Human Interaction and Emerging Technologies (IHIET 2024), Vol. 157, 2024, 196–206
https://doi.org/10.54941/ahfe1005478

An AI-Driven User-Centric Framework Reinforced by Autonomic Computing: A Case Study in the Aluminium Sector

Ramon Angosto Artigues[1], Andrea Gregores Coto[1], Jonathan Josue Torrez Herrera[2], Fernando Lou Tomás[2], Sabrina Verardi[3], Mattia Giuseppe Marzano[3], and Andrea Fernández Martínez[1]

[1]AIMEN Technology Centre, 36418, O Porriño (Pontevedra), Spain
[2]Ibérica de Aleaciones Ligeras (IDALSA), 50668, Remolinos (Zaragoza), Spain
[3]Engineering Ingegneria Informatica S.p.A, 00144 - Roma (RM) - Italy

ABSTRACT

The integration and deployment of AI in the industry faces several challenges, involving not only the need for robust and accurate AI models, but also their seamless integration with existing systems, while ensuring an intuitive user experience for workers. Furthermore, it is critical for AI solutions to be continuously managed for data governance, performance optimization, and the mitigation of risks, among other factors. This paper presents a service-oriented application that explores the integration of Machine Learning algorithms by adopting Human-in-the-Loop (HITL) strategies to enhance user-technology interactions in an Aluminium industrial environment. The proposed application exploits the use of data-driven Autonomic Computing techniques in AI Data Pipelines to promote the development of self-managed, adaptive systems that support dynamic interactions between technology and workers. Through the implementation of a web interface, workers are provided with seamless access to real-time data analysis and intelligent solutions within the user-empowered application.

Keywords: Artificial intelligence, Autonomic computing, AI data pipeline, Human-centric design, Human-in-the-loop (HITL)

INTRODUCTION

In today's rapidly evolving industrial landscape, the integration of Artificial Intelligence (AI) in the industry represents a paradigm shift, offering unprecedented opportunities to enhance efficiency, reduce environmental impact, and streamline operations, what is essential to maintain EU Industry competitiveness and sustainability (Mhlanga, 2022). Nevertheless, the complexities of modern industrial processes necessitate robust, flexible AI solutions that can be seamlessly integrated into existing infrastructures without disrupting current production workflows, while responding to the dynamics of the process, particularly under uncertain scenarios. To guarantee the successful adoption of AI technologies in industrial settings, it is essential to keep the human in control and at the centre of all developments, actively involving workers in the design, implementation, and refinement of the technology. Indeed, AI systems have shown promising potential in supporting workers in their daily operational activities, enhancing their skills and knowledge of the process, and enabling faster-informed decisions (Rožanec et al., 2023), but their engagement with the technology is key to fully exploit the AI capabilities.

This paper presents a novel AI-driven application through the implementation of an AI Data Pipeline with Autonomic Computing capabilities to dynamically adapt to the needs of the industry, facilitating human-machine interactions and technology management, shown in Figure 1. The proposed user-centric application is integrated and evaluated in the Aluminium sector to demonstrate its ability to create a smart, human-centric production environment. By incorporating Human-In-The-Loop (HITL) strategies in a user-friendly interface, the proposed application enables a collaborative environment between humans and technology, ensuring that AI implementations are both effective and ethically sound.

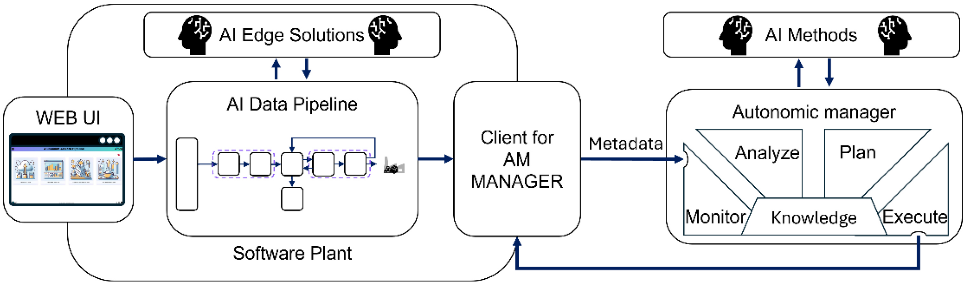

Figure 1: Diagram of the AI framework linking a data pipeline with an autonomic manager (AM) for industrial efficiency. The AM, based on the MAPE-K[1] model, leverages metadata to enhance autonomy.

THE AI DATA PIPELINE

The integration of AI in industrial processes requires of methodical strategies to manage the flow of data through various computational tasks. Some of these tasks include the ingestion and transformation of data, as well as real-time data analysis. AI Data Pipelines are essential to ensure that data moves efficiently and reliably across different stages, transforming raw data into actionable information. Furthermore, this type of infrastructure must ensure the proper management and processing of datasets from different sources to enable AI applications managing industry factors—both internal and external—effectively (Krismentari et al., 2022). In this paper, we present an innovative AI Data Pipeline consisting of five main components that efficiently manages the end-to-end process of collecting, processing, and analysing data to train and deploy AI models effectively in industrial

[1]MAPE-K: (Monitor-Analyze-Plan-Execute-Knowledge)

scenarios. The pipeline has been designed and engineered with the aim of improving operational agility and performance in the industry. The AI Data Pipeline and the interrelations among components are depicted in Figure 2. The five components of the pipeline are described in more detail in the following subsections.

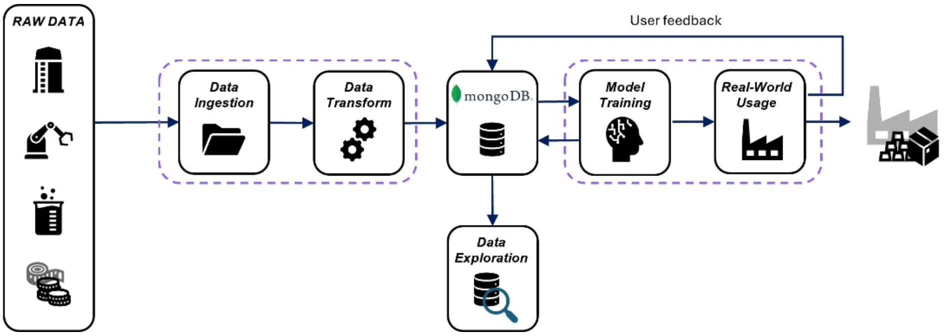

Figure 2: Workflow diagram illustrating the process from raw data to application, detailing components of pipeline: ingestion, transformation, MongoDB exploration, model training, and deployment.

Data Ingestion: The initial stage of the AI Data Pipeline involves collecting data from various sources, such as sensors, external databases, and/or third-party platforms and applications. At this stage, raw data is read and structured as necessary to prepare it for subsequent processing, ensuring its accessibility in a usable data format or schema. In industrial environments, customized ingestion procedures that cover the full spectrum of data might be required to handle the various data sources and formats.

Data Transformation: This component processes the ingested data, applying transformations such as normalization, filtering, and feature extraction to make the data suitable for data analysis. It is in this stage where data is merged and prepared to be stored in a proper data storage and management system e.g., MySQL, or MongoDB. Techniques to handle missing or incomplete data are implemented by this component as well e.g., inference and statistical methods from historical data.

Data Exploration: Data visualization and analysis are vital to enhance human knowledge of the process and its parameters, as well as elucidate intrinsic data patterns and analyze its evolution over time. This component can provide data insights in form of tables, charts, and/or pre-defined data analysis techniques e.g., via descriptive statistics and statistical inference methods.

Model Training: In this stage, ML models are trained, fine-tuned, and evaluated using the processed datasets from the Data Ingestion & Transformation components. Some of the functions comprised by the Model Training component includes the selection of appropriate ML algorithms according to the data nature and the problem addressed, the selection and fine-tuning of hyperparameters and architecture of the ML models, and the evaluation of their performance to ensure that the specified criteria is accurately met.

Real-World Usage (RWU): Trained ML models can be deployed and integrated into production environments through various methods and platforms, such as cloud platforms, containers (Docker), or edge computing. In addition, monitoring ML models is essential to detect performance degradation or potential failures, leading to necessary re-training or model updates by the previous Model Training component. The RWU component often incorporates *feedback mechanisms* that enable continuous improvement for latest data inputs, as well as the integration of HITL strategies to further enhance the on-going refinement of AI models.

A HUMAN-CENTRIC SERVICE-ORIENTED APPLICATION

For AI Data Pipelines to enable organizations extract insights and value from their data effectively, a seamless deployment and integration of the pipeline with the AI solutions is essential. A Service-Oriented Architecture (SOA) is an architectural approach wherein distinct, reusable software components, referred as services, communicate with each other over a network to achieve specific business functionalities (Schall et al., 2008). This architectural framework can facilitate the integration of AI Data Pipelines by presenting its various components as modular, reusable services. By following this approach, the system achieves loose coupling, thereby enhancing the flexibility and agility of the pipeline by facilitating independent development, deployment, and update of the individual components. The integration of the proposed AI Data Pipeline based on the SOA framework using APIs to expose services' functionalities provides a robust and scalable architecture that promotes modularity, interoperability, and the enforcement of security and monitoring mechanisms in the system. These characteristics are crucial to enable more effective HITL interactions through the simplification of complex processes into more manageable services. Moreover, this framework also enables the smooth integration of feedback mechanisms across the various services exposed, fostering continuous improvement and personalized user experiences.

In fact, HITL integration is imperative for any data-driven application to achieve full usability and functionality in real-world industrial settings, ensuring complete symbiosis between humans and Cyber-Physical Systems (CPS) (Adel, 2022). However, the transition towards the human-centric smart factory concept introduces challenges such as a prevalent skills gap where workers might lack the necessary IT expertise to effectively manage advanced systems (Tan et al., 2019). To this end, Autonomic Computing techniques can play a pivotal role closing this gap, providing technology with self-managing capabilities to mitigate the increasing complexity of computing systems (Gil et al., 2019). Additionally, the cognitive load imposed by evaluating large datasets, and the need for rapid decision-making under uncertain circumstances, can overwhelm workers without well-designed interfaces or adequate support tools. Hence, it is vital to prioritize human-centric strategies through the design and development of the technology to ensure that it is not only intuitive and easy to use, but it is also tailored to the needs and preferences of users to foster their engagement, satisfaction, and productivity. Section *The AI-Driven Web Application In The Aluminium Use Case* provides more information about the proposed user-friendly web application.

AUTONOMIC COMPUTING FOR SELF-REGULATED AI SYSTEMS

Autonomic Computing (AC) is a computing paradigm initiated by IBM in 2001 and originally inspired by the human autonomic nervous system that aims to develop computing systems with self-managing characteristics. The ultimate goal of this paradigm is to reduce human intervention in the rapidly growing complexity of software systems management, freeing humans from low-level management tasks, while still maintaining their central role in providing high-level guidance for their self-management (Parashar & Hariri, 2005).

A common framework to achieve self-management in AC is the MAPE (Monitor, Analyze, Plan, Execute) loop, which intends to provide systems with the necessary abilities to autonomously adapt to changing conditions, recover from failures, and optimize their performance. The Monitor module is responsible for continuously monitoring the environment and collecting relevant data about the system's state. This data is then processed, analyzed, and interpreted by the Analyze module to identify potential deviations or anomalies, as well as opportunities for the optimization of the system. Subsequently, based on the previous data analysis, strategies to maintain, adjust, and/or improve the system behaviour and attributes are determined by the Plan module. Ultimately, the planned actions are implemented as automated responses by the Execute module. The MAPE loop can be extended to the MAPE-K framework with the addition of the Knowledge component, which incorporates knowledge-driven reasoning and decision-making capabilities to the system to operate more intelligently and effectively. By incorporating the MAPE-K framework, the four main properties of AC systems can be realized as follows (Vizcarrondo et al., 2017).

Self-configuration: AC systems can automatically configure themselves based on their environment and requirements to adapt to changing conditions. For instance, AI algorithms can fine-tune their hyperparameters and model architectures based on the nature and characteristics of data, as well as the task to be addressed.

Self-optimization: AI systems can continuously monitor their performance and resource usage, adjusting their parameters to improve their efficiency. Optimization methods such as Bayesian optimization or evolutionary algorithms can be leveraged to fine-tune models. Other algorithms, such as in Reinforcement Learning, can directly learn, adapt, and optimize their strategies from their experiences in the environment. Ultimately, techniques such as re-training (continuous training) can be applied to existing models to incorporate information from new data without affecting the model architecture and parameters.

Self-Healing: Autonomic AI systems can be designed to automatically detect and diagnose failures and errors in real-time to execute corrective actions and restore the system functionality. Anomaly detection algorithms can detect unusual behaviour and data patterns by monitoring the system status. These algorithms cover statistical methods (e.g., z-score), ML methods like isolation Forest and one-class SVM, density methods (e.g., LOF), and state-of-the-art time series anomaly detection algorithms like Seasonal Hybrid ESD and Prophet.

Self-Protection: To protect systems from security threats and attacks, mechanisms to automatically mitigate risks and vulnerabilities can be implemented in AC systems. For this purpose, AI-powered cybersecurity systems and techniques such as adversarial training or anomaly detection algorithms can be deployed to detect, prevent, adapt, and recover from adversarial attacks. Regarding the performance of AI systems, adversarial training can also be used to improve AI robustness by preventing algorithms of being deceived by fake or perturbed data.

METADATA-DRIVEN AUTONOMIC MANAGER FOR AUTONOMIC COMPUTING CAPABILITIES IN THE AI DATA PIPELINE

To effectively adopt the MAPE-K framework and implement the autonomous (self) abilities previously described -self-configuration, self-optimization, self-healing, and self-protection- the Autonomic Manager (AM) serves as the brain of the AC system. The AM can perform a variety of activities, including enforcing policies that govern the behaviour and operation of the AC system, monitoring and analysing the system to detect anomalies or opportunities for improvement, engaging in decision-making and real-time problem-solving processes, and consolidating learning strategies to enable continuous improvement and adaptation. In complex distributed systems, the AM can also coordinate and collaborate with other system components to exchange information, or coordinate actions (Vizcarrondo et al., 2017).

In the proposed framework, the AM assumes a central role serving as an autonomous AI Data Pipeline coordinator and decision-maker, thereby guaranteeing the AI Data Pipeline Governance. In this way, a continuous MAPE flow based on Knowledge of the AI Data Pipeline is implemented. The AM is responsible for continuously monitoring and analysing the state of the AI Data Pipeline to plan and execute corrective and optimization actions via the triggering of the autonomous (self) abilities. Thus, the implementation of the abilities is executed directly on the AI Data Pipeline for its self-management upon notification from the AM triggers.

As part of the practical implementation of our framework, the Monitoring module of the AM interacts with the AI Data Pipeline through an Orion Context Broker via APIs. The Analyze and Plan modules integrate a rule-based engine whose targets and goals are set by operators considering process requirements and specifications. Ultimately, the Execution module interacts with the AI Data Pipeline by triggering the autonomous (self) abilities when needed, as well as by sending notifications and alarms when abnormal behaviour is detected. To complement the rule-engine, AI methods based on temporal series analysis combined with statistical measures are also available for anomalies detection. The temporal analysis considers weekly- and monthly- time windows of data for analysing the evolution of specific parameters over time. Figure 3 depicts the communication flow between the different elements of our framework.

For a proper analysis, planning, and execution of corrective and optimization actions, it is crucial to properly select the data that is shared from the AI Data Pipeline to the AM. One of the key innovations of the proposed framework is the use of metadata from the components of the AI Data Pipeline to better abstract their behaviour and performance.

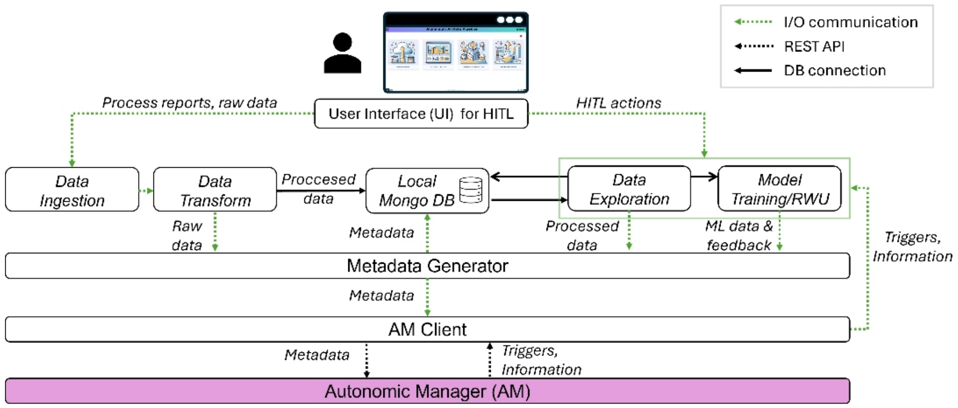

Figure 3: Architecture of the HITL system for autonomic computing, illustrating the flow from data ingestion to model training and deployment, with a local MongoDB for storage and exploration.

By providing metadata the AM can get additional context about the performance of the underlying components, enabling a higher-level conceptualization and understanding. Metadata is based on descriptive statistics of new raw data. Each component of the pipeline can share its own set of metadata with the AM, hence the distributed nature of the framework. As an example, the Data Exploration component might send the average and standard deviation of parameters, e.g., gas and oxygen consumption, whereas the Model Training component might send the errors obtained during training e.g., MAE. Furthermore, to guarantee that users are at the centre of all developments, human feedback is sent to the AM as part of the data. The rule-engine of the AM includes rules assessing the conformity of humans with the established AI solutions, serving as triggering events that activate the autonomous (self) abilities on these solutions.

THE AI-DRIVEN WEB APPLICATION IN THE ALUMINIUM USE CASE

Although the proposed framework based on the AI Data Pipeline with Autonomic Computing abilities has been designed and engineered as a technology-agnostic, domain-independent solution, a preliminary integration in a real (recycled-based) aluminium industrial setting has been performed as a validation of functionality and assessment of performance to enable its continuous improvement.

The recycled aluminium-making process starts with the reception of materials from different sources. Upon reception, materials are sampled for proper sorting and storage based on their chemical composition and characteristics. Based on customer orders on products, which are characterized by the product norms and their quantities, scraps are selected to be processed via primary and secondary melting. This selection is based on the chemical composition of scraps, their availability at the plant, and their cost. In secondary melting, alloys can be introduced to refine the melted mixture prior its moulding and cooling to produce the final aluminium products. The different stages of the process are depicted in Figure 4.

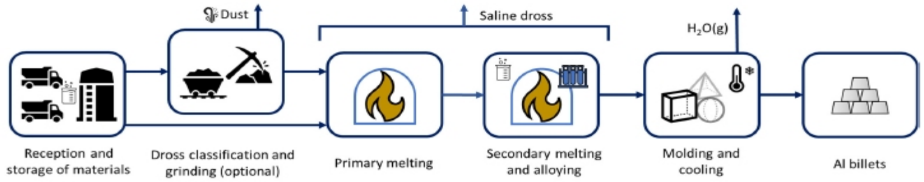

Figure 4: Schematic diagram of the recycled aluminium production process: from material reception to final product formation.

The AI solutions developed within the project aim to support operators in the decision-making process of scraps and alloys (aluminium recipe) via a Random Forest Regressor for the estimation of the chemical composition of aluminium mixtures, and a Generative Reinforcement Learning Framework to generate new recipes based on customer requests as a multi-criteria optimization problem. The details of the AI solutions are out of the scope of this paper.

To facilitate the interaction between users and the AI Data Pipeline and the Autonomic Manager, a web-based interface has been developed following a human-centric design strategy. By incorporating the targeted users in the design and development phases of the interface, the application was unequivocally built around the users' needs, preferences, and behaviours. There are two components of the AI Data Pipeline that hold special significance in terms of HITL interactions: Data Exploration and RWU components. The Data Exploration component can be accessed through two different web pages according to the user preferences. On one hand, users can access data in form of tables, filtering the available datasets according to their origin, heat process code, dates, and number of samples shown. In addition, tables can be downloaded in*.csv* or*.xlsx* format. On the other hand, users can visualize the evolution of a set of pre-defined parameters through charts. Currently, these charts also display statistical measures of the data shown for visualization e.g., mean and standard deviation. Figure 5 shows the two discussed pages of the web application for the Data Exploration component.

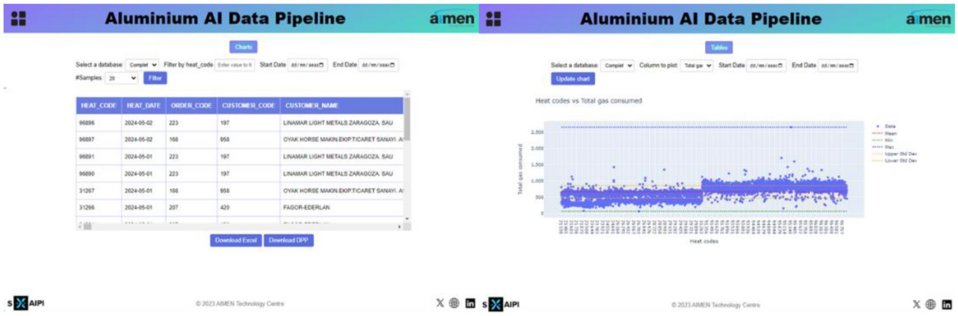

Figure 5: Data exploration interfaces in the web application: tabular and graphical representations for comprehensive analysis of industrial aluminium production parameters.

The RWU component can also be accessed through two different pages according to the AI solution to be invoked, as shown in Figure 6. To collect human feedback on the performance of those solutions, different strategies have been implemented. In the case of the AI solution that estimates the chemical composition of scrap-based mixtures, a thumbs-up/thumbs-down button is available for acquiring the overall satisfaction of users with the predictions. For the recipe generation algorithm, a multi-criteria evaluation via a 5-star scoring is integrated to allow users assess the recipes according to the final chemical composition estimated, the scraps used, and the overall cost. As previously described, human feedback is used as a satisfaction score that can trigger the activation of the self-abilities through the rule-engine of the AM to optimize the AI solutions in the pipeline.

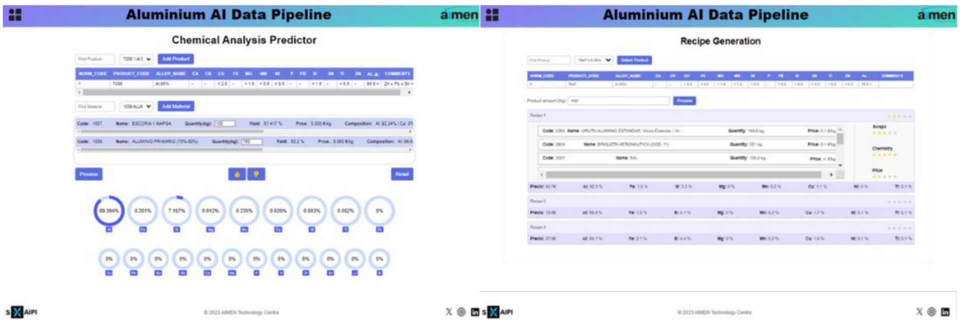

Figure 6: Interactive AI solutions in aluminium production: incorporating user feedback for chemical composition prediction and recipe generation.

The front-end technologies for the web interface development are HTML, CSS, and JavaScript. For the back-end framework, Django is used as the backbone that handles user authentication, database management (MongoDB), and web-server logic. To isolate potential problems with the AI solutions, a FastAPI server has been deployed, which is also responsible for handling communications with the Orion Context Broker of the AM. The complete application is deployed on a private server configured with Docker compose to guarantee that all software dependencies and requirements are met. The presented set-up, depicted in Figure 7, ensures a seamless interaction and easy deployment of all the elements of the application, enabling a smooth communication with the client side and real-time data processing and analysis.

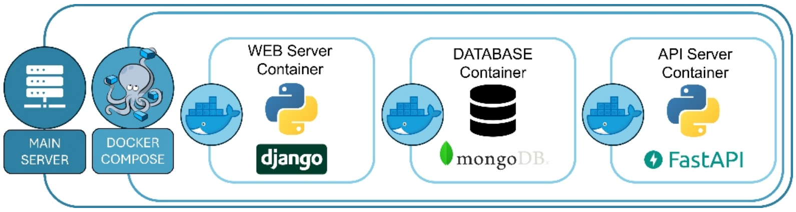

Figure 7: Dockerized web application architecture: integrating Django, MongoDB, and FastAPI for efficient data management and real-time processing.

CONCLUSION

This paper presents a novel AI-driven application embedding an AI Data Pipeline with Autonomic Computing capabilities to strategically integrate advanced AI technologies in real industrial settings. By employing a SOA that facilitates the modular integration of AI components, the various services can be developed and deployed independently, enhancing the flexibility and scalability of the system. This approach also enables the integration of human feedback loops to refine the AI functionalities, guaranteeing the system reliability and AI robustness. By facilitating an intuitive interface for workers following a user-centric design, the system simplifies human-machine interactions leveraging real-time data analysis for better decision-making support. The preliminary integration of the proposed framework through a web application in the Aluminium sector has proven the system to be robust and capable of real-time adaptation and optimization.

ACKNOWLEDGMENT

This work has been supported by the project “self-X Artificial Intelligence for European Process Industry digital transformation” (s-X-AIPI), which has received funding from the European Union’s Horizon Europe research and innovation programme under grant agreement No. 101058715.

REFERENCES

Adel, A. (2022). Future of industry 5.0 in society: human-centric solutions, challenges and prospective research areas. *Journal of Cloud Computing 2022 11:1*, *11*(1), 1–15. https://doi.org/10.1186/S13677-022-00314-5

Gil, M., Albert, M., Fons, J., & Pelechano, V. (2019). Designing human-in-the-loop autonomous Cyber-Physical Systems. *Int. J. Hum. Comput. Stud.*, *130*, 21–39. https://doi.org/10.1016/J. IJHCS.2019.04.006

Krismentari, et al (2022). Data Pipeline Framework for AIS Data Processing. *2022 7th International Conference on Informatics and Computing, ICIC 2022*. https://doi.org/10.1109/ICIC56845.2022.10006941

Mhlanga, D. (2022). Human-Centered Artificial Intelligence: The Superlative Approach to Achieve Sustainable Development Goals in the Fourth Industrial Revolution. *Sustainability*, *14*(13). https://doi.org/10.3390/SU14137804

Parashar, M., & Hariri, S. (2005). Autonomic computing: An overview. *Lecture Notes in Computer Science*, *3566*, 257–269. https://doi.org/10.1007/11527800_20

Rožanec, et al. (2023). Human-centric artificial intelligence architecture for industry 5.0 applications. *International Journal of Production Research*, *2023*(20), 6847–6872. https://doi.org/10.1080/00207543.2022.2138611

Schall, D., Truong, H. L., & Dustdar, S. (2008). Unifying human and software services in web-scale collaborations. *IEEE Internet Computing*, *12*(3), 62–68. https://doi.org/10.1109/MIC.2008.66

Tan, Q., et al. (2019). Anthropocentric Approach for Smart Assembly: Integration and Collaboration. *Journal of Robotics*, *2019*. https://doi.org/10.1155/2019/3146782

Vizcarrondo, J., Aguilar, J., Exposito, E., & Subias, A. (2017). MAPE-K as a service-oriented architecture. *IEEE Latin America Transactions*, *15*(6), 1163–1175. https://doi.org/10.1109/TLA.2017.7932705