A Modular LLM-Agent System for Transparent Multi-Parameter Weather Interpretation

Daniil Sukhorukov , Andrei Zakharov , Nikita Glazkov , Katsiaryna Yanchanka , Vladimir Kirilin , Maxim Dubovitsky , Roman Sultimov , Yuri Maksimov , Ilya Makarov

arXiv:2512.11819v1 [cs.CY] 28 Nov 2025

Abstract

Weather forecasting is not only a predictive task but an interpretive scientific process requiring explanation, contextualization, and hypothesis generation. This paper introduces AI-Meteorologist, an explainable LLM-agent framework that converts raw numerical forecasts into scientifically grounded narrative reports with transparent reasoning steps. Unlike conventional forecast outputs presented as dense tables or unstructured time series, our system performs agent-based analysis across multiple meteorological variables, integrates historical climatological context, and generates structured explanations that identify weather fronts, anomalies, and localized dynamics. The architecture relies entirely on in-context prompting, without fine-tuning, demonstrating that interpretability can be achieved through reasoning rather than parameter updates. Through case studies on multi-location forecast data, we show how AI-Meteorologist not only communicates weather events but also reveals the underlying atmospheric drivers, offering a pathway toward AI systems that augment human meteorological expertise and support scientific discovery in climate analytics.

Code —
https://anonymous.4open.science/r/ai-meteorologist/

1 Introduction

Large Language Models (LLMs) are increasingly being applied in meteorology, showing promise in tasks ranging from climate question-answering to tailored forecast advice (Engelmore and Morgan 1986). However, translating raw meteorological data into actionable insights presents significant challenges. Traditional approaches often neglect to convert numerical variables into interpretable narratives of events and their consequences (Engelmore and Morgan 1986). While existing systems focus on extracting insights from textual corpora or generating brief recommendations, they fail to address the end-to-end processing of structured numerical forecast data with rich contextual reasoning (Peláez-Rodríguez et al. 2024). We identify two main challenges:

Meteorological Knowledge Gap: Most LLMs lack understanding of meteorological contexts, including parameter relationships, atmospheric processes, and regional climatology. Without essential background information, LLMs cannot provide a meaningful analysis of forecast data.

Complexity of Weather Data: Weather forecasts involve multiday, high-frequency predictions across numerous parameters. The LLM must understand temporal dependencies and parameter interactions for accurate interpretation.

In response to these challenges, we propose AI-Meteorologist, a modular agent-based system that translates numerical weather forecasts into structured natural language reports with visualizations. The system employs multiple specialized LLM agents to process structured forecast data, reason about emerging weather patterns, and generate human-readable, explainable reports. By decomposing the task into specialized steps, AI-Meteorologist enables interpretation of detailed forecasts while addressing the scale of the data and the need for trustworthy explanations.

2 Related Work

LLM-agent systems for scientific reasoning. Recent work has demonstrated the emergence of LLM-agent systems as a new paradigm for scientific reasoning. Early systems such as AutoGPT (Significant-Gravitas 2023) and MetaGPT (Hong et al. 2023) introduced autonomous multi-agent collaboration frameworks, while AutoGen (Wu et al. 2024) formalized conversational agent orchestration for complex analytical tasks. In scientific domains, ChemCrow (M. Bran et al. 2024) and AI Scientist (Lu et al. 2024) demonstrated how LLM agents can autonomously design experiments, retrieve literature, and validate hypotheses. These systems collectively illustrate the growing shift from static language models to interactive, tool-augmented scientific agents capable of structured reasoning and knowledge generation.

LLMs in meteorology and weather forecasting. Recent work has begun leveraging LLMs to translate structured meteorological data into human-interpretable weather narratives and interactive tools. For example, ECMWF's DestinE chatbot (ECMWF 2025) to make high-resolution weather and climate data accessible via conversational interfaces, CLLMate (Li et al. 2024) to enable event-based forecasting of weather and climate phenomena in text form, and GPT-based forecasting (Franch et al. 2025) to perform precipitation nowcasting by tokenizing radar imagery and generating ensemble forecasts via language-model driven frameworks. More recent works have explored LLM-agent frameworks

that integrate iterative querying, reflection, and domain-specific validation to improve the scientific accuracy and robustness of weather reports (Varambally et al. 2025).

3 System Description

The system employs a modular agent-based LLM architecture designed for the automated generation of explainable weather forecast reports from structured prediction data. Operating as a pipeline of specialized agents, each component handles the distinct aspects of transforming raw meteorological input into human-readable bulletins. Implementation leverages the GPT-4o API and processes high-resolution time-series data retrieved from the OpenWeather One Call 2.5 API, providing hourly forecasts up to five days ahead.

Data Acquisition and Preprocessing. The pipeline begins with data acquisition for a selected location, retrieving forecast time series via the OpenWeather API. This includes more than ten meteorological parameters, including temperature, humidity, wind, precipitation, and visibility at a resolution of one hour. The system simultaneously gathers contextual information, comprising geographic metadata from OpenStreetMap (such as region type and urban/rural classification) and climatological norms from the Meteostat API, specifically 20-year monthly averages of temperature and precipitation. All collected data are serialized into a structured EXTERNAL INFO block and propagated downstream to the reasoning agent.

Meteorologist Agent. The core analytical component generates a structured output containing four key fields: a natural language summary outlining expected weather evolution; a causal explanation grounded in meteorological reasoning; a self-assessed confidence level; and warnings describing potentially hazardous or anomalous conditions detected through deviation from climatology.

Writer Agent. Building on the meteorologist's output, the Writer agent creates the report's textual structure: a title (location/parameters), an introduction synthesizing outlook with summary, proof, and warnings, and a weather_params block with key variable descriptions. The text emulates official bulletins, optionally incorporating user preferences.

Illustrator Agent. Based on the selected parameters, this agent generates Python code using matplotlib to produce plots of forecast data such as temperature, wind, and precipitation, ensuring that visual representations align with the textual narrative.

Report Compilation and Output. The final report integrates all content blocks into a structured PDF document, comprising a general synopsis, detailed forecast summary and reasoning, visual plots, and optional warnings with climatological comparisons. This comprehensive output provides users with both textual explanations and visual representations of the forecast data.

Sample Prompt: *You are a professional meteorologist tasked with interpreting a time series of weather data for a specific location. Your goal is to write a clear, concise, and scientifically grounded summary of the weather conditions over the given period.*

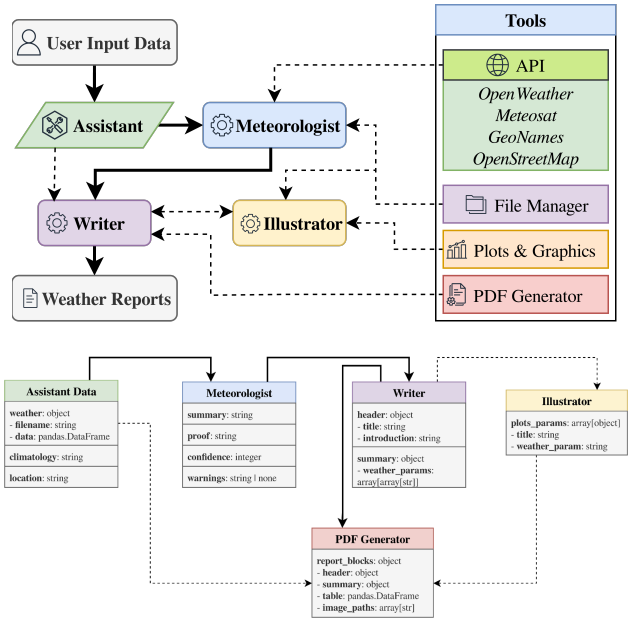


Figure 1: Interaction of individual system modules for generating weather reports. Core workflow is supported by weather APIs, file management, visualization capabilities, and end-to-end report generation.

*You must: - Describe the evolution of key meteorological parameters such as temperature, wind, precipitation, humidity, and pressure over time. - Identify and explain significant *weather phenomena*, including heat waves, cold fronts, thunderstorms, or abrupt changes in conditions. - Highlight temporal *trends*, such as warming, cooling, increasing cloudiness, or shifting wind directions. - Provide *interpretive insights*, not just numerical repetitions — describe what the numbers imply about the atmosphere. - If provided, use climatological normals or geographic context to comment on anomalies (e.g., "unusually cold for this region"). - Avoid simply listing values or copying table entries. Instead, synthesize the data into meaningful meteorological observations. - Use CLIMATOLOGY INFO from input data for describing anomaly events.*

Use formal meteorological language but remain understandable to non-specialists. Structure your output in coherent paragraphs, typically 4–6 sentences, covering the full time period.

*You are NOT generating a forecast. You are summarizing and interpreting **past or present** observed or modeled weather data.*

4 Results

We evaluate the AI-Meteorologist system on real-time multi-parameter forecast data across coastal, inland, and high-altitude locations to assess its ability not only to generate reports, but to conduct interpretable scientific reasoning. Figure 1 illustrates the interaction between the agent modules during inference, and Figure 2 presents an example forecast report automatically generated without fine-tuning.

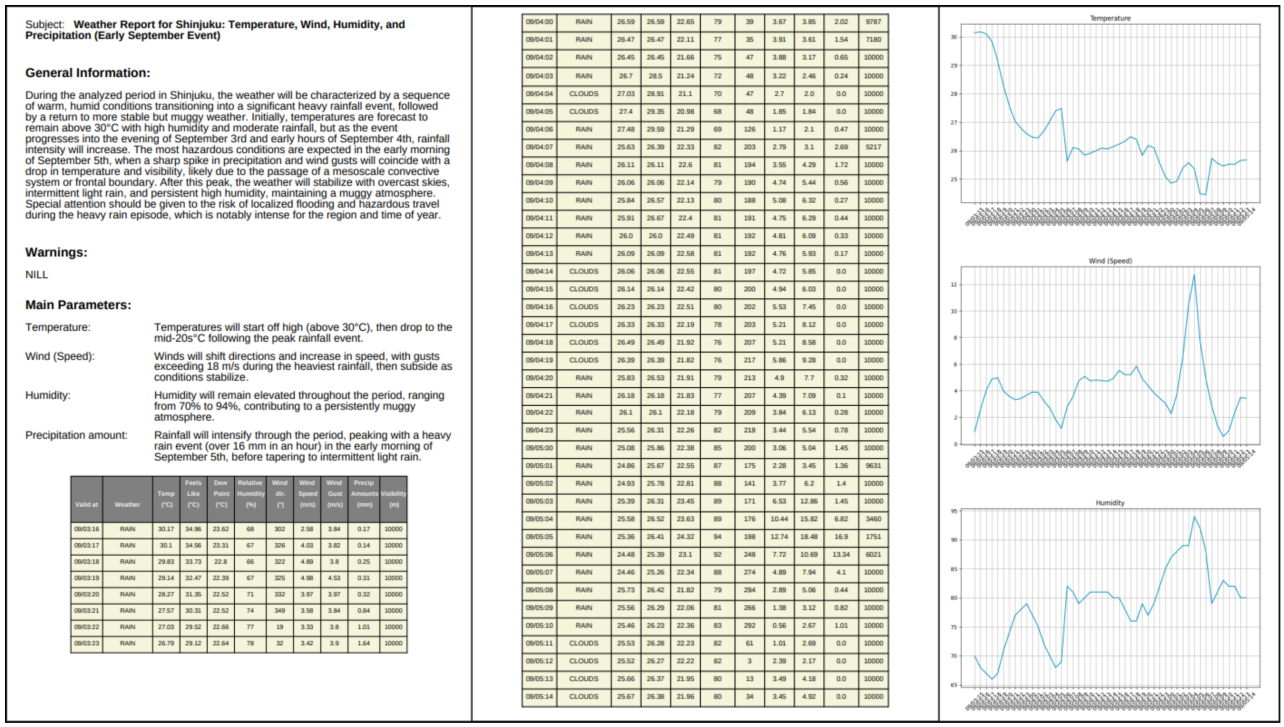


Figure 2: Example of a report generated by the system with meteorological insights, detailed multi-parameter forecasts, and time-series visualizations of key weather variables.

Across all test cases, the system consistently identified key meteorological phenomena through explainable diagnostic reasoning steps. For a coastal location (Synoptic Front Detection) the agent detected an approaching cold front by jointly analyzing pressure gradients, wind direction rotation, and a rapid decrease in temperature. Rather than merely reporting numerical changes, the system generated a scientifically grounded explanation describing how frontal boundaries form due to interactions between maritime and continental air masses. It further contextualized the event using local climatological baselines and automatically generated a confidence score indicating forecast reliability. This demonstrates the system's ability to move beyond narrative reporting toward hypothesis-driven meteorological interpretation. For an inland region (Anomaly and Hazard Reasoning), the system identified anomalous precipitation by comparing short-term forecasts with long-term climatological percentiles. The agent explained the deviation from expected seasonal patterns and issued a data-driven risk statement regarding potential flooding. Importantly, the explanation included its reasoning chain, referencing historical anomalies and presenting the conclusion as a structured argument rather than a black-box output.

These results demonstrate that the AI-Meteorologist is not merely a language interface, but a reasoning-capable scientific assistant. By combining hierarchical interpretation, semantic keyword extraction, and climatological knowledge integration, the system provides transparent and reproducible insights—fulfilling the goals of explainable AI for advancing scientific understanding of weather systems.

5 Discussion

The AI-Meteorologist system demonstrates several key advances in automated weather forecasting and explanation. First, it addresses the critical challenge of translating dense numerical meteorological data into human-interpretable narratives without requiring domain-specific fine-tuning of the underlying LLM. This approach significantly reduces the computational resources and expertise typically needed for specialized AI-weather systems (Zhang et al. 2025).

Second, the system's modular agent-based architecture provides flexibility and extensibility, enabling incremental improvements without disrupting the overall workflow. Third, the integration of multiple data sources enables comprehensive contextual reasoning that references both immediate weather patterns and the long-term climatic context.

However, several limitations warrant consideration. The performance of the system depends on the quality and coverage of the underlying APIs and data sources. Additionally, while the current implementation demonstrates strong capabilities in generating textual explanations, the quantitative evaluation of forecast accuracy and explanation quality remains an area for future work.

Conclusion

We presented AI-Meteorologist, a modular LLM-agent system designed to generate explainable weather forecast reports from structured meteorological data. Our approach integrates time-series reasoning, climatological context, and geospatial metadata to produce high-quality bulletins that go beyond numerical repetition, offering interpretability, causal justifications, and warnings about anomalous conditions.

Our demonstration illustrates that even general-purpose LLMs, such as GPT-4o, can reason over dense hourly forecast tables when provided with properly serialized inputs and structured prompts. The modularity of our architecture allows for extensibility, including multilingual support, user-driven customization of the report structure, and incorporation of additional forecast sources.

Future directions include comparative evaluation of generated reports against human-authored bulletins, integration with probabilistic ensemble data, and exploration of self-refinement loops for improved robustness. By bridging the gap between raw meteorological data and human-interpretable explanations, AI-Meteorologist represents a significant step toward more accessible and trustworthy automated weather forecasting systems.

References

ECMWF. 2025. Development Seed to create a climate and weather chatbot in DestinE. "News" article on the Destination Earth website. https://destine.ecmwf.int/news/development-seed-to-create-a-climate-and-weather-chatbot-in-destine/?utm_source=chatgpt.com (accessed: 2025-10-22).

Engelmore, R.; and Morgan, A., eds. 1986. *Blackboard Systems*. Reading, Mass.: Addison-Wesley.

Franch, G.; Tomasi, E.; Wanjari, R.; Poli, V.; Cardinali, C.; Alberoni, P. P.; and Cristoforetti, M. 2025. GPTCast: a weather language model for precipitation nowcasting. *Geoscientific Model Development*, 18(16): 5351–5371.

Hong, S.; Zhuge, M.; Chen, J.; Zheng, X.; Cheng, Y.; Wang, J.; Zhang, C.; Wang, Z.; Yau, S. K. S.; Lin, Z.; et al. 2023. MetaGPT: Meta programming for a multi-agent collaborative framework. In *The Twelfth International Conference on Learning Representations*.

Li, H.; Wang, Z.; Wang, J.; Wang, Y.; Lau, A. K. H.; and Qu, H. 2024. CLLMate: A Multimodal Benchmark for Weather and Climate Events Forecasting. *arXiv preprint arXiv:2409.19058*.

Lu, C.; Lu, C.; Lange, R. T.; Foerster, J.; Clune, J.; and Ha, D. 2024. The ai scientist: Towards fully automated open-ended scientific discovery. *arXiv preprint arXiv:2408.06292*.

M. Bran, A.; Cox, S.; Schilter, O.; Baldassari, C.; White, A. D.; and Schwaller, P. 2024. Augmenting large language models with chemistry tools. *Nature Machine Intelligence*, 6(5): 525–535.

Peláez-Rodríguez, C.; Pérez-Aracil, J.; Marina, C. M.; Prieto-Godino, L.; Casanova-Mateo, C.; Gutiérrez, P. A.; and Salcedo-Sanz, S. 2024. A general explicable forecasting framework for weather events based on ordinal classification and inductive rules combined with fuzzy logic. *Knowl. Based Syst.*, 291: 111556.

Significant-Gravitas. 2023. AutoGPT. https://github.com/Significant-Gravitas/AutoGPT. Version v0.6.33, accessed: 2025-10-22.

Varambally, S.; Fisher, M.; Thakker, J.; Chen, Y.; Xia, Z.; Jafari, Y.; Niu, R.; Jain, M.; Manivannan, V. V.; Novack, Z.; et al. 2025. Zephyrus: An Agentic Framework for Weather Science. *arXiv preprint arXiv:2510.04017*.

Wu, Q.; Bansal, G.; Zhang, J.; Wu, Y.; Li, B.; Zhu, E.; Jiang, L.; Zhang, X.; Zhang, S.; Liu, J.; et al. 2024. Autogen: Enabling next-gen LLM applications via multi-agent conversations. In *First Conference on Language Modeling*.

Zhang, J.; Clairmont, C.; Que, X.; Li, W.; Chen, W.; Li, C.; and Ma, X. 2025. Streamlining geoscience data analysis with an LLM-driven workflow. *Applied Computing and Geosciences*, 25: 100218.

Supplementary Material

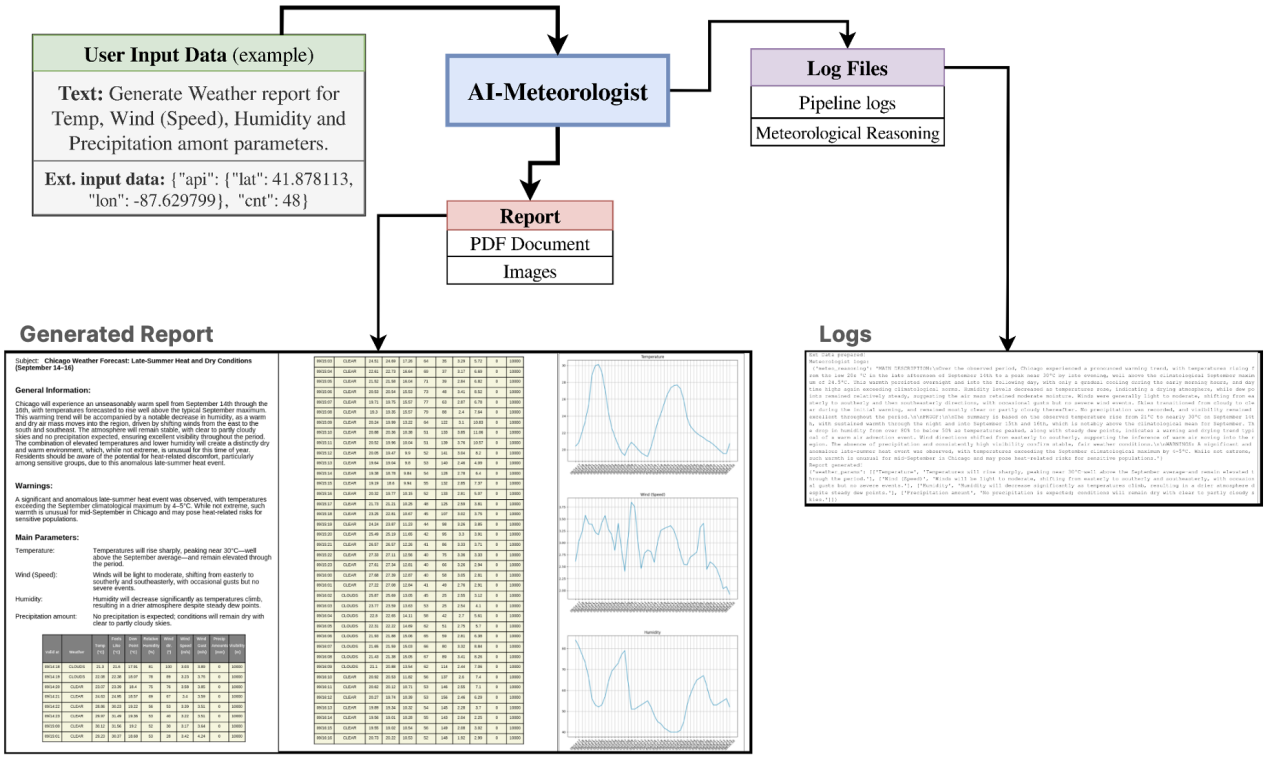


Figure 3: Schematic overview of the weather report generation process.