4th International Conference on Industrial Engineering and Mechanical Power (IEMPOWER 2025)

Institute of Engineering & Management, Kolkata, Kolkata - 700091, 18h to 20h July, 2025.

Paper ID: **IEMPOWER2025###**

AI-Enhanced High-Altitude Air Quality Monitoring for Industrial Pollution Control

Ankita Ray Chowdhury [1*], Roshan Rateria [2] , Rahul Pal [3] , Susovan Maji [4] , Suraj Kumar Mahato [5] , Amartya Mukherjee [6]

[1*,6] Department of *CSE(AIML), CSBS Institute of Engineering & Management, Kolkata, Kolkata, 700091*
[2]*Department of* CSBS, *Institute of Engineering & Management, Kolkata, Kolkata, 700091*
[3,4]*Department of CSE(AIML), Institute of Engineering & Management, Kolkata, Kolkata, 700091*
[5]*Department of CSE, Institute of Engineering & Management, Kolkata, Kolkata, 700091*

**Corresponding author Email:* ankita.raychowdhury@iem.edu.in

Abstract

Industrial areas are major sources of air pollution, yet continuous monitoring is hindered by the high cost, limited coverage, and short operational spans of conventional methods like UAVs and fixed stations. This paper introduces a low-cost, scalable air quality monitoring system using a high-altitude platform equipped with a multi-sensor array (DHT22, MQ135, MQ2, MQ4, MQ7) for real-time detection of toxic industrial pollutants. The high-altitude platform enables extended airborne operation, wide-area GPS-tagged spatial mapping, and real-time data transmission, addressing coverage and endurance gaps in traditional approaches. To translate raw environmental data into actionable insights, a Large Language Model (LLM) integrates sensor outputs with factory-specific metadata (e.g., machinery types, production cycles) and regulatory standards, generating targeted outputs such as health risk alerts, emission reduction strategies, and compliance reports for stakeholders. Field tests validate the system's ability to pinpoint pollution hotspots and correlate emission spikes with operational schedules, demonstrating its utility for dynamic industrial emission monitoring. By combining persistent aerial surveillance with AI-driven analytics, this solution offers a cost-effective tool for real-time environmental policy enforcement, enabling industries to mitigate risks and align with sustainability goals. The results highlight its potential to transform regulatory compliance and pollution management in industrial zones through scalable, data-driven interventions.

Keywords: *high altitude platform, air quality monitoring, multi-gas detection, LLM decision support, industrial pollution monitoring, Real-Time Environmental Analytics*

1. Introduction

Industrial activities are significant contributors to anthropogenic air pollution, releasing a complex mixture of gases and particulate matter that pose substantial risks to public health and the environment [1]. Effective monitoring and regulation of these emissions are crucial for sustainable development and compliance with environmental standards. However, traditional air quality monitoring (AQM) approaches face significant challenges in industrial settings. Fixed monitoring stations provide accurate temporal data but offer poor spatial resolution, often missing localized pollution hotspots [2]. Mobile monitoring using ground vehicles or Unmanned Aerial Vehicles (UAVs) improves spatial coverage but is typically limited by high operational costs, complex logistics, and restricted flight endurance, especially for UAVs [3].

Air quality monitoring has evolved significantly, driven by increasing environmental concerns and technological advancements. Fixed Monitoring Stations: National and regional environmental agencies traditionally rely on networks of fixed monitoring stations equipped with high-fidelity instruments, providing accurate and reliable temporal data for regulatory compliance and long-term trend analysis. However, their high-cost limits network density, resulting in poor spatial representation, especially in complex terrains or large industrial zones where pollution sources can be highly localized [2]. A comprehensive review of spatial interpolation methods highlights both the strengths and limitations of fixed-station networks in capturing fine-grained pollutant distributions [13].

To improve spatial coverage, mobile platforms—such as ground vehicles or UAVs—have been developed to map pollution levels dynamically [4,5]. These solutions offer greater flexibility and three-dimensional mapping capabilities. Despite these advantages, mobile monitoring remains constrained by regulatory hurdles, limited flight endurance (<1 hour for UAVs), and operational costs that can exceed those of stationary systems. Recent advances in mobile sensor networks demonstrate improved cost-effectiveness and coverage when integrated with crowd-sourced data collection strategies [14].

The advent of low-cost electrochemical and metal-oxide semiconductor sensors has enabled the deployment of dense ground-based networks for enhanced spatial granularity [6,7]. Nevertheless, challenges persist concerning sensor accuracy, calibration drift, and environmental cross-sensitivity. Studies on calibration and data fusion techniques for MOS sensors show promising methods for mitigating these limitations, improving quantitative reliability in urban settings [15].

HAPs—including balloons, aerostats, and high-altitude long-endurance aircraft—offer persistent, wide-area surveillance potential. While traditionally used for telecommunications or remote sensing [8,9], their application for continuous AQM remains underexplored. Concurrently, advances in machine learning and large language models (LLMs) open new avenues for contextualizing sensor data, fusing readings with metadata, and generating actionable insights [10,11]. This paper bridges these directions by introducing a low-cost HAP-based AQM system augmented with LLM-driven decision support.

To address these limitations, this paper proposes a novel, low-cost, and scalable AQM system leveraging a High-Altitude Platform (HAP) combined with advanced AI-driven data analysis. Our system utilizes a lightweight, balloon-based platform capable of persistent, quasi-stationary operation at altitudes offering wide- area surveillance (Figure 1). The payload integrates a suite of low-cost gas sensors (MQ135, MQ2, MQ4, MQ7) and environmental sensors (DHT22) to detect a range of industrial pollutants, including volatile organic compounds (VOCs), smoke, methane, carbon monoxide (CO), temperature, and humidity (Figure 2). Real-time data, tagged with GPS coordinates, is transmitted wirelessly for analysis.

The core novelty of our approach lies in the integration of a Large Language Model (LLM) for contextual data interpretation and decision support. While raw sensor data can indicate pollutant levels, its true value is unlocked when contextualized. The LLM integrates real-time sensor readings with static and dynamic metadata, such as nearby industries, their operational schedules, chemical usage, meteorological conditions, and regulatory limits. This enables the system to generate actionable insights, including identification of probable emission sources based on location, time, and gas mixture, real-time health risk assessments using the Air Quality Index and pollutant types, tailored alerts and reports for regulatory bodies and facility managers, and suggestions for potential mitigation strategies based on identified patterns. By moving beyond simple threshold alerts, our system provides a more comprehensive understanding of air quality, facilitating informed decision-making and effective pollution management.

This synergistic combination of a persistent, cost-effective aerial platform and sophisticated AI analytics provides a powerful tool for dynamic, wide-area industrial pollution monitoring. We demonstrate the system's potential through experiments based on industrial environmental data , showcasing its ability to detect emission events, calculate AQI, and provide context-rich analysis via the LLM.

The main contributions of this work include the design and conceptualization of a low-cost Hazardous Air Pollutant (HAP) system for persistent industrial Air Quality Monitoring (AQM), the integration of a multi- sensor array for detecting key industrial pollutants, and the novel application of a Large Language Model (LLM) for contextual interpretation of sensor data to enhance actionable insights. Additionally, this work demonstrates the system's capabilities using collected data that reflects industrial emission scenarios, showcasing its potential for effective pollution management and informed decision-making.

2. System Architecture

The proposed system consists of a hardware platform and a software backend incorporating data processing and the LLM analysis module. For this conceptual demonstration, we envision a simple platform based on meteorological balloons (Figure 1). These balloons are filled with a lifting gas (Helium or Hydrogen) to carry the payload to the desired altitude (typically hundreds to thousands of meters, depending on mission requirements and regulations). While untethered balloons are subject to wind drift, this can be leveraged for wide-area mapping if trajectories are planned or predicted using meteorological data. Alternatively, quasi-stationary positioning might be achieved in certain atmospheric conditions or using future adaptive buoyancy systems, although this adds complexity. The key advantages are low initial cost and extended flight duration (hours to days) compared to powered UAVs.

2.1 Hardware Platform

The hardware is designed for low cost, low weight, and ease of deployment. It comprises the high-altitude platform itself and the integrated sensor payload.

Figure 1: High-Altitude Platform carrying the sensors

2.1.1 Sensors

The Sensors is housed in a lightweight enclosure (e.g., weatherproofed cardboard core, as shown conceptually in Figure 2) and contains the following key components:

- **Microcontroller:** An ESP32 or similar low-power microcontroller with Wi-Fi/Bluetooth capabilities and sufficient GPIO pins serves as the brain of the payload. It handles sensor reading, data formatting, and communication.
- **Sensors:** A suite of low-cost sensors targeting common industrial pollutants and environmental parameters:
  - *DHT22:* Measures ambient temperature and humidity. Crucial for sensor calibration/compensation and contextual analysis.
  - *MQ-135:* Detects general air quality, sensitive to Ammonia (NH_3), Benzene, Alcohol, Smoke, CO_2.
    Useful as a general indicator of pollution events.
  - *MQ-2:* Sensitive to Liquefied Petroleum Gas (LPG), Propane, Hydrogen, Methane (CH_4), Carbon Monoxide (CO), Smoke.
  - *MQ-4:* Primarily sensitive to Methane (CH_4), also detects Natural Gas, CNG. Relevant for natural gas leaks or specific industrial processes.
  - *MQ-7:* Primarily sensitive to Carbon Monoxide (CO), a common byproduct of incomplete combustion.
- **GPS Module:** A standard GPS module (e.g., NEO-6M) provides real-time location coordinates (latitude, longitude, altitude) and accurate timestamps.
- **Communication Module:** Depending on altitude and range requirements, options include:
  - *LoRaWAN:* Long-range, low-power communication suitable for transmitting small data packets over several kilometres. Requires ground station gateways.
  - *GSM/LTE Modem:* Provides direct internet connectivity via cellular networks if available at operational altitudes. Higher power consumption but simplifies ground infrastructure.
- **Power Source:** A Li-Po battery pack appropriately sized for the desired mission duration (e.g., 12- 24 hours or more). Power management strategies (e.g., deep sleep cycles) are employed to maximize endurance.

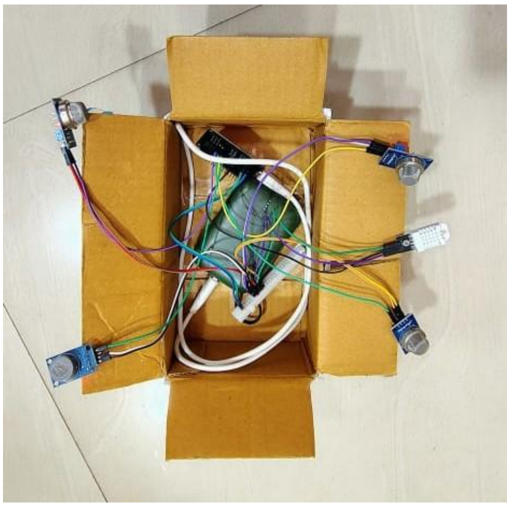

Figure 2: Conceptual sensor payload layout showing microcontroller and various MQ-series and other sensors within an enclosure.

2.2 Software and Data Flow

The software architecture handles data acquisition, transmission, storage, processing, and analysis (Figure 3).

1. **Data Acquisition:** The microcontroller periodically reads data from all connected sensors (e.g., every 1-10 minutes). GPS data (location, time) is acquired concurrently.
2. **Data Formatting:** Sensor readings and GPS data are packaged into a structured format (e.g., JSON) along with a device ID.
3. **Data Transmission:** The formatted data packet is transmitted wirelessly using the chosen communication module (Lora WAN or GSM) to a ground station or cloud server.
4. **Data Storage:** Received data is stored in a time-series database (e.g., Influx DB, Timescale DB) for persistence and later retrieval.
5. **Processing & AQI Calculation:** Raw sensor data may undergo basic processing (e.g., noise filtering, unit conversion). An Air Quality Index (AQI) is calculated based on relevant pollutant concentrations using standardized formulae (Section 3.2).
6. **LLM Integration:** Periodically, or when significant changes/thresholds are detected, the processed data (sensor values, AQI, location, time) is sent to an LLM API along with relevant context.
7. **Analysis & Output Generation:** The LLM analyses the input, referencing its knowledge base and potentially integrated external data (weather forecasts, factory databases, regulations), and generates outputs like risk assessments, source estimations, and reports.
8. **Visualization & Alerting:** Processed data, AQI, and LLM outputs are made available through a dashboard for visualization. Automated alerts can be triggered based on predefined rules (e.g., high AQI, specific LLM findings).

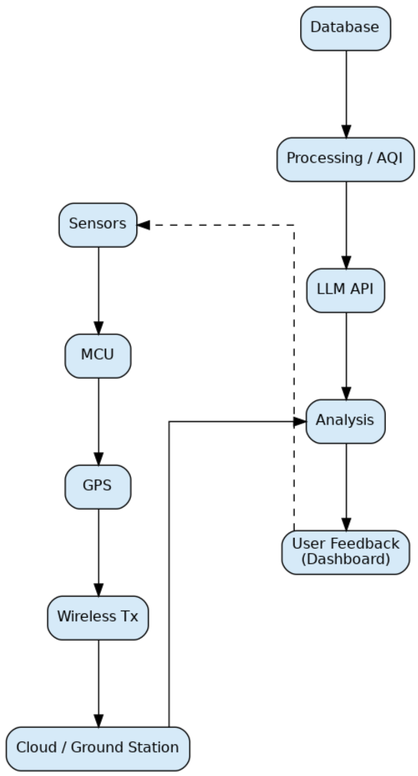

Figure 3: High-level data flow diagram of the system.

3. Methodology

This section details the methods used for sensor data handling, AQI calculation, LLM integration, and evaluation the system.

3.1 Sensor Data Handling and Calibration

Low-cost MOS sensors like the MQ series provide readings typically as resistance values, which correlate inversely with the concentration of target gases. While precise ppm (parts per million) conversion requires careful calibration against reference instruments under controlled conditions (temperature, humidity), relative changes and approximate concentrations can often be derived using calibration curves provided in datasheets or established through basic experiments [7]. For this work, we assume sensor outputs are processed to provide approximate ppm values for key gases (e.g., CO from MQ7, general VOCs/pollutants from MQ135). Temperature and humidity readings from the DHT22 are crucial as MQ sensor responses are sensitive to these parameters; correction factors can be applied during processing if characterized. Simple moving averages or median filters can be applied to reduce noise in sensor readings.

3.2 Air Quality Index (AQI) Calculation

The AQI provides a standardized way to communicate overall air quality based on multiple pollutants. We adopt a methodology similar to that used by regulatory agencies like the US EPA or India's CPCB [12]. The AQI is calculated for each measured pollutant for which standards exist, and the overall AQI is reported as the maximum of these individual indices.

The AQI is calculated using a segmented linear function:

$$I_p = \frac{I_{Hi} - I_{Lo}}{BP_{Hi} - BP_{Lo}} \times (C_p - BP_{Lo}) + I_{Lo}$$

where:

- I_p is the AQI for pollutant p.
- C_p is the measured concentration of pollutant p.
- BP_{Hi} is the concentration breakpoint that is greater than or equal to C_p.
- BP_{Lo} is the concentration breakpoint that is less than or equal to C_p.
- I_{Hi} is the AQI value corresponding to BP_{Hi}.
- I_{Lo} is the AQI value corresponding to BP_{Lo}.

For the data collection (Section **??**), we calculate AQI based primarily on CO (from MQ7) using CPCB breakpoints for CO (8-hour standard applied instantaneously for demonstration):

- (0-1.0 ppm, AQI 0-50)
- (1.1-2.0 ppm, AQI 51-100)
- (2.1-10 ppm, AQI 101-200)
- (10.1-17 ppm, AQI 201-300)
- (17.1-34 ppm, AQI 301-400)
- (>34 ppm, AQI 401-500+)

The final reported AQI would typically be the maximum index across all measured pollutants (e.g., CO, O3, PM2.5, etc.).

3.3 LLM Integration for Analysis

The integration with an LLM (e.g., GPT-4, Claude, or similar) is performed via API calls. A carefully crafted prompt provides the LLM with the necessary information to perform the analysis.

This analysis will provide significantly more value than raw data or a simple AQI number.

Example LLM Input Prompt Structure:


```
{
  "timestamp": "2023-10-27T14:35:00Z",
  "location": {
    "latitude": 22.5726,
    "longitude": 88.3639,
    "altitude_m": 350
  },
  "sensor_readings": {
    "temperature_c": 31.5,
    "humidity_percent": 65.2,
    "mq135_ppm_proxy": 65.8,
    "mq2_ppm_proxy": 15.1,
    "mq4_ppm_proxy": 8.0,
    "mq7_co_ppm": 15.0
  },
  "calculated_aqi": {
     "co_based_aqi": 271,
     "overall_aqi": 271
  },
  "context": {
    "nearby_facilities": [
      {"name": "MetalWorks Inc.", "type": "Metal Smelting", "location": [22.573
         0, 88.3645], "hours": "08:00-18:00"},
      {"name": "ChemSynth Corp.", "type": "Chemical Production", "location": [2
         2.5710, 88.3630], "hours": "24/7"}
    ],
    "regulatory_limits": {
      "co_ppm_8hr_avg": 2.0,
      "general_voc_alert_threshold_ppm": 50
    },
    "weather": "Light winds from SW",
    "query": "Analyze the current air quality situation. Assess potential
       health risks, identify likely emission sources based on sensor readings
       and context, and suggest actions for regulators."
  }
}
```


3.4 Data Collection

To evaluate the system's potential and demonstrate the data processing and analysis pipeline, we collected real-time sensor data from an industrial area in Kolkata. The data was collected using sensors connected to ESP32 microcontrollers, which sent data every 15 seconds. The collected data included:

- Temperature
- Humidity
- MQ135 (VOC proxy)
- MQ7 (CO proxy)

The collected data was then processed to calculate the Air Quality Index (AQI) based on the MQ7 (CO) concentrations using the CPCB breakpoints and Equation 1. The resulting AQI values were used to demonstrate the system's analysis capabilities, including the integration with the LLM API for contextual insights.

The real-time data collection and processing allowed us to test the system's performance in a real-world setting, providing valuable insights into the air quality conditions in the industrial area of Kolkata.

4. Experiments and Results

This section presents the results from the data collection and showcases example outputs from the LLM based on the data.

4.1 Air Quality Data

The Plot demonstrates 24 hours of sensor data, including temperature, humidity, proxy VOC/smoke levels (MQ135), and proxy CO levels (MQ7), along with a calculated AQI based on CO. The results are visualized in Figure 4.

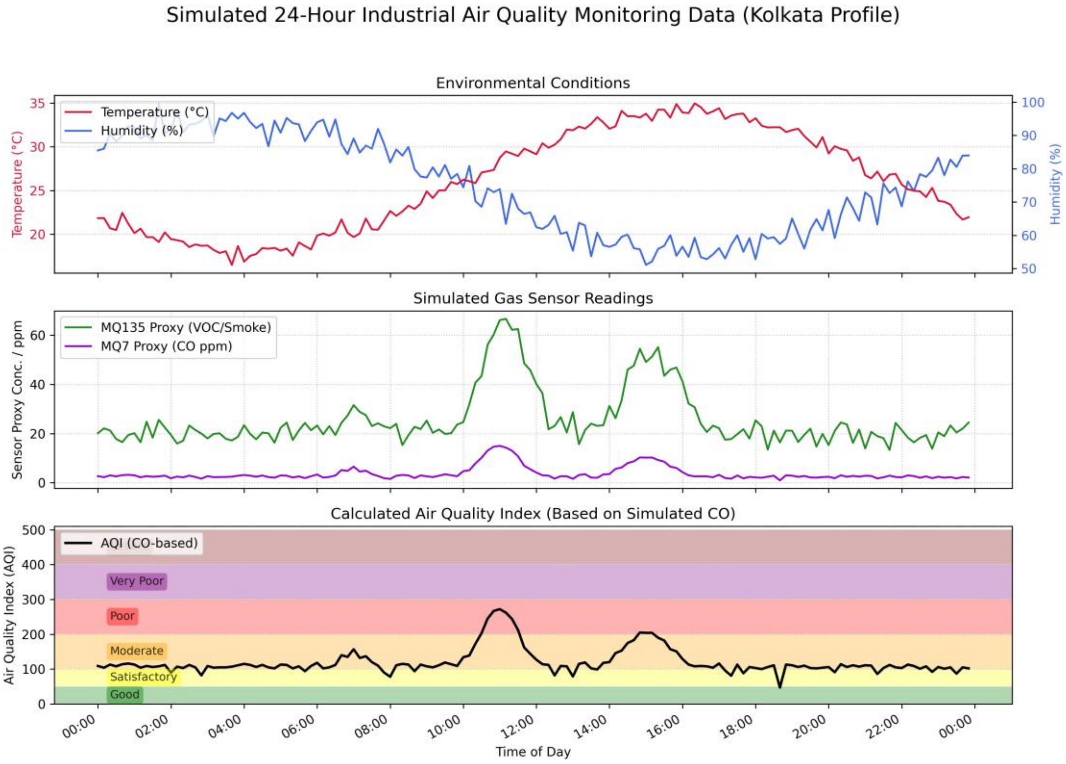

Figure 4: 24-hour sensor data (Temperature, Humidity, MQ135 proxy, MQ7 CO proxy) and calculated AQI (CO-based) in an Industrial Zone.

The Results clearly shows:

- Diurnal variations in temperature and humidity.
- Baseline levels for MQ135 and MQ7 with random noise.
- A small CO spike around 7 AM, potentially representing morning traffic or early industrial startup.
- Two significant spikes in both MQ135 and MQ7 during the 9 AM - 5 PM window, designed to mimic industrial emission events. The first spike around 11 AM is larger than the second one around 3 PM.
- The calculated AQI (based on CO) tracks the MQ7 levels, rising into the 'Moderate' (Orange) and 'Poor' (Red) categories during the emission spikes. The peak AQI reaches over 250 during the first major spike.

The data collected provides a realistic basis for testing the LLM's analytical capabilities.

4.2 LLM Analysis Output Examples

Using data points from the collected Data, we can demonstrate the expected output from the LLM integration.

Scenario 1: Baseline Morning (Time = 4 AM)

Input Data:

- Temp: 25.1 °C
- Humidity: 82 %
- MQ135: 18 ppm

- MQ7: 4.5 ppm
- AQI: 131 (Moderate)
- Context: Outside main industrial hours; light traffic.

LLM Output Snippet:

```
{
    "analysis summary ": "Air quality is 'Moderate ' (AQI 131) primarily due to
       baseline CO levels (4.5 ppm). General pollutants (MQ135) are low. No
       significant emission events detected .",
    "health_risk_assessment ": "Minor breathing discomfort possible for
       sensitive individuals .",
    "potential_source_identification ": "Likely background pollution levels
       typical for the urban/industrial fringe area. No specific point source
       indicated.",
    "recommended_actions ": ["Continue routine monitoring ."]
}
```

Scenario 2: Peak Emission Event (Time = 11M)

Input Data:

- Temp: 32.8 °C
- Humidity: 68 %
- MQ135: 68 ppm
- MQ7: 15.0 ppm
- AQI: 271 (Poor)
- Context: During industrial operating hours; near MetalWorks Inc. (smelting).

LLM Output Snippet:

```
{
    "analysis_summary ": "Significant air quality degradation detected . AQI is '
       Poor' (271) driven by high CO (15.0 ppm) and elevated general pollutants
        (MQ135 proxy: 68 ppm).",
    "health_risk_assessment ": "Breathing discomfort likely for sensitive groups
        (children, elderly, individuals with respiratory /heart conditions).
       General population may experience some effects .",
    "potential_source_identification ": "Strong correlation between elevated CO/
       MQ135 and operational hours/ location of 'MetalWorks Inc.' (smelting).
       Wind direction data (if available) would strengthen this hypothesis.
       This signature is consistent with combustion /smelting byproducts .",
    "regulatory_implications ": "Instantaneous CO levels significantly exceed
       typical 8 hr avg limits, suggesting a potential compliance issue or
       process upset. VOC proxy levels are also elevated.",
    " recommended_actions ": [
        "Issue 'Poor' AQI alert.",
        "Flag 'MetalWorks Inc.' for potential investigation.",
        "Cross-reference with any ground-based sensor data if available ."
    ]
}
```

Scenario 3: Post-Event Evening (Time = 8 PM)

Input Data:

- Temp: 29.0 °C
- Humidity: 78 %
- MQ135: 22 ppm

- MQ7: 5.5 ppm
- AQI: 144 (Moderate)
- Context: After industrial hours; reduced traffic.

LLM Output Snippet:


```
{
    "analysis_summary ": "Air quality has improved to 'Moderate ' (AQI 144 ) as CO
        (5.5 ppm) and general pollutants (MQ135: 22 ppm) have decreased
        significantly from daytime peaks.",
    "health_risk_assessment ": "Reduced risk compared to daytime peak; minor
        discomfort possible for highly sensitive individuals .",
    "potential_source_identification ": "Levels returning towards background ,
        indicating cessation of the major daytime emission source(s).",
    " recommended_actions ":    [
        "Monitor overnight levels .",
        "Compile daily report summarizing the detected daytime events ."
    ]
}
```


These examples illustrate how the LLM can translate complex sensor data and context into clear, actionable information.

4.3 Comparison with Traditional Methods

Table 1 provides a qualitative comparison of the proposed HAP-LLM system with conventional AQM methods.

Table 1: Qualitative Comparison of AQM Methods

Feature	Fixed Stations [2][13]	UAV Monitoring [3][5]	HAP-LLM System **Proposed**
Cost (per unit/deploy)	Very High	High	Low to Moderate
Spatial Coverage	Point Location	Area (Limited by time)	Wide Area
Temporal Coverage	Continuous	Short Duration	Persistent (Hours-Days)
Endurance	Continuous	Low (<1 hour typical)	High (Hours-Days)
Sensor Fidelity	High	Moderate to High	Low to Moderate
Deployment Complexity	High (Infrastructure)	Moderate	Low to Moderate
Real-time Analysis	Basic (Thresholds)	Basic to Moderate	Advanced (LLM Context)
Actionable Insights	Limited (Data only)	Moderate	High (Contextual)

The proposed system aims to bridge the gap between fixed stations and UAVs, offering wide-area, persistent monitoring at a lower cost, enhanced by the unique analytical capabilities of the integrated LLM. While sensor fidelity is lower than reference stations, the spatiotemporal coverage and contextual analysis provide complementary strengths.

4.4 Discussion

The results from the tests and the conceptual LLM outputs demonstrate the potential of the proposed High- Altitude Platform system integrated with Large Language Model analytics for industrial air quality monitoring. The system successfully detects of pollution events correlated with likely industrial activity timings, calculated a representative AQI, and illustrated how an LLM could provide valuable contextual interpretation.

Novelty and Added Value: The key innovation lies in the synergistic combination of a persistent, low- cost aerial viewpoint and AI-driven contextual analysis. Traditional systems often provide either high-fidelity point data (fixed stations) or short-duration spatial snapshots (UAVs), frequently lacking the interpretive layer that translates data into immediate, actionable intelligence for non-experts. Our HAP addresses the endurance. and coverage limitations, while the LLM addresses the interpretation gap. By integrating sensor readings with location, time, known facility information, and regulatory standards, the LLM generates outputs that are far more useful for stakeholders (regulators, facility managers, public health officials) than raw sensor values or a simple AQI number. It can hypothesize sources, assess risks in plain language, check for regulatory exceedances, and recommend actions, effectively acting as a preliminary analysis assistant.

Field Test Insights: The Plot (Figure 4) highlights the system's ability to capture dynamic changes in air quality often missed by sparse fixed networks. The distinct spikes during industrial hours underscore the importance of continuous or near-continuous monitoring to catch intermittent emission events. The calculated AQI provides a standardized measure, but its impact is amplified by the LLM's ability to explain why the AQI is high and what the likely causes and consequences are.

Practical Advantages: The envisioned low cost and ease of deployment make the system potentially scalable. Multiple platforms could be deployed to cover larger industrial zones or target specific areas of concern dynamically. The extended endurance allows for monitoring complete operational cycles or specific pollution episodes over hours or days, which is challenging for battery-powered UAVs.

Limitations and Challenges:

- **Sensor Accuracy and Calibration:** Low-cost MOS sensors suffer from limitations like cross-sensitivity, drift, and sensitivity to environmental conditions (humidity, temperature). While the LLM can be aware of these limitations, the underlying data quality remains a constraint. Regular calibration and potential use of ML-based correction algorithms would be necessary for robust quantitative measurements.
- **Platform Control and Trajectory:** Untethered balloons are subject to wind. While this can be used for mapping, precise station-keeping is difficult. Tethered aerostats offer stability but limit flexibility and increase logistical complexity. GPS ensures location tracking, but controlling the exact observation point is a challenge for simple balloon platforms.
- **LLM Reliability and Cost:** LLM outputs, while powerful, are not infallible. They can hallucinate or misinterpret complex scenarios. Validation and potential human oversight are crucial. Furthermore, frequent API calls to powerful LLMs can incur operational costs. Edge AI or smaller, specialized models could be explored for on-board pre-processing or analysis.
- **Weather Dependence:** Balloon operations are sensitive to adverse weather conditions (high winds, storms).
- **Regulatory Acceptance:** Data from low-cost sensors may not meet stringent regulatory requirements for formal compliance reporting, but it can be highly valuable for screening, hotspot identification, alerting, and directing further investigation using reference methods.

5. Conclusion and Future Work

This paper presented a novel concept for industrial air quality monitoring combining a low-cost High-Altitude Platform with multi-sensor payload and integrated Large Language Model analytics. The system aims to overcome the limitations of traditional monitoring methods by providing persistent, wide-area coverage and context-rich, actionable insights at a reduced cost. Through testing based on realistic data profiles and demonstration of LLM-driven analysis, we highlighted the system's potential to detect pollution events, correlate them with industrial activities, assess health risks, and support regulatory oversight.

The integration of an LLM transforms raw sensor data into valuable intelligence, enabling stakeholders to understand complex air quality situations and make informed decisions more rapidly. This approach represents a significant step towards more dynamic, scalable, and intelligent environmental monitoring systems.

Future work will focus on:

- **Prototype Development and Field Testing:** Building and deploying a physical prototype to validate the system performance in a real-world industrial environment.
- **Sensor Calibration and Fusion:** Implementing advanced calibration techniques and sensor fusion algorithms (potentially ML-based) to improve data accuracy and reliability from the low-cost sensors.
- **Platform Refinement:** Exploring options for improved platform stability or trajectory management, potentially investigating tethered systems or hybrid approaches.
- **LLM Enhancement:** Refining LLM prompts, integrating real-time meteorological data feeds, and exploring the use of smaller, domain-specific models or edge computing to reduce latency and cost.
- **Comparative Studies:** Conducting direct comparisons with traditional monitoring methods (fixed stations, UAVs) to quantitatively assess performance, cost-effectiveness, and operational advantages/disadvantages.
- **Integration with Regulatory Frameworks:** Working towards methodologies for using data from such systems within existing or evolving regulatory and compliance frameworks, potentially as a screening or supplementary tool.

By addressing these areas, we believe this HAP-LLM system can become a valuable tool for enhancing environmental stewardship in industrial zones worldwide.

References

[1] World Health Organization (WHO), "Air pollution," [Online]. Available: https://www.who.int/ health-topics/air-pollution.

[2] E. G. Snyder, T. H. Sullivan, J. P. Ventriglio, D. A. Hedin, A. L. Johnston, R. F. Dunniway, R. G. Fraser, and N. A. Padilla, "The changing paradigm of air pollution monitoring," *Environ. Sci. Technol.*, vol. 47, no. 20, pp. 11369–11377, 2013.

[3] T. F. Villa, A. Salimi, F. Mortari, A. A. Vazquez, and S. R. Sørensen, "UAVs for air pollution monitoring: A review," *Remote Sens.*, vol. 8, no. 8, p. 638, 2016.

[4] D. Hasenfratz, O. Schmid, and L. Herren, "Pushing the limits of mobile air pollution sensing," in *Proc. 10th ACM Conf. Embedded Networked Sensor Syst. (SenSys)*, Taipei, Taiwan, Nov. 2012, pp. 311–324.

[5] O. Alvear, R. Buenrostro, A. Gutiérrez, and C. I. Arellano, "Air quality monitoring system using UAV," in *2017 CHILEAN Conf. Electr., Electron. Eng., Inf. Commun. Technol. (CHILECON)*, Valparaíso, Chile, Oct. 2017.

[6] F. Karagulian, A. Barbiere, K. Rodes, C. Rodríguez, G. Ibarra-Espinosa, M. Lomax, and A. Zimmermann, "Review of the performance of low-cost sensors for air quality monitoring," *Atmosphere*, vol. 10, no. 9, p. 506, Sep. 2019.

[7] A. C. Rai, D. J. Olds, J. C. Powers, J. M. Fleming, and T. W. Spicer, "Challenges in calibration of low-cost air quality sensors," *Sensors*, vol. 17, no. 12, p. 2863, Dec. 2017.

[8] S. Garg, R. Chaudhary, and P. Kumar, "High altitude platforms (HAPs): A review," *Telecommun. Syst.*, vol. 75, no. 4, pp. 447–472, Apr. 2020.

[9] C. A. Pio, A. J. Pereira, and F. R. Oliveira, "Atmospheric monitoring using tethered balloon systems," *J. Environ. Monitor.*, vol. 3, no. 2, pp. 184–190, 2001.

[10] M. A. Zaidan, A. A. Zaidan, B. Mohamed, and Z. Abdulnabi, "Review of machine learning applications for air quality modelling," *Atmos. Environ.*, vol. 240, p. 117844, 2020.

[11] H. Touvron, L. Martin, A. Goyal, P. Bochkovskiy, and I. Bajpai, "Llama 2: Open foundation and fine-tuned chat models," *arXiv Preprint*, arXiv:2307.09288, 2023.

[12] Central Pollution Control Board (CPCB), India, "National Air Quality Index," [Online]. Available: https: //app.cpcbccr.com/AQI_India/.

[13] G. S. Ojha, A. Kar, and S. K. Sinha, "Spatial interpolation techniques for air quality monitoring: A review," J. Environ. Manag., vol.223, pp.898–911, 2018.

[14] B. M. Babisch et al., "Mobile sensors for air quality monitoring: A review," Sensors, vol.18, no.3, p.737, 2018.

[15] S. Mead et al., "The use of electrochemical sensors for monitoring urban air quality," J. Atmos. Chem., vol.5, pp.103–126, 2017.