Enabling Seamless HRC Integration: The AI-PRISM Reference Architecture

Miguel A. Mateo-Casalí [1], Josep Rueda[2], Samuel Olaiya Afolaranmi[3], Jose Luis Martinez Lastra[3], Davide Boscaini[4], Laura Moya-Ruiz[1], and Francisco Fraile[1]

[1] *Research Centre on Production Management and Engineering (CIGIP), Universitat Politècnica de València, Camino de Vera s/n, 46022, Valencia (Spain).*
[2] *IKERLAN Technology Research Centre, Basque Research and Technology Alliance (BRTA), Arrasate, Basque Country (Spain).*
[3] *Future Automation Systems and Technologies Laboratory (FAST-Lab), Tampere University, Tampere, Finland 4 University 1, Address, City, Index, Country*

Abstract

This article addresses the challenges of deploying AI solutions for human-robot collaboration (HRC) in the context of industry 5.0, introducing the AI-PRISM framework as a solving proposal. This framework presents a high-level reference architecture which is designed to provide efficient and adaptive robotic systems in industrial environments, using Kubernetes clusters managed through Rancher to ensure reliability and easy maintenance and monitorization. The four tiers identified in this architecture start from the lower levels with the physical devices in the Device Tier, followed by the Ambient Network Tier to ensure that real-time communication required in any human-robot collaboration environment. The Fog Tier involves more processing capacity and resources, but keeping almost real-time responses, therefore, modules related with reasoning, perception and simulation are here. The Enterprise Tier focuses on Continuous Integration and Deployment (CI/CD), allowing to deploy components, and non real-time modules. In addition to the Reference Architecture components, the AI-PRISM Open-Access Platform is a web-based gateway, that provides stakeholders access to tools, technologies, and resources, encouraging collaboration, experimentation and testing in AI-based collaborative robotics.

Keywords

Cloud manufacturing, Microservices, Architectures

1. Introduction

Human-robot collaboration (HRC) is a prominent feature of modern Industry 5.0 applications [1]. HRC technologies maintain human involvement in workplaces by leveraging robotics to alleviate mental and physical stress. Artificial Intelligence (AI) can enhance the perception and reasoning capabilities of robotic agents to achieve paramount performance and accuracy. However, the deployment of AI models in industrial environments poses significant challenges, particularly in scenarios with strict timing constraints, where models are deployed in environments with high processing power, and results must reach actuators (cobots) immediately.

To address these challenges, it is necessary to adopt architectural frameworks that allow the correct and efficient management of IT resources, and that are able to leverage edge computing and real time communications in an effective manner to reduce latency and ensure overall performance [2]. Besides timing constraints, the correct functioning of the system requires frequent updates of different software components (including AI models), and at the same time, ensure that these continuous updates do not cause any interruptions in the HRC processes. Adherence to standards and compliance with the regulatory frameworks and industry best practices, particularly on security and data protection, are also important aspects to consider. Moreover, the rapid advancements in collaborative robotics have led to a wide range of physical equipment (robots, sensors, camera, edge devices) that is available to build HRC applications. This implies that it is

12th International Conference on Interoperability for Enterprise Systems and Applications (I-ESA24), April 10th 12th, 2024, Chania, Crete, Greece
mmateo@cigip.upv.es (M. A. Mateo-Casali); jrueda@ikerlan.es (J. Rueda), samuel.afolaranmi@tuni.fi (S.O. Afolaranmi); jose.martinezlastra@tuni.fi (JL. Martinez Lastra); dboscaini@fbk.it (D. Boscaini); ffraile@cigip.upv.es (F. Fraile) 0000-0001-5086-9378((M. A. Mateo-Casali), 0000-0002-8132-3017 (S.O. Afolaranmi); 0000-0001-6227-3408 (JL. Martinez Lastra); 0000-0003-4887-2038 (D. Boscaini); 0000-0003-0113-7976(L. Moya-Ruiz); 0000-0003-0852-8953(F. Fraile)
© 2024 Copyright for this paper by its authors.
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).
CEUR Workshop Proceedings ceur-ws.org ISSN 1613-0073
CEUR-WS.org/Vol-4167/paper8.pdf

important to incorporate standardized interfaces and protocols to ensure interoperability. Finally, network and connectivity for HRC must be considered, as bandwidth optimization and proper bandwidth management are necessary to cope with the transmission of data. As a response to these challenges, this paper presents the reference architecture of the AI Powered Human-Centered Robot Interactions for Smart Manufacturing (AI-PRISM) project. The AI-PRISM project is focused project is focused on enhancing smart manufacturing through AI powered, HRC applications, and the AI-PRISM reference architecture provides a high-level architectural representation of AI enhanced HRC systems which can lead to more adaptive, efficient, and scalable robotic systems in industrial environments.

Throughout section 2 the AI-PRISM Reference Architecture will be explained, detailing its parts. Within each tier, the different modules that can be deployed and its role, justifying its position, will be included.

2. AI-PRISM Reference Architecture

The AI-PRISM project provides a reference architecture that provides a structured solution for the technical implementation. This architecture offers a framework or guidance for the technical elements and the internal communications of AI-PRISM HRC applications, and a framework for the concrete implementation of the different demonstrators and project industrial scenarios or pilots. One of the most important aspects of the architecture is the definition of connections between hardware components, communication networks, execution environments and software components. **Figure *1*** provides a detailed breakdown of each software component, including an in-depth description of the functional components mapped to the AI-PRISM reference architecture (AI-PRISM-RA) and indications of anticipated input and output interfaces.

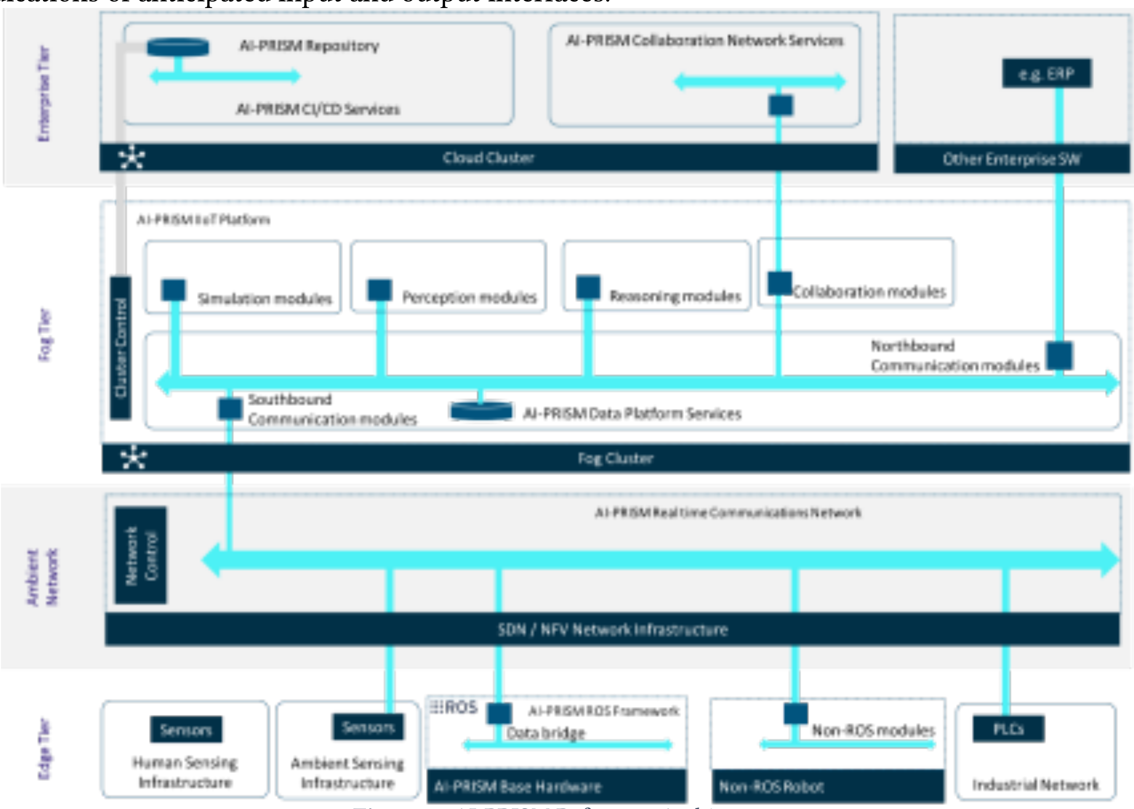

Figure 1. AI-PRISM Reference Architecture

The AI-PRISM-RA is based on architectures and reference models established under the ISO/IEC/IEE 42010 standard for system and software engineering architecture descriptions [3]. This standard provides different definitions and structures for complex architectures. An illustrative example of the use of this standard is the Industrial Internet Architecture Framework (IIRA), as well as other relevant reference architectures and models for digital manufacturing [4].

The AI-PRISM project elaborates its architecture around four different tiers, the Edge tier, Environment tier, Fog tier, and Enterprise tier. Each of these tiers encompasses different hardware components providing computational resources to execute the different software components of AI-PRISM HRC applications. AI-PRISM adopts a distributed cluster architecture, in which edge devices and other computational resources are treated as independent Kubernetes clusters [5], managed centrally by a cluster management and control component. This component facilitates the installation, management and monitoring of clusters and containers, offering an easy-to-use interface to discover and install components and applications from the AI-PRISM repository. This Kubernetes based independent cluster architecture not only enables continuous updates using Continuous Integration and Delivery (CI/CD) infrastructure, but it also provides a layer of reliability in case of failure as it is possible to move workloads from one cluster to another as a fallback mechanism. Furthermore, AI-PRISM components are independent microservices. With this design, modules can be tested prior to deployment and, in case of failure, can be easily reverted or restarted without compromising the rest of the modules.

2.1. Edge Tier

The Edge Tier groups all the different devices that play a role in the interaction with the ambient, or human-robot collaboration environment. It includes sensors, robots, other manufacturing equipment, and the controller that commands them. Each device or set of devices is coupled with a controller. This conjunction of devices and their controller are meant to perform a task, for example, a couple of cameras and their controller performing a 3D reconstruction of the environment.

Within this tier, the follwing modules can be found:

(i) **Ambient and Human Sensing Infrastructure**, which encompasses all the equipment capable of perceiving the environment, such as cameras, sensors, and robots equipped with sensors.

(ii) **Base Hardware**, covering all non-sensorial hardware like actuators and robots, as well as the hardware that will host the execution of controllers to direct all the mentioned devices in this Device Tier.

(iii) **ROS Framework** includes the set of modules facilitating communication among other modules based on ROS. Additionally, it's crucial to consider two additional components which are integrated into the ROS framework using ad-hoc modules: a. **Non-ROS Modules**, which operate independently of the ROS Framework and ooffer adiverse range of complementary functions.

b. **Programmable Logic Controllers (PLCs) in the industrial network**. These PLCs are programmable logic devices that play a pivotal role in industrial automation by managing communication and controlling equipment within manufacturing environments, contributing to the seamless operation of the overall system.

All these modules and infrastructure, along with ROS Framework and Non-ROS modules, constitute the "Device Tier", are managed through the central cluster management component, built on top of Rancher [5], an open-source tool designed to supervise computational unit clusters in Kubernetes. Users can deploy containers on individual units, assign workloads to an entire cluster, and monitor performance through an intuitive interface provided by Rancher. It's worth highlighting that, by allocating a workload to a device cluster, any residual computational capacity can contribute to additional tasks, enhancing both performance and reliability compared to non-clustered distributions. While the primary function of each device is to manage its specific set, it doesn't remain underutilized, as it can collaborate with other devices in their tasks or perform common functions not directly related. Furthermore, the Edge Tier plays a crucial role in executing critical workloads, emphasizing the adaptability of Kubernetes cluster architecture to dynamically scale with low latency. In this context, the ROS Framework addresses the hardware interoperability challenge, ensuring that the Kubernetes infrastructure is compatible with various types of robots and edge devices used in HRC (Human-Robot Collaboration).

2.2. Ambient Network Tier

The Ambient Network encloses all the infrastructure that conforms the real-time communication network between the different Edge Tier devices ensuring that their time sensitive needs are met, as well as the tools to monitor and control the traffic inside said network. The Ambient Tier relies on a SDN (Software-Defined Networking) controller which allows configuring the low-level behavior of each device therefore allowing us to disregard the low-level functionalities during the development of new applications. This tier also connects Edge devices with those modules executed in the Fog Tier.

2.3. Fog Tier

The Fog Tier groups those AI-PRISM modules that are deployed between the Edge Tier and the Cloud Tier. Some modules, traditionally, with real-time constraints or with minimal computational requirements will be running on a controller on the same workspace, while other modules with no real-time constraints as well as data storage services will be running on a cloud environment. Therefore, two kinds of communication platforms can be found in the Fog cluster: the ROS DDS communication, which connects the devices with real-time requirements; and the IIoT Platform, which is used to share data between these resource expensive modules. In the same way as in the Device Tier, the deployment and management of modules into clusters is carried out using Rancher considering that this Fog Cluster, as stated before, will be running these two clearly differentiated kinds of modules.

This tier encompasses various modules to form a cohesive system that enhances the capabilities of the Tier:

(i) **Data Platform Service**, a fundamental component, plays a crucial role in efficiently managing the data collected from the edge through the IIoT Platform. Its primary functions include storing and making this data easily accessible to other modules, ensuring a flow of information within the system, as well as controlling the quality of service of communications through the SDN network APIs.

(ii) **Simulation Modules**, providing users with tools to simulate scenarios, facilitating tests of the developed components. This feature empowers users to optimize solutions by subjecting them to different test scenarios, improving their adaptability and performance in real-world applications.

(iii) **Perception Modules**, another integral part of this tier, leverage data acquired through the sensing infrastructure of the Device Tier. These modules are responsible for creating a digital reconstruction of the working environment, providing a representation of the actors. Additionally, Perception Modules excel in describing actors within the environment, identifying objects from CAD models, and using neural networks for specialized tasks, showcasing the advanced cognitive capabilities of the system.

(iv) **Reasoning Modules** contribute to the dynamism of the collaborative environment by coordinating available agents. These modules command actions based on ongoing tasks and awareness derived from the environment, ensuring synchronized system operation.

(v) **Collaboration Modules** refers to the modules that while ensuring safe interactions between human and robot agents in the ambient, provide interfaces between the human workers and the robotic agents allowing the human to teach the system a task that is desired to be executed by the robot through demonstration. This teaching can be done by observing the human user or by graving the robotic arm and guiding it through the process while we track the motion. Then the robot executes said action, which could be collaborative in such a way that the human could instruct the robot by using commands, which could include any kind of human-robot interaction, like shouting a verbal directive or physically interrupting the robot motion. And, in the same manner, the robot can provide information about its current state to the user using acoustic or visual indicators.

In the same way that there is a centralized management of the deployed modules in the Device Tier, the fog cluster or clusters are also managed through Rancher, keeping a centralized control of the whole distributed architecture.

2.4. Enterprise Tier

The Enterprise Tier groups the components that enable CI/CD (Continuous Integration and Continuous Deployment) of solutions, and the components that enable open access of the AI PRISM solutions to external collaborators. These components can be found in the cloud cluster where the AI-PRISM partners, who are connected to said cluster, will have access to the AI-PRISM tools allowing them to deploy the desired components to their working environment, and they will also be contributing to the aforementioned solutions if any relevant module or piece of data can be used to improve the AI-PRISM capabilities. The Open Access Platform services provide a layer of hyper-connectivity enabling virtual access to remote resources and are described in detail in the next section.

2.4.1. AI-PRISM Open-Access Platform

The AI-PRISM Open-Access Platform (also known as the Network suite) is a web-based platform through which AI-PRISM technologies, modules, components, and infrastructure can be accessed. The platform acts as the gateway for the AI-PRISM stakeholders (industry experts, researchers, developers, manufacturers etc.) to access AI-PRISM tools and utilize the available resources. Through the platform services, end-users can gain remote access to various AI-PRISM offerings (software, algorithms, and datasets), industrial infrastructure (industrial facilities and testing sites) and resources such as robotic systems and industrial equipment. The platform consists of a set of components that aid the orchestration and management of access to AI-PRISM offerings to interested parties and stakeholders. These components, also called "engines" includes user interface (UI) engine (for accessing the platform), user access control engine (for managing the users' rights and access), training and certification engine (for teaching and education), planning and scheduling engine (for handling user visits to testing sites and matching users to resources), contracts and finance engine (for managing financial agreement between stakeholders) and virtual pilot engine. The services provided by the platform through the different engines enhances the visibility of the available AI-PRISM technologies and resources.

The Open-Access Platform ensures access and facilitates the findability of tools and resources needed by the AI-PRISM stakeholders. For instance, System Admins can utilize the platform to discover a series of AI tools, ascertain and deploy modules compatible with their local infrastructure. Also, shopfloor operators working with collaborative robots can utilize the platform to access the AI-PRISM simulation environment to model the human-robot work layout as a means of assessing the collaborative interactions and monitoring the efficiency of the workflows. With this, the open-access platform facilitates the collaboration between infrastructure providers and the end-users, who utilize the available infrastructure and resources to test and validate the different AI-PRISM technologies to assess their performance and functionality. In addition, the end-users can leverage the available resources to develop new AI based collaborative robotics and human-centered solutions that can enhance their business operations. This collaboration increases the exploitation potential of the newly developed solution and brings added value such early experimentation, possibility to test the solutions before investing money as well as certification and training. The collaboration also ensures that the infrastructure providers can benefit through equipment utilization, capitalization of unused resources and acceleration of upskilling for their workforce. In general, the AI-PRISM Open Access Platform encourages the development of new AI-based collaborative robotics solutions, facilitates knowledge sharing and transfer, enhances upskilling of workforce regarding the application of AI in manufacturing, and promotes the ethical use of AI.

2.5. Edge Tier

In conclusion, the AI-PRISM reference architecture presents a seamless solution to the challenges of deploying AI-driven human-robot collaboration (HRC) systems in industrial environments. By leveraging a distributed cluster architecture and adopting standardized frameworks such as Kubernetes, ROS, and PLC, AI-PRISM ensures efficient management of hardware components, communication networks, and software modules. AI-PRISM provides a structured approach to deploying and managing HRC applications, addressing latency, reliability, and scalability issues case-by-case through its Edge, Ambient Network, Fog, and Enterprise tiers. In addition, the AI PRISM open-access platform facilitates collaboration and knowledge sharing among stakeholders, enhancing the development and adoption of AI-based solutions while ensuring ethical practices and regulatory compliance. The AI-PRISM reference architecture offers a promising path towards a more adaptable, efficient, and scalable robotic system in industrial environments, stepping into the era of

Industry 5.0 applications.

Acknowledgements

AI-PRISM has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101058589.

Declaration on Generative AI

The author(s) have not employed any Generative AI tools.

References

[1] Ranavolo, A., Chini, G., Draicchio, F., Silvetti, A., Varrecchia, T., Fiori, L., Tatarelli, A., Rosen, P.H., Wischniewski, S., Albrecht, P., Vogt, L., Bianchi, M., Averta, G., Cherubini, A., Fritzsche, L., Sartori, M., Vanderborght, B., Govaerts, R., & Ajoudani, A. (2021). Human-Robot Collaboration (HRC) Technologies for Reducing Work-Related Musculoskeletal Diseases in Industry 4.0. Proceedings of the 21st Congress of the International Ergonomics Association (IEA 2021). https://doi.org/10.1007/978-3-030-74614-8_40

[2] Semeraro, F., Griffiths, A., & Cangelosi, A. (2021). Human-Robot Collaboration and Machine Learning: A Systematic Review of Recent Research. Robotics Comput. Integr. Manuf., 79, 102432.

[3] "ISO/IEC/IEEE Systems and software engineering -- Architecture description," in ISO/IEC/IEEE 42010:2011(E) (Revision of ISO/IEC 42010:2007 and IEEE Std 1471-2000), vol., no., pp.1-46, 1 Dec. 2011, doi: 10.1109/IEEESTD.2011.6129467.

[4] Fraile, F.; Sanchis, R.; Poler, R.; Ortiz, A. Reference Models for Digital Manufacturing Platforms. Appl. Sci. 2019, 9, 4433. https://doi.org/10.3390/app9204433 [5] Kubernetes. (2023). Kubernetes documentation. Retrieved February 4, 2024, from https://kubernetes.io/docs/

[6] Rancher Labs. (2023). Introduction to Rancher. Retrieved February 4, 2024, from https://ranchermanager.docs.rancher.com/