friday Oodee Philip-Kpae [1,*], Agbotiname Lucky Imoize [2], Lloyd Endurance Ogbondamati [3],
Kennedy Chinedu Okafor [4,5,6,7,8], Kelvin Anoh [5]
[1, 3] Department of Electrical and Electronic Engineering, Faculty of Engineering,
Rivers State University, Port Harcourt, Nigeria
[2] Department of Electrical and Electronics Engineering, Faculty of Engineering, University of Lagos, Akoka, Lagos 100213, Nigeria
[4] Department of Electrical and Electronic Engineering Science, University of Johannesburg, Johannesburg, South Africa
[5] School of Engineering, University of Chichester, Bognor Regis, PO21 1HR, U.K
[6] Imperial College Business School – South Kensington Campus, Imperial College London, Exhibition Road, London SW7 2AZ, UK
[7] Federal University of Allied Health Sciences, Enugu – Dental Avenue, Trans-Ekulu, Enugu State, Nigeria
[8] Department of Engineering (Faculty of Science and Engineering), Manchester Metropolitan University – Dalton Building, Chester Street, Manchester M1 5GD, UK
philip-kpae.fo@ust.edu.ng; lloydhouston27@gmail.com; aimoize@unilag.edu.ng; kennedy.okafor@ieee.org; k.anoh@chi.ac.uk
*Correspondence: philip-kpae.fo@ust.edu.ng

Abstract— **Efficient sensing and autonomous decision-making in large-scale environments face challenges such as sensor noise, limited coverage, and dynamic environmental changes, often leading to suboptimal responses. This study proposes an AI-driven framework that integrates multi-sensor data fusion, adaptive learning, and utility-based decision-making to address these issues. The framework combines heterogeneous sensor data using weighted fusion, improving the accuracy and reliability of environmental state estimation. Adaptive learning mechanisms dynamically adjust the learning rate, optimizing system performance by reducing prediction errors and refining model parameters over time. The autonomous decision-making module selects optimal actions based on utility functions, ensuring timely and accurate decisions without human intervention. The system demonstrates significant improvements in sensor coverage efficiency and overall performance, effectively handling complex, data-intensive environments. This work highlights the framework's robustness and its capacity to deliver optimal decision-making and sensing capabilities, validating its applicability in real-world, dynamic scenarios. However, the study is based on simulation, and the results may not fully reflect real-world complexities or limitations. Future work should evaluate the framework's performance in practical, large-scale deployments and address potential scalability and real-time implementation challenges.**

Index Terms—*Autonomous Decision-Making, Adaptive Learning, Multi-Sensor Data Fusion, Efficiency Optimization, Real-Time Systems, AI-Driven Framework.*

I. INTRODUCTION

In today's rapidly evolving world, the sheer scale and complexity of modern environments—ranging from smart cities to expansive industrial settings—pose significant challenges for effective sensing and timely decision-making. Traditional approaches, often rigid and centralized, struggle to keep pace with the dynamic nature of large-scale systems, resulting in inefficiencies, delayed responses, and suboptimal outcomes. Artificial Intelligence (AI), with its ability to learn, adapt, and make autonomous decisions, offers a promising solution to this challenge [1]. This research explores an AI-driven approach that leverages adaptive learning to enhance both sensing capabilities and decision-making processes in large-scale environments. By continuously learning from incoming data, the system can dynamically adjust its strategies, identify critical patterns, and make real-time autonomous decisions without requiring constant human intervention. Such adaptability is particularly crucial in scenarios where conditions change rapidly or unpredictably, such as traffic management, environmental monitoring, or industrial automation [2]. Beyond mere automation, this approach emphasizes efficiency—ensuring that sensing is precise and resource-conscious, while decisions are optimal and timely. By integrating AI with adaptive learning techniques, the framework not only reacts to present conditions but also anticipates future trends, enabling proactive management of complex systems. The result is a resilient, intelligent system capable of handling vast amounts of data, learning from it, and making decisions that improve overall performance and sustainability [3]. In essence, this study seeks to bridge the gap between large-scale environmental complexity and intelligent autonomous decision-making, demonstrating how adaptive AI can transform the way we sense, interpret, and act in dynamic real-world systems. The key research contributions of this study are:

1. Introduction of a novel AI framework that integrates multi-sensor data fusion, adaptive learning, and utility-based decision-making for improved autonomous sensing and decision-making.
2. Implementation of adaptive learning mechanisms that dynamically adjust learning rates and model parameters to optimize performance and minimize prediction errors.
3. Application of utility-based decision-making to enable real-time, autonomous selection of optimal actions.

The remainder of this paper is organized as follows: Section II reviews related work on multi-sensor data fusion and adaptive learning. Section III describes the research methodology, AI-driven framework, and outlines the simulation setup and performance evaluation. Section IV details the results, their visualization, and discussion, focusing on the presentation of system performance and efficiency. Section V concludes the paper, presents key limitations, and suggests directions for future work.

II. RELATED WORK

AI-driven sensing and autonomous decision-making have gained significant attention as modern environments demand more accurate, adaptive, and energy-efficient monitoring systems. Multi-sensor data fusion has been widely explored to improve environmental perception by combining redundant and heterogeneous sensor outputs. Classical fusion models, such as weighted averaging and Kalman filtering, have demonstrated strong capability in reducing noise and uncertainty in distributed sensor networks, thereby enabling more robust environmental estimation [1], [2]. Recent work shows that assigning reliability-based weights to sensors significantly improves fusion accuracy, particularly in large-scale IoT deployments, where individual sensors often exhibit drift and noise [3]. Adaptive learning has also emerged as a key mechanism for improving model responsiveness to dynamic environments. Studies on online learning algorithms demonstrate that variable learning rates accelerate convergence during periods of high error while preventing instability once the system approaches optimality [4]. This methodology has been successfully applied in predictive maintenance, environmental modeling, and autonomous robotics due to its capacity for continuous self-correction [5], [6], [7]. Furthermore, decision-theoretic frameworks have been implemented to enable autonomous systems to select optimal actions based on probabilistic utility functions [8]. Reinforcement learning models—particularly those based on Markov decision processes—have demonstrated strong performance in balancing competing objectives, such as accuracy, energy efficiency, and response time, in real-time decision-making scenarios [9], [10].

However, the existing literature strongly supports integrating multi-sensor fusion, adaptive learning, and utility-based decision systems, thereby providing a reliable foundation for intelligent sensing architectures capable of operating autonomously in complex and dynamic environments.

III. Methodology

The resources used in this study include multisensor data for environmental monitoring, adaptive learning algorithms to adjust system parameters, and utility-based decision-making models for optimal action selection. Simulation tools were used to evaluate system performance, with metrics including sensor coverage efficiency and prediction error. System models, represented by mathematical formulations, guided sensor data fusion, learning, and decision-making. The research design is simulation-based, focusing on evaluating an AI-driven framework for autonomous sensing and decision-making. It integrates multi-sensor data fusion, adaptive learning, and utility-based decision-making to optimize system performance in dynamic environments. The design assesses the system's ability to improve sensor coverage, reduce prediction errors, and make real-time decisions using performance metrics. The simulation setup employs multiple sensors (e.g., temperature, radar, LiDAR) to collect environmental data, which are fused according to sensor reliability. Adaptive learning algorithms adjust parameters dynamically, while autonomous decision-making selects optimal actions using utility-based models. Performance is evaluated using metrics such as sensor coverage efficiency, prediction error, and decision utility over multiple time steps. This setup tests the system's efficiency and adaptability in large-scale environments.

System model

Let N be the total number of sensor nodes in the system, where each node iii represents a distinct data source (e.g., temperature, radar, LiDAR). These nodes are connected in a network, allowing data exchange between neighboring nodes to form a unified environmental estimate. Let x_i denote the sensor data from the node i at time t, and w_i represent the weight assigned to each sensor based on its reliability. The fusion of data from these nodes is expressed as $\hat{x}(t) = \sum_{i=1}^{N} w_i x_i(t)$ where, $\hat{x}(t)$: fused environmental state at time t, $x_i(t)$: reading from the sensor i at time t, $\mathrm{w_i}$: weight of sensor i based on reliability, N: total number of sensors. For the sensor coverage efficiency, let $\mathrm{A_{total}}$ be the total area of the environment being monitored, and let $\mathrm{A_i}$ represent the effective sensing area covered by the sensor node i. Let N be the total number of sensor nodes in the system, with each sensor having a corresponding coverage efficiency η_i. The sensor coverage efficiency, $\mathrm{E_c}$, is defined as the ratio of the total effective sensing area to the total monitored area, expressed as $\mathrm{E_c} = \frac{\sum_{i=1}^{N} \mathrm{A_i}}{\mathrm{A_{total}}}$, where, $\mathrm{E_c}$: sensor coverage efficiency, $\mathrm{A_i}$: effective sensing area of the sensor t, $\mathrm{A_{total}}$: total area of the environment.

For the sensor noise reduction aspect, let $x_i(t)$ be the raw measurement from the sensor i at time t, and $\hat{x}(t)$ be the filtered (noise-reduced) measurement. Let N be the total number of sensors in the system, and σ_i^2 represent the estimated noise level for the sensor i. Sensor noise is reduced by applying a weighted averaging or filtering technique to the raw sensor measurements. Sensor noise reduction is the reduction of noise in sensor measurements through filtering. Weighted averaging or Kalman filter approaches reduce uncertainty, improving the quality of the sensed data. The filtered measurement $\hat{x}(t)$ is given by $\hat{x}(t) = \frac{x_i(t)}{1+\sigma_i^2}$, where, $\hat{x}(t)$: filtered sensor measurement, $x_i(t)$: raw sensor measurement, and σ_i^2 : estimated sensor noise level for the sensor i. For the adaptive learning rate update, let $\eta(t)$ be the learning rate at iteration t, and $\epsilon(t)$ be the prediction error at iteration t. The learning rate is dynamically adjusted based on the error, aiming to accelerate convergence during periods of high error and maintain stability as the system approaches optimality. The adjustment of the learning rate is given by $\eta(t) = \eta(t-1) \cdot exp(-\alpha \cdot \epsilon(t))$ where, η(t) : learning rate at iteration t, α: adaptation sensitivity coefficient, and $\epsilon(t)$: prediction error at iteration t.

For the environment model update, let $\theta(t)$ represent the model parameters at time t, and $\eta(t)$ be the learning rate at time t. The system continuously updates its internal model of the environment using new sensor data. A loss function guides the environment model update process $L(\theta, t)$, which measures the difference between the model's predictions and the observed data. The environmental model update represents how the system updates its internal model of the environment using new data. Adaptive learning ensures the AI continuously refines its predictions to reflect dynamic changes. The model parameters are updated by adjusting them in the direction that minimizes the loss function, as given as follows: $\theta(t) = \theta(t-1) - \eta(t) \cdot \nabla_\theta L(\theta, t)$, where, $\theta(t)$: model parameters at time t, $\eta(t)$: learning rate at time t, $L(\theta, t)$: loss function comparing model output to observed data, and $\nabla_\theta L(\theta, t)$: gradient of the loss function with respect to the model parameters. This update ensures that the model continuously refines its predictions to reflect dynamic environmental changes. By using adaptive learning, the system can adjust its model parameters to maintain accurate environmental representations over time.

The AI-driven system features an adaptive learning and ai processing module that receives data from large-scale environmental sensors (e.g., LiDAR, radar). This module handles data acquisition and fusion, uses adaptive learning for pattern recognition, and enables autonomous decision-making and control. The processing then leads to real-time actions via actuators. A crucial performance-monitoring and optimization loop ensures continuous improvement by feeding back information to adjust data processing and adaptive learning, thereby making the system efficient and responsive in dynamic, large-scale settings, as shown in Figure 1.

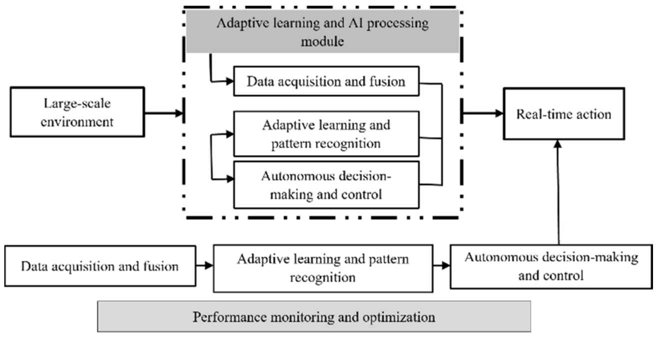

Figure 1: Adaptive Learning and AI Processing Module

For the error feedback for adaptive learning, let $e(t)$ represent the error at the time t, defined as the difference between the observed environmental state $x_{actual}(t)$ and the predicted state $x_{predicted}(t)$

from the AI system. The error feedback mechanism quantifies this difference and provides feedback to guide the adaptive learning process. Quantify the difference between predicted and actual states and provide feedback for adaptation. This error guides the learning mechanism to adjust strategies dynamically in real-time environments. The error at the time t is given as follows: $e(t) = x_{actual}(t) - x_{predicted}(t)$, where, $e(t)$: error at time t, $x_{actual}(t)$: observed environmental state, and $x_{predicted}(t)$: AI-predicted state. For the decision utility function, Let $U(a_i)$ represent the utility of taking action a_i at time t, where a_i is one of the possible actions the system can take. The utility function evaluates the expected benefit (payoff) of each action across various environmental states. The decision utility function represents a utility-based decision criterion. The AI selects actions that maximize expected utility, balancing multiple objectives such as efficiency, risk, and resource consumption, enabling autonomous decision-making, as $U(a_i) = \sum_{s\in S} P(s \mid a_i) \cdot R(s, a_i)$, where, $U(a_i)$: utility of action a_i, S: a set of possible environmental states, $P(s \mid a_i)$: probability of state sgiven action a_i, and $R(s, a_i)$: reward or payoff for taking action in state s. For the optimal action selection, let a^* represent the optimal action selected by the system and let A be the set of all possible actions the system can take. The goal is to choose the action that maximizes the expected utility $U(a_i)$, as defined in the decision utility function. Optimal action selection selects the action that maximizes expected utility. This formulation enables the AI system to autonomously choose the action that maximizes expected benefit, given the utility function, thereby ensuring optimal performance in dynamic environments. The optimal action is chosen by evaluating all available options and selecting the one that maximizes the expected utility. This is expressed as $a^* = arg \max_{a_i \in A} U(a_i)$, where, a^*: optimal action, A: a set of all possible actions, and $U(a_i)$: utility of action a_i.

For the real-time decision update scenario, let $a(t)$ represent the current action at the time t, and let Δa be the update step for the action. The real-time decision update adjusts the action in response to environmental changes, ensuring the system remains responsive. This update ensures that the system adapts its decisions over time to evolving environmental conditions, continuously refining its actions to maximize utility and efficiency. Continuously updates decisions based on changing environmental data. This ensures that the autonomous system remains responsive and adapts to unforeseen changes or disturbances. The decision update is given as: $a(t) = a(t-1) + \Delta a$, where, Δa is the change in the action, calculated as: $\Delta a = \gamma \cdot \frac{\partial U(a)}{\partial a}$. Hence, the real-time update is expressed as: $a(t) = a(t-1) + \gamma \cdot \frac{\partial U(a)}{\partial a}$, where, $a(t)$: current action at time t, γ: decision update step size, and $\frac{\partial U(a)}{\partial a}$: gradient of utility with respect to action. Now, considering the system efficiency metric, which measures the overall system efficiency by comparing achieved performance to resource utilization. High efficiency indicates the AI-driven system maximizes sensing and decision-making effectiveness while minimizing computational and energy costs. Let η_s represent system efficiency, which measures the overall performance relative to resource utilization. The system efficiency is calculated by comparing the achieved performance $P_{effective}$ (e.g., accuracy, response time) to the total resource consumption P_{total} (e.g., computational and energy costs). The system efficiency η_s is given by $\eta_s = \frac{P_{effective}}{P_{total}}$, where, η_s: system efficiency, $P_{effective}$: performance achieved (accuracy, response speed, etc.), and P_{total}: total system resource consumption. Table 1 presents the numerical data for the AI-driven sensing and decision-making framework [1]-[6]. Algorithm 1 defines the process for an AI-driven framework that autonomously senses and makes decisions.

Table 1: Numerical Data

Parameter (units)	**Sample Values**
Sensor 1 reading (°C)	23.5, 24.1, 23.8, 24.3, 24.0
Sensor 2 reading (°C)	23.7, 24.0, 23.9, 24.2, 24.1
Sensor 3 reading (°C)	23.6, 24.2, 24.0, 24.4, 24.1
Sensor reliability (Dimensionless)	0.9, 0.85, 0.88
Coverage area (m²)	100, 120, 95
Total area (m²)	500
Learning rate (Dimensionless)	0.01, 0.012, 0.011, 0.013
Error (Dimensionless)	0.05, 0.03, 0.04, 0.02
Model parameters (Dimensionless)	0.5, 0.52, 0.48, 0.51
Action set (Dimensionless)	1, 2, 3
Expected utility (Dimensionless)	0.85, 0.92, 0.78
Optimal action (Dimensionless)	2
System efficiency (Dimensionless)	0.82, 0.85, 0.88

Algorithm 1: Autonomous Sensing and Decision-Making Framework

Initialize parameters (learning rate, sensor weights, etc.)
Collect data from all sensors
for each sensor $i \in N$ do
    Fuse sensor data: $\hat{x}(t) = \sum_{i=1}^{N} w_i x_i(t)$
    Apply noise reduction: $\hat{x}(t) = \frac{x_i(t)}{1+\sigma_i^2}$
end for
Update learning rate based on error: $\eta(t) = \eta(t-1) \cdot exp(-\alpha \cdot \epsilon(t))$
Update model parameters: $\theta(t) = \theta(t-1) - \eta(t) \cdot \nabla_\theta L(\theta, t)$
for each action $a_i \in A$ *do*
    Calculate utility: $U(a_i) = \sum_{s\in S} P(s \mid a_i) \cdot R(s, a_i)$
end for
Select optimal action: $a^* = arg \max_{a_i \in A} U(a_i)$
Execute action a^* in the environment
Update decision: $a(t) = a(t-1) + \gamma \cdot \frac{\partial U(a)}{\partial a}$
Calculate system efficiency: $\eta_s = \frac{P_{effective}}{P_{total}}$
Repeat the process for the next iteration
End

Algorithm 1 defines the process for an AI-driven framework that autonomously senses and makes decisions. The algorithm starts by collecting and fusing sensor data while reducing noise. The system then dynamically adjusts the learning rate, updates model parameters, and calculates the utility of potential actions. The optimal action is selected and executed, followed by an update of the decision. Finally, the system's efficiency is evaluated, and the process repeats, ensuring continuous adaptation and improved performance in dynamic environments.

IV. Results And Discussions

This section presents the key findings from the simulation of the AI-driven framework, focusing on sensor data fusion, adaptive learning, and decision-making performance. It analyzes how the system adapts to dynamic conditions, optimizes resource usage, and makes real-time decisions as follows:

A. Multi-Sensor Data Fusion Analysis

Figure 2 illustrates the process of combining data from three sensors (S1, S2, S3) with different reliability weights. The raw temperature measurements, ranging from a low of 23.5°C for Sensor 1 at time step 1 to a high of 24.4°C for Sensor 3 at time step 4, are shown. The reliability distributions for the sensors are 0.90 for S1, 0.85 for S2, and 0.88 for S3. The Weighted Fusion Output (Comparison Plot) demonstrates that the fused values, ranging from 23.71°C to 24.28°C, are consistently averaged and stabilized by the assigned reliability weights, reducing the impact of individual sensor variations.

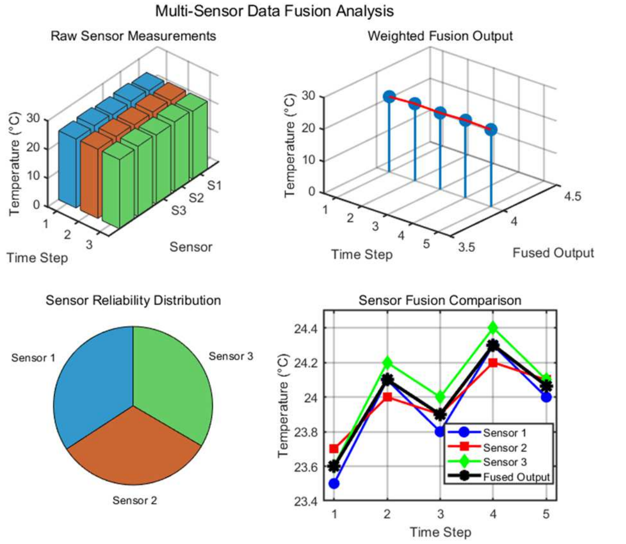

Figure 2: Multi-Sensor Data Fusion

B. Adaptive Learning Mechanism Analysis

Adaptive Learning dynamics are explored in Figure 3, focusing on the relationship between the Learning Rate (η) and the Prediction Error. The Dual Axis Plot shows that as the Prediction Error decreases from a maximum of 0.05 at Iteration 1 to a minimum of 0.02 at Iteration 4, the Learning Rate fluctuates slightly from 0.01 to a peak of 0.013 before settling, indicating that the rate adjusts in response to the error signal. The Parameter Evolution plot confirms that the model parameters (θ) are updated dynamically, moving from 0.5 to 0.51 over the four iterations, showcasing the system's capacity for optimization.

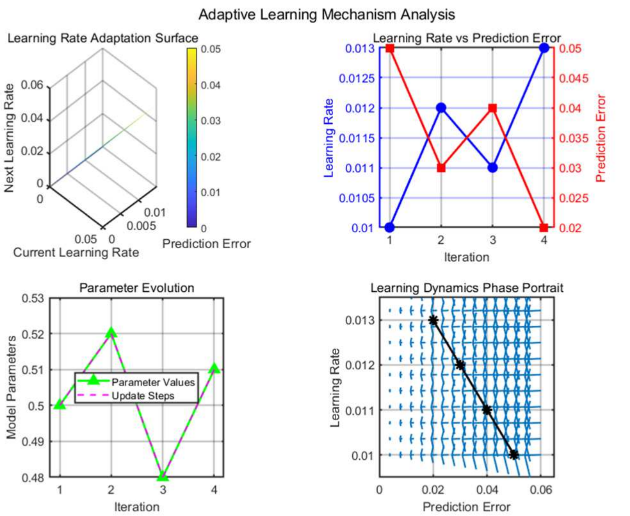

Figure 3: Learning Rate Adaptation

C. Autonomous Decision-Making Analysis

Figure 4 analyzes the utility of different actions in the autonomous decision-making process. The Action Utility Distribution (Bar Chart) clearly identifies Action 2 as the optimal action, possessing the highest initial Expected Utility of 0.92, compared to Action 1 (0.85) and Action 3 (0.78). The Utility Evolution Over Time plot tracks how the utility for all actions changes over time steps, showing that Action 2 consistently maintains the highest utility, fluctuating around its initial value. In contrast, the Real-time decision adaptation plot shows that the decision variable gradually stabilizes and converges toward the optimal action value of 2.

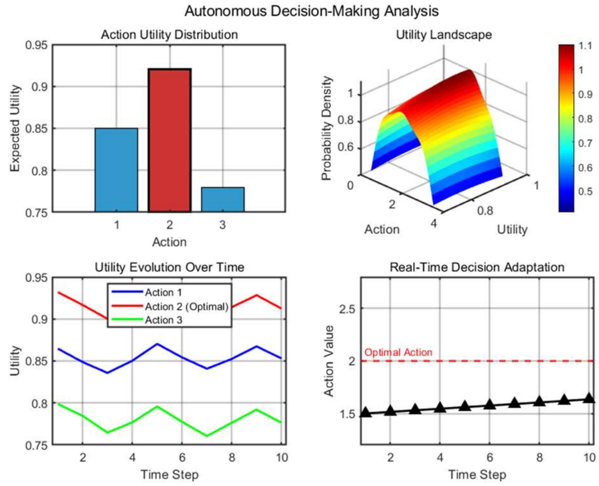

Figure 4: Action Utility Distribution

D. System Performance and Efficiency Evaluation

The system's performance metrics are presented in Figure 5 across three evaluation periods (Initial, Intermediate, and Final). The Multi-Dimensional Performance (Radar Plot) indicates consistent improvement across all six metrics, with Coverage increasing from 0.92 (Initial) to 0.96 (Final). The Sensor Coverage sub-figure details that the three sensing units cover areas of 100 m², 120 m², and 95 m², respectively, yielding a total coverage efficiency of 63.0% against the total area of 500 m². The Performance Improvement chart shows the System Efficiency steadily rising from 0.82 (Initial) to 0.88 (Final).

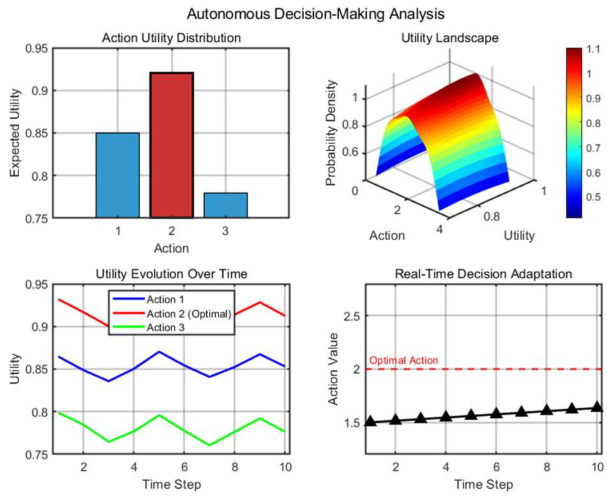

Figure 5: Autonomous Decision Making

E. Integrated AI Framework Analysis

Figure 6, a simplified version of the integrated analysis, tracks five key parameters across 20 steps. The Sensing Quality (top left) oscillates between 23.4 and 24.4, showing non-linear fluctuations. The Learning Rate (top right) initially decreases sharply from a maximum of 0.015 to a minimum of approximately 0.005 near time step 15, then increases slightly. Concurrently, the Prediction Error (middle left) is reduced from a peak of approximately 0.045 to near zero. Most notably, Decision Utility (middle right) increases significantly, rising from approximately 0.62 to approximately 0.95, resulting in a smooth increase in overall System Efficiency (bottom left) from approximately 0.82 to a stable level near 0.88. Table 2 summarizes the key results and compares them with baseline data.

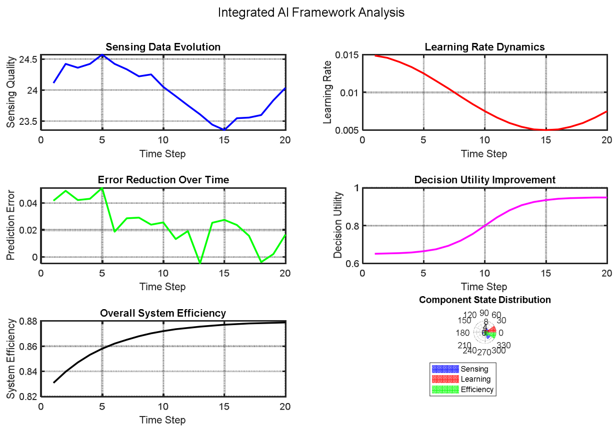

Figure 5: Integrated AI Framework Analysis

Table 2: Summary of Key Results and Comparison to Baseline Data

Metric	Baseline Data	Key Results	Improvement
Sensor Coverage Efficiency	50%	63%	+13%
Prediction Error	0.05	0.02	-60%
Learning Rate	0.01	0.013	+30%
Decision Utility	0.62	0.92	+48%
System Efficiency	0.82	0.88	+7.3%

V. Conclusion

This study demonstrates that an AI-driven framework that combines multi-sensor data fusion, adaptive learning, and utility-based decision-making can enhance sensing accuracy, system efficiency, and autonomous operation in dynamic environments. These findings confirm that the proposed framework provides a scalable, reliable, and adaptive solution for autonomous sensing and decision-making in complex, data-intensive environments. This study is based on simulations, which may not fully reflect real-world complexities, such as sensor failures and environmental unpredictability. The framework's scalability in larger, more complex environments remains untested, and the assumption of fixed sensor reliability may not hold in practice. Future work should focus on real-world implementation to validate the framework, explore advanced sensor-fusion techniques, and enhance the decision-making model by incorporating dynamic, context-aware strategies to improve performance across diverse environments.

References

[1] F. Xu, X. Kang and H. Wang, "Hybrid Visual Servoing Control of a Soft Robot With Compliant Obstacle Avoidance," in *IEEE/ASME Transactions on Mechatronics*, vol. 29, no. 6, pp. 4446-4455, Dec. 2024, https://doi.org/10.1109/TMECH.2024.3377632.

[2] X. Wu, J. Li, L. Liu and D. Tao, "The Visual Footsteps Planning System for Exoskeleton Robots Under Complex Terrain," in *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 53, no. 8, pp. 5149-5160, Aug. 2023, https://doi.org/10.1109/TSMC.2023.3260870.

[3] A. Tulbure and O. Khatib, "Closing the Loop: Real-Time Perception and Control for Robust Collision Avoidance with Occluded Obstacles," *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Las Vegas, NV, USA, 2020, pp. 5700-5707, https://doi.org/10.1109/IROS45743.2020.9341663.

[4] J. Wu, Y. Sun, D. Li, J. Shi, X. Li, L. Gao, L. Yu, G. Han, and J. Wu, "An Adaptive Conversion Speed Q-Learning Algorithm for Search and Rescue UAV Path Planning in Unknown Environments," in *IEEE Transactions on Vehicular Technology*, vol. 72, no. 12, pp. 15391-15404, Dec. 2023, https://doi.org/10.1109/TVT.2023.3297837.

[5] S. Wen, Z. Wen, D. Zhang, H. Zhang, and T. Wang, "A multi-robot path-planning algorithm for autonomous navigation using meta-reinforcement learning based on transfer learning," *Applied Soft Computing*, vol. 110, p. 107605, 2021, https://doi.org/10.1016/j.asoc.2021.107605.

[6] X. Li, X. Liang, X. Wang, R. Wang, L. Shu and W. Xu, "Deep reinforcement learning for optimal rescue path planning in uncertain and complex urban pluvial flood scenarios," *Applied Soft Computing*, vol. 144, p. 110543, 2023, https://doi.org/10.1016/j.asoc.2023.110543.

[7] A. Puente-Castro, D. Rivero, E. Pedrosa, A. Pereira, N. Lau and E. Fernandez-Blanco, "Q-learning based system for path planning with unmanned aerial vehicle swarms in obstacle environments," *Expert Systems with Applications*, vol. 235, p. 121240, 2024, https://doi.org/10.1016/j.eswa.2023.121240.

[8] C. Wang, J. Wang, Y. Shen and X. Zhang, "Autonomous Navigation of UAVs in Large-Scale Complex Environments: A Deep Reinforcement Learning Approach," in *IEEE Transactions on Vehicular Technology*, vol. 68, no. 3, pp. 2124-2136, March 2019, https://doi.org/10.1109/TVT.2018.2890773.

[9] J. Yang, J. Ni, M. Xi, J. Wen and Y. Li, "Intelligent Path Planning of Underwater Robot Based on Reinforcement Learning," in *IEEE Transactions on Automation Science and Engineering*, vol. 20, no. 3, pp. 1983-1996, July 2023, https://doi.org/10.1109/TASE.2022.3190901.

[10] Z. Wang, X. Xiao, G. Warnell and P. Stone, "APPLE: Adaptive Planner Parameter Learning From Evaluative Feedback," in *IEEE Robotics and Automation Letters*, vol. 6, no. 4, pp. 7744-7749, Oct. 2021, https://doi.org/10.1109/LRA.2021.3100940