SmartPilot:Agent-Based CoPilot for Intelligent Manufacturing

Demonstration Track

Chathurangi Shyalika
Artificial Intelligence Institute,
University of South Carolina
Columbia, SC, USA
jayakodc@email.sc.edu

Renjith Prasad
Artificial Intelligence Institute,
University of South Carolina
Columbia, SC, USA
kaippilr@mailbox.sc.edu

Alaa Al Ghazo
Artificial Intelligence Institute,
University of South Carolina
Columbia, SC, USA
alaaalghazo@gmail.com

Darssan L. Eswaramoorthi
Artificial Intelligence Institute,
University of South Carolina
Columbia, SC, USA
darssan@email.sc.edu

Sara Shree Muthuselvam
Artificial Intelligence Institute,
University of South Carolina
Columbia, SC, USA
muthuses@email.sc.edu

Amit Sheth
Artificial Intelligence Institute,
University of South Carolina
Columbia, SC, USA
amit@sc.edu

ABSTRACT

In the dynamic landscape of Industry 4.0, achieving efficiency, precision, and adaptability is essential for optimizing manufacturing operations. SmartPilot is a neurosymbolic and agent-based CoPilot designed to enhance real-time decision-making capabilities in manufacturing. The system addresses three key challenges: anomaly prediction, production forecasting, and domain-specific question answering through an agent-based framework. SmartPilot leverages multimodal data and a compact architecture optimized for edge devices. This paper highlights its innovative combination of agent-based design and neurosymbolic reasoning to enable contextual decision-making in complex environments. The demonstration video[1], datasets, and supplementary materials are available at https://github.com/ChathurangiShyalika/SmartPilot.

KEYWORDS

Smart Manufacturing, CoPilot, Agent-Based Architecture, Multimodality

ACM Reference Format:
Chathurangi Shyalika, Renjith Prasad, Alaa Al Ghazo, Darssan L. Eswaramoorthi, Sara Shree Muthuselvam, and Amit Sheth. 2025. SmartPilot:Agent-Based CoPilot for Intelligent Manufacturing: Demonstration Track. In *Proc. of the 24th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2025), Detroit, Michigan, USA, May 19 – 23, 2025*, IFAAMAS, 3 pages.

1 INTRODUCTION

Manufacturing processes in the Industry 4.0 era rely heavily on data-driven technologies for decision making. However, challenges such as predictive analytics, supply chain disruptions, and inventory discrepancies hinder operational efficiency [3, 6]. Although foundational models (FMs) such as large language models (LLMs) have demonstrated success in the text and image domains, their generalization to sensor data is restricted by challenges such as heterogeneous datasets, data leakage, and the scarcity of labeled data. Hence, FMs focused on time series such as Time-MOE [4] and MOIRAI [7] often fail to generalize in manufacturing-specific tasks, preventing their practical utility [5]. The rise of smaller and more efficient models [2] offers an opportunity for a CoPilot system tailored for manufacturing, but such systems are currently unavailable in the industry. SmartPilot is designed to address industry-specific challenges by serving as a neurosymbolic and agent-based CoPilot specifically tailored for manufacturing. It focuses on three critical tasks: anomaly prediction, production forecasting, and domain-specific question answering. The system incorporates three specialized agents, each dedicated to one of these key functions. *PredictX* agent is responsible for anomaly prediction and focuses on identifying irregularities in operational data to prevent disruptions. *Foresight* agent forecasts next-hour production and anticipates future production requirements to optimize resources. *InfoGuide* agent facilitates real-time information retrieval tailored to manufacturing-specific questions. All three agents within SmartPilot are designed with lightweight models optimized for deployment on edge devices. This design ensures that individual agents can perform real-time operations effectively in resource-constrained environments. The system incorporates neurosymbolic techniques, blending statistical methods with symbolic reasoning. Multimodal data serves as input for the statistical methods, while manufacturing ontologies provide the basis for symbolic reasoning. This gives SmartPilot the ability to provide precise and contextually aware decision-making capabilities. SmartPilot is currently deployed in a rocket assembly use case, where it predicts anomalies caused by the absence of specific rocket components, forecasts the rockets produced, and assists with answering domain-specific questions. Further details are provided in the video demo.

2 DESIGN AND IMPLEMENTATION

2.0.1 Core Architecture: Figure 1 illustrates the core system framework, highlighting the key technical components of SmartPilot. Key attributes of the system encompass the agent-based system and multimodal data integration.

2.0.2 Agent-Based Implementation: The agent-based component in Figure 1 highlights the agents and the interactions between them. In ***PredictX***, we introduce a decision-level fusion approach for multimodal anomaly prediction in assembly pipelines, currently

[1] Video URL: https://tinyurl.com/2hurd6nd

Proc. of the 24th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2025), Y. Vorobeychik, S. Das, A. Nowé (eds.), May 19 – 23, 2025, Detroit, Michigan, USA. © 2025 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org).

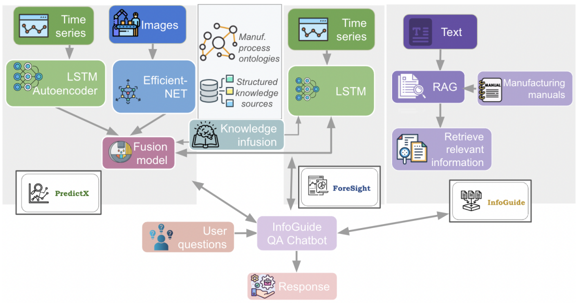

Figure 1: Core system framework that includes the technical components of SmartPilot

trained on multimodal data in a rocket assembly process [1]. It integrates time-series data, processed via an autoencoder, with image data analyzed using a fine-tuned EfficientNet and combines the outputs of these models through a fully connected network. We also use manufacturing-based process ontologies[2] to infuse domain-specific knowledge (about sensor ranges, cycle states, robots, and machinery status) in the model training process and to provide user-level explanations for the predictions. ***ForeSight*** utilizes a Long Short-Term Memory (LSTM)-based model for production forecasting, using 30-timestep historical data from target variables as inputs. The architecture includes two LSTM layers (100 units each) to capture temporal dependencies, with additional features on sensor ranges and robots rotating angles infused at the final dense layer. It outputs predictions for the target variables, trained with the Adam optimizer and mean squared error. This approach ensures accurate, context-aware forecasting tailored to complex production processes. ***InfoGuide*** leverages meaningful texts derived from manufacturing manuals to provide answers to contextually rich manufacturing-specific questions. When a user submits a query, the system uses retrieval-augmented generation (RAG) to find the top k relevant contexts based on similarity measures: Neural context retrieval (using cosine similarity with BERT embeddings) and Symbolic context retrieval (using Jaccard similarity with tokenized keywords). If the similarity score exceeds a set threshold, the user query, agent-specific template and retrieved context are sent to the Mixtral language model to generate a coherent response. If the score falls below the threshold, the system refrains from providing an answer, effectively mitigating the risk of generating hallucinated responses.

2.0.3 Inter-Agent Connectivity: PredictX feeds anomaly-related insights into ForeSight, allowing the forecasting model to adjust with the evolving production system. This interconnected approach enables real-time, dynamic responses to user queries, ensuring both anomaly prediction and production forecasting are aligned with changing operational conditions. The model is further integrated with InfoGuide for effective information retrieval, enhancing the system's ability to provide actionable insights. To handle the interconnection between the InfoGuide and the other two agents, we fine-tune the DistilBERT language model on the outputs of PredictX and Foresight, employing Low-Rank Adaptation (LoRA). The fine-tuned model is integrated with InfoGuide, enabling effective responses to user queries about anomaly prediction and production forecasting through real-time information retrieval.

[2]https://github.com/revathyramanan/Dynamic-Process-Ontology

3 DEMONSTRATION

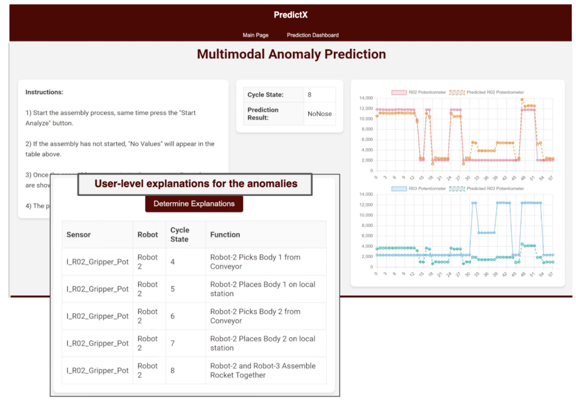

Figure 2: Real-time anomaly prediction and user-level explanations given by PredictX agent, with a similar design applied to ForeSight agent

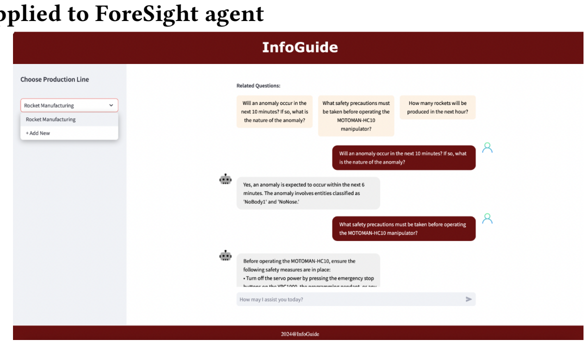

Figure 3: User interface of InfoGuide: Providing real-time responses to user queries

3.0.1 User Interfaces and Interactivity: Upon accessing SmartPilot, the main interface provides three buttons to access the three agents. PredictX offers two options: the prediction dashboard for real-time anomaly prediction and the explanation dashboard for detailed insights into the predictions, as shown in Figure 2. ForecastX interface features the forecasting dashboard, enabling users to monitor and forecast future production. InfoGuide interface serves as a real-time question-and-answer chatbot, allowing users to input domain-specific questions and receive immediate responses as shown in Figure 3.

3.0.2 Real-time Integration with Backend Systems: The prediction and forecasting dashboards in PredictX and ForecastX enable them to connect with an OPC UA server, allowing them to visualize real-time sensor data for anomaly prediction and production forecasting, respectively. InfoGuide is similarly connected to the OPC UA server, utilizing real-time data to provide accurate responses to queries related to anomaly prediction and production forecasting.

3.0.3 User-level Explanation of Predictions: Once anomalies are predicted, the second window of PredictX provides user-level explanations, including details on (i) which variables are responsible for the anomaly, (ii) the functions performed by the robots during that state and (iii) the expected values of those variables.

ACKNOWLEDGMENTS

This work is supported in part by NSF grant #2119654, “RII Track 2 FEC: Enabling Factory to Factory (F2F) Networking for Future Manufacturing”.

REFERENCES

[1] Ramy Harik, Fadi El Kalach, Jad Samaha, Devon Clark, Drew Sander, Philip Samaha, Liam Burns, Ibrahim Yousif, Victor Gadow, Theodros Tarekegne, et al. 2024. Analog and Multi-modal Manufacturing Datasets Acquired on the Future Factories Platform. *arXiv preprint arXiv:2401.15544* (2024).

[2] Shengding Hu, Yuge Tu, Xu Han, Chaoqun He, Ganqu Cui, Xiang Long, Zhi Zheng, Yewei Fang, Yuxiang Huang, Weilin Zhao, et al. 2024. Minicpm: Unveiling the potential of small language models with scalable training strategies. *arXiv preprint arXiv:2404.06395* (2024).

[3] Michael Nieberl, Alexander Zeiser, and Holger Timinger. 2024. A Review of Data-Centric Artificial Intelligence (DCAI) and its Impact on manufacturing Industry: Challenges, Limitations, and Future Directions. In *2024 IEEE Conference on Artificial Intelligence (CAI)*. IEEE, 44–51.

[4] Xiaoming Shi, Shiyu Wang, Yuqi Nie, Dianqi Li, Zhou Ye, Qingsong Wen, and Ming Jin. 2024. Time-MoE: Billion-Scale Time Series Foundation Models with Mixture of Experts. *arXiv preprint arXiv:2409.16040* (2024).

[5] Chathurangi Shyalika, Harleen Kaur Bagga, Ahan Bhatt, Renjith Prasad, Alaa Al Ghazo, and Amit Sheth. 2024. Time Series Foundational Models: Their Role in Anomaly Detection and Prediction. *arXiv preprint arXiv:2412.19286* (2024).

[6] Mahipal Singh, Rekha Goyat, and Renu Panwar. 2024. Fundamental pillars for industry 4.0 development: implementation framework and challenges in manufacturing environment. *The TQM Journal* 36, 1 (2024), 288–309.

[7] Gerald Woo, Chenghao Liu, Akshat Kumar, Caiming Xiong, Silvio Savarese, and Doyen Sahoo. 2024. Unified Training of Universal Time Series Forecasting Transformers. *arXiv preprint arXiv:2402.02592* (2024).