Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-5 | May-2026 | Page 967-991

Human-Centric AI Framework for Smart Decision Support in Industry 5.0

[1]Shyam Kumar Singh, [2]Dr. Shikha Tiwari
[1]Student, [2]Associate Professor
[1,2]Amity University Chhattisgarh, India
[1]shyam.singh3@s.amity.edu, [2]shkhtiwari583@gmail.com

Abstract
Industry 5.0 represents the next evolution in manufacturing, integrating advanced automation with human creativity, resilience, and sustainability. This paper presents a human-centric artificial intelligence (AI) framework designed to support smart decision-making in Industry 5.0 environments. The proposed framework emphasizes explainability, user feedback, and human-AI collaboration. It combines real-time data analytics, predictive modeling, and interactive visualization to empower human operators in decision processes. Key components include advanced AI modules (forecasting, active learning, explainable AI), a simulated-reality environment for testing scenarios, and adaptive user interfaces. A conceptual system architecture is presented, and its operation is illustrated with hypothetical manufacturing case studies. Results from a simulated implementation demonstrate improved decision accuracy and trust compared to a baseline automated system. The findings underline that integrating human expertise with AI yields robust, transparent decision support, aligning with Industry 5.0's human-centric goals.

Keywords: Industry 5.0, Human-centric AI, Decision Support, Smart Manufacturing, Explainable AI

I. INTRODUCTION

Artificial intelligence (AI) is a field of study that focuses on creating systems or machines capable of performing tasks that require human intelligence. These tasks include machine learning, problem-solving, pattern recognition, natural language processing, and more. In general, AI aims to develop systems that can simulate some of the human capabilities of learning and reasoning, such as visual perception, voice recognition, and linguistic translation, to perform specific tasks autonomously or semi-autonomously. John McCarthy specifically described AI as the scientific and technological competence for the development of intelligent computer programmers [1].

Machine learning (ML) and deep learning (DL) are two of the most used methods of artificial intelligence [2]. These models rely on data and are used to develop predictive models by individuals, companies, and governmental organizations. Currently, methods of automated learning are being developed capable of handling the complexity and unpredictability of information in various industrial sectors such as food, biomedical, and aerospace [3]. By combining the principles of ML and DL with advanced optimization techniques into industrial

ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-5 | May-2026 | Page 967-991

processes, practitioners are empowered to navigate modern challenges with heightened precision and efficacy [4,5,6,7,8].

At the industrial level, this has led to the development of advanced approaches to cognitive computing and deep learning for automated applications such as visual inspection, fault detection, and maintenance in manufacturing systems. Deep learning approaches are actively used in production systems, from supply chain to manufacturing programs, ensuring the highest quality and safety in multiple sectors [9], including public policy and public administration.
AI helps to monitor and manage production processes comprehensively and not only to replace humans in risky operations but also to support them, placing them at the center of the process through a specific approach called human- centered. Numerous examples include the adoption of human- centered approaches by major companies such as Apple, Google, and Microsoft during the development of artificial intelligence software [10]. Mercedes-Benz has replaced standard robots with AI-based collaborative robots (cobots), enabling the production of customized cars more efficiently [11]. IBM Watson has proposed a system that could recommend cancer treatments in line with the doctor's recommendations most of the time [12]. Furthermore, AI and sensors offer real-time technological advancements that support the shipping industry, enhancing safety, reducing costs, thereby boosting productivity in international trade and embracing a more intelligent and sustainable future [13,14].
This underlines the importance of a multidisciplinary approach that allows different sectors and roles to communicate and highlight positive aspects and diverse perspectives of AI use. This article explores the key challenges associated with the human-centered artificial intelligence (HCAI) approach in Industry 5.0 and the circular economy, as well as the role of decision-makers. The methodological approach will be integrative and comprehensive, employing qualitative methods to examine the impact and applications of HCAI in these fields.
This involves conducting an extensive literature review, critically analyzing major challenges with a focus on smart manufacturing and the role of policymakers, and identifying future research trends and needs. As a guiding framework for our study, we set the following research questions (RQs):

1. What does the human-centered AI approach mean in the context of the industrial sector, in general, and in the manufacturing sector in particular?
2. What is the associated literature of the human- centered AI approach when realizing the vision of Industry 5.0 and the main challenge to address?
3. What role do decision-makers play in the successful adoption and implementation of HCAI in Industry 5.0 and the circular economy?
4. How can interdisciplinary collaboration enhance the effectiveness of HCAI applications in industry?
5. What are the future directions for research and development in the field of HCAI? The remainder of the paper unfolds as follows: the theoretical framework is outlined in Section 2, while Section 3 details the HCAI and Industry 5.0 and provides empirical evidence on the perspectives of HCAI for additive manufacturing (AM). Section 4 discusses the challenges for

Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-5 | May-2026 | Page 967-991

Government 5.0 in the light of the HCAI approach. Section 5 proposes some suggestions for further research.

Figure 1. Automated conveyor assembly environment in a modern smart factory. Human operators collaborate with AI- enabled machines to ensure efficient and flexible production.

II. RELATED WORKS

1.Industry 5.0

Industry 4.0 was introduced at the Hannover Trade Fair in 2011, aiming to introduce new technologies into manufacturing with the purpose of achieving high levels of operational efficiency and productivity (Sanchez, Exposito, and Aguilar (2020)). While technology is emphasized as a means toward greater efficiency and productivity to enhance competitiveness in the global market, some emphasis have been placed on using such technology to reach certain level of human-centricity through the concept of Operator 4.0. Operators 4.0 are operators who will be assisted by systems providing relief from physical and mental stress, without compromising the production objectives (Romero et al. (2016); Romero, Stahre, and Taisch (2020); Kaasinen et al. (2020)). Industry 5.0 is envisioned as a co-existing industrial revolution (Xu et al. (2021)), for which two visions have emerged: (i) one that refers to human-robot co-working, and (ii) a second one as a bioeconomy where renewable biological resources are used to transform existing industries (Demir, Döven, and Sezen (2019)). In this work, we focus on Industry 5.0 as a value-driven manufacturing paradigm and revolution that highlights the importance of research and innovation to support the industry while placing the well- being of the worker at the center of the production process (Xu et al. (2021)). Such a revolution must attempt to satisfy the needs placed in the Industrial Human Needs Pyramid, which range from workplace safety to the development of a trustworthy relationship between humans and machines that enables the highest level of self-esteem and self-actualization, realizing and fulfilling their potential (Lu et al. (2022)). It aims to intertwine machines and humans in a synergistic collaboration to increase productivity in the manufacturing industry while retaining human workers. Furthermore, it seeks to develop means that enable humans to unleash their critical thinking, creativity, and domain knowledge. At the same time, the machines can be

Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-5 | May-2026 | Page 967-991

trusted to autonomously assist on repetitive tasks with high efficiency, anticipating the goals and expectations of the human operator, and leading to reduced waste and costs (Nahavandi (2019); Demir, Döven, and Sezen (2019); Maddikunta et al. (2021)). Such communication and collaborative intelligence enable the development of trustworthy coevolutionary relationships between humans and machines. To foster the development of trustworthy coevolutionary relationships, interfaces must consider the employee's characteristics (e.g., age, gender, and level of education, among others) and the organizational goals. One example of collaboration between humans and machines is realized with cobots, where the cobots share the same physical space, sense and understand the human presence, and can perform tasks either independently, simultaneously, sequentially, or in a supportive way (El Zaatari et al. (2019)). In order to realize the Industry 5.0 vision, the focus must be shifted from individual technologies to a systematic approach rethinking how to (a) combine the strengths of humans and machines, (b) create digital twins of entire systems, and (c) widespread use artificial intelligence, with a particular emphasis generation of actionable items for humans. While research regarding Industry 5.0 is incipient, it has been formally encouraged by the European Commission through a formal document released back in 2021 (Ind (n.d.)).

2.Considering standards and regulations

In order to realize the vision laid out for Industry 5.0, constraints and directions imposed by existing regulations must be considered. Furthermore, standards should be taken into account to ensure that the fundamental blocks can be universally understood and adopted to achieve compatibility and interoperability.

Cybersecurity is considered a transversal concern in the architecture presented in this work. Among the standards and regulations that relate to it we must mention the ISO 27000 family of standards (ISO (n.d.)), the USA Cybersecurity Information Sharing Act (CISA) (CIS (n.d.)), the EU Cybersecurity Act (cyb (n.d.)), and the EU Network and Information Security Directive II (NIS II) (NIS (n.d.)). The ISO 27000 standards defined a common vocabulary and provided an overview of information security management systems. The NIS II directive aimed to force certain entities and sectors of the European Union to take measures to increase the overall cybersecurity level in Europe. The European Union Cybersecurity Act provided complementary legislation by establishing a cybersecurity certification framework for products and services and granted a permanent mandate to the EU agency for cybersecurity (ENISA) to inform the public regarding certification schemas and issue the corresponding certificates. Finally, the CISA established a legal ground for information sharing between the USA government agencies and non-government entities for cyberattack investigations. When considering data management, much emphasis is being put on privacy. The General Data Protection Regulation (GDPR) (GDP (n.d.)), ePrivacy directive (ePr (n.d.)), or Data Governance Act (dat (n.d.)), issued by the European Union, are relevant when managing and sharing data, especially personal or sensitive data. The GDPR establishes a legal framework setting guidelines to process and collect personal information of persons living in the European Union. The ePrivacy directive regulates data protection and privacy, emphasizing issues related to confidentiality of information, treatment of spam, cookies, and traffic data. Finally, the Data Governance Act

Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-5 | May-2026 | Page 967-991

promotes wider re-use of data, using secure processing environments, data anonymization techniques (e.g., differential privacy), and synthetic data creation; and establishes a licensing regime for data intermediaries between data holders and data users. While these regulations and directives must be considered, we provide no systemic solution from an architectural point of view.

Finally, given the increasing adoption of artificial intelligence, a legislative effort is being made to regulate its use. For example, the Artificial Intelligence Act (AIA (n.d.b)), issued in the European Union, was the first law of this kind issued by a significant regulator worldwide. The law categorizes artificial intelligence applications into three risk categories: (a) unacceptable risk (e.g., social scoring systems), which are banned, (b) high-risk (e.g., resume scanning applications), which are subject to specific legal requirements, and (c) applications that do not fall into categories (a) and (b), which remain unregulated. Another example is a law issued by the Federative Republic of Brazil (AIA (n.d.a)), which establishes the principles, obligations, rights, and governance instruments regarding the use of artificial intelligence.

While the abovementioned list is not exhaustive, it provides a high-level view of the main concerns and topics that must be considered.

3. Enabling technologies

In order to realize a human-centric artificial intelligence architecture for Industry 5.0 applications, a set of technologies that enable a human-centric approach we consider a set of technologies must be taken into account. We consider five of them related to the field of artificial intelligence: (i) active learning, (ii) explainable artificial intelligence, (iii) simulated reality, (iv) conversational interfaces, and (v) security. Below we introduce some related work regarding each of them, and in Section 3 describe the corresponding architecture building blocks.

3.1. Active learning

The adoption of artificial intelligence in manufacturing and the complementarity of the machine and human capabilities is reshaping jobs, and human-machine cooperation opportunities are emerging. One way to realize such human- machine cooperation is through the Active Learning paradigm, which considers an artificial intelligence model can be improved by carefully selecting a small number of data instances to satisfy a learning objective (Settles (2009)). Active Learning is built upon three assumptions: (i) the learner (artificial intelligence model) can learn by asking questions (e.g., request a target variable's data), (ii) there is an abundance of questions that can be asked (e.g., data, either gathered or synthetically created, without a target value), and (iii) there is a constrained capacity to answer such questions (and therefore, the questions must be carefully selected) (Elahi, Ricci, and Rubens (2016)). Therefore, applied research is focused on how to structure use case solutions so that through a human-in-the-loop, artificial intelligence models can benefit from human expertise to make decisions and provide valuable input, which is later used to enhance the models (Kumar and Gupta (2020); Schröder and Niekler (2020); Budd, Robinson, and Kainz (2021)).

Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-5 | May-2026 | Page 967-991

We discriminate between data obtained from real sources and synthetic data (created through some procedure) regarding the source of the data. Synthetic data is frequently used to enlarge the existing data or to generate instances that satisfy specific requirements when similar data is expensive to obtain. While many techniques and heuristics have been applied in the past to generate synthetic data, the use of Generative Adversarial Networks (GANs) has shown promising results and been intensely researched Zhu and Bento (2017); Mahapatra et al. (2018); Sinha, Ebrahimi, and Darrell (2019); Mayer and Timofte (2020)). Strategies related to data selection are conditioned by how data is generated and served. If the data is stored, data instances can be scanned and compared, and some latency can be tolerated to make a decision. On the other hand, decisions must be made at low latency in a streaming setting, and the knowledge is constrained to previously seen instances. Data selection approaches must consider informativeness (quantifying the uncertainty associated to a given instance, or the expected model change), representativeness
(number of samples similar to the target sample), or diversity criteria (selected samples scatter across the whole input space) (Wu (2018)). Popular approaches for classification problems are the random sampling, query-by- committee (Seung, Opper, and Sompolinsky (1992)), minimization of the Fisher information ratio (Padmanabhan et al. (2014)), or hinted sampling with Support Vector Machines (Li, Ferng, and Lin (2015)).

Active learning has been applied to several manufacturing use cases. Nevertheless, applied research in the manufacturing sector remains scarce (Samsonov et al. (2019); Meng et al. (2020)), but its relevance increases along with the proliferation of digital data and democratization of artificial intelligence. In the scientific literature, authors report using Active Learning to tackle quality control, predictive modeling, and demand forecasting. For example, active learning for quality control was applied to predict the local displacement between two layers on a chip (Dai et al. (2018)) or gather users' input in visual quality inspection of printed company logos on the manufactured products (Trajkova et al. (2021)). In predictive modeling, it was applied in the aerospace industry to assist a model in predicting the shape control of a composite fuselage (Yue et al. (2020)). Finally, in the demand forecasting use case, the authors explored using active learning to recommend media news and broaden the logisticians' understanding of the domain while informing relevant events that could affect the demand, to reach better decisions (Zajec et al. (2021)). Regardless of the successful application in several manufacturing use cases, active learning is not widely adopted in manufacturing and could be applied to enhance cybersecurity capabilities and fatigue monitoring systems (Li et al. (2019)), among others.

3.2.Explainable Artificial Intelligence

While artificial intelligence was applied to manufacturing problems in the past (Bullers, Nof, and Whinston (1980)), it has become increasingly common to rely on artificial intelligence models to automate certain tasks and provide data-based insights (Chien et al. (2020)). When human decision-making relies on artificial intelligence models' outcomes, enough information regarding the models' rationale for such a forecast must be provided (Ribeiro, Singh, and

Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-5 | May-2026 | Page 967-991

Guestrin (2016); Lundberg and Lee (2017)). Such information enables the user to assess the trustworthiness and soundness of the provided forecast and therefore ensure decisions are made responsibly (Das and Rad (2020)). Research on techniques and approaches that convey information regarding the rationale behind the artificial intelligence models, or an approximation to it, and how such information is best presented to the users, is done in a sub-field of artificial intelligence, known as eXplainable Artificial Intelligence (XAI) (Henin and Le Métayer (2021)). Such approaches can be classified according to different taxonomies. Among them, there is consensus that artificial intelligence models can be considered either white-box models (inherently interpretable models), or black-box models (models that remain opaque to the users) (Loyola-Gonzalez (2019)).

Regarding the characteristics of the explanation, Angelov et al. (2021) divide XAI methods into four groups, considering whether (i) the explanations are provided at a
local (for a specific forecast) or global (for the whole model) level, (ii) the models are transparent or opaque to the users, (iii) the explainability techniques are model-specific or model-agnostic, and (iv) the explanations are conveyed through visualizations, surrogate models or taking into account features relevance.

The scientific literature reports an increasing amount of use cases where explainable artificial intelligence is applied. Meister et al. (2021) applied deep learning models to automate defect detection on composite components built with a fiber layup procedure. Furthermore, the authors explored using three Explainable Artificial Intelligence techniques (Smoothed Integrated Gradients (Sundararajan, Taly, and Yan (2017)), Guided Gradient Class Activation Mapping (Shrikumar, Greenside, and Kundaje (2017)) and DeepSHAP (Selvaraju et al. (2017))), to understand whether the model has learned and thus can be trusted that it will behave robustly. Senoner, Netland, and Feuerriegel (2021) developed an approach to creating insights on how production parameters contribute to the process quality based on the estimated features' relevance to the forecast estimated with the Shapley additive explanations technique (Lundberg and Lee (2017)). Finally, Serradilla et al. (2020) implemented multiple machine learning regression models to estimate the remaining life of industrial machinery and resorted to the Local Interpretable Model-Agnostic Explanations technique (Ribeiro, Singh, and Guestrin (2016)) to identify relevant predictor variables for individual and overall estimations. Given research had little consideration for the complexity of interactions between humans and their environment in manufacturing, much can be done to develop XAI approaches that consider anthropometrics, physiological and psychological states, and motivations to not only provide better explanations, but also enhance the workers' self-esteem and help them towards their self-actualization (Lu et al. (2022)).

3.3.Simulated reality

Under simulated reality, we understand any program or process that can generate data resembling a particular aspect of reality. Such a process can take inputs and produce outputs, such as synthetic data or outcomes that reflect different scenarios or process changes.

Machine learning models can solve complex tasks only if provided with data. Acquiring high-quality data can be a complex and expensive endeavor: lack of examples concerning faulty

Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-5 | May-2026 | Page 967-991

items for defect detection systems, wearing down and damaging a robotic system during data collection, or human errors when labeling the data are just some examples. Synthetic data is envisioned as a solution to such challenges. Much research is invested in making it easy to generate while avoiding annotation pitfalls, ethical and practical concerns and promising an unlimited supply of data (de Melo et al. (2021)). In addition, much research was invested in the past regarding synthetic data generation to cope with imbalanced datasets.

Nevertheless, the development of GANs opened a new research frontier, leading to promising results (Creswell et al. (2018)). GANs consist of two networks: a generator (trained to map some noise input into a synthetic data sample) and a discriminator (that, given two examples, tries to distinguish the real from the synthetic one). This way, the generator learns to generate higher-quality samples based on the discriminator's feedback. While they were first applied to images (Goodfellow et al. (2014)), models have been developed to enhance the quality of synthetic images and to apply them to other types of data, too (Patki, Wedge, and Veeramachaneni (2016); Xu et al. (2019)).

Simulated reality can be considered a key component of Reinforcement Learning. The reinforcement learning agent can explore an approximation of the real world through the simulator and learn efficient policies safely and without costly interactions with the world. Furthermore, by envisioning the consequences of an action, simulations can help to validate desired outcomes in a real-world setting (Amodei et al. (2016)). Simulated reality has been applied in a wide range of manufacturing use cases. Neural Style Transfer (Wei et al. (2020)) has been successfully used to generate synthetic samples by fusing defect snippets with images of non-defective manufactured pieces. Such images can be later used to enhance the algorithm's predictive capacity. Simulators have been widely applied to train Reinforcement Learning models in manufacturing. Mahadevan and Theocharous (1998) used them to simulate a production process and let the RL algorithm learn to maximize the throughput in assembly lines, regardless of the failures that can take place during the manufacturing process. Oliff et al. (2020) simulated human operators' performance under different circumstances (fatigue, shift, day of week) so that behavioral policies could be learned for robotic operators and ensure they provided an adequate response to the operators' performance variations. Finally, Johannink et al. (2019) used Reinforcement Learning to learn robot control and evaluated their approach in both real-world and simulated environments. Bridging the gap from simulated to real-world knowledge remains a challenge.

3.4.Intention Recognition in Manufacturing Lines

The development of Industrial Internet of Things (IIoT) technologies and the availability of low-cost wearable sensors have enabled access to big data and the utilization of the sensors for the manufacturing industry (Jeschke et al. (2017)). Recently, various deep learning-based methods have been proposed to learn valuable information from big data and improve the effectiveness and safety of the manufacturing lines. In particular, worker's activity and intention recognition can be used for quantification and evaluation of the worker's performance and safety in the manufacturing lines. In addition, with the introduction of autonomous robots for

Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-5 | May-2026 | Page 967-991

effective manufacturing, efficient collaboration between robots and workers and the safety of workers are also becoming increasingly important.

Thanks to the miniaturization and reduction of costs, the adoption of wearable sensors has also been growing in the industrial context to investigate workers' conditions and well- being. Worker's health is a key factor in determining the organization's long-term competitiveness, and it is also directly related to production efficiency. The cumulative effect of positive impacts on the human factor brings economic benefit through productivity increase, scrap reduction, and decreased absenteeism. Few research works have been recently developed, where workers' physiological data are used to infer the insurgence of phenomena such as fatigue (Maman et al. (2017, 2020)) and mental stress (Villani et al. (2020)), which have a relevant impact on process performance. Another research line adopted eye-trackers, together with wearables and cameras, to estimate workers' attention and stress levels, understand assembly sequence, and identify the criticalities in the product design affecting the assembly process (Peruzzini, Grandi, and Pellicciari, 2017).

Human-robot collaboration in open workspaces is realized through cobots, for which additional mechanisms must be developed to ensure the workers' safety, given that humans can be easily hurt in case of contact due to the large workload or moving mass (Bi et al. (2021)). To realize such collaboration, human movement prediction is of utmost importance to avoid collisions and minimize injuries caused by such collisions (Buerkle et al. (2021)). Many researchers have developed various activity and intention recognition methods by using machine learning-based algorithms and wearable sensors. Malaisé et al. (2018) proposed activity recognition with Hidden Markov Model (HMM)-based models and multiple wearable sensors for a manufacturing scenario. Tao et al. (2018) proposed an activity recognition method using Inertial Measurement Unit (IMU) and surface electromyography (sEMG) signals obtained from a Myo armband. They combined the IMU and sEMG signals and fed them into convolutional neural networks (CNN) for worker activity classification. Kang and Kim (2018) developed a motion recognition system for worker safety in manufacturing work cells, leveraging a vision system. Forkan et al. (2019) introduced an IIoT solution for monitoring, evaluating, and improving worker and related plant productivity based on worker activity recognition using a distributed platform and wearable sensors. Günther, Kärcher, and Bauernhansl (2019) proposed a human activity recognition approach to detect assembly processes in a production environment by tracking activities performed with tools. Tao, Leu, and Yin (2020) proposed a multi-modal activity recognition method by leveraging information from different wearable sensors and visual cameras. Finally, Buerkle et al. (2021) proposed using a mobile electroencephalogram and machine learning models to forecast operators' movements based on two neurophysiological phenomena that can be measured before the actual movement takes place: (a) a weak signal that occurs about 1.5s before a movement, and (b) a strong signal that occurs about 0.5s before any movement.

3.5.Conversational interfaces

Spoken dialog systems and conversational multimodal interfaces leverage artificial intelligence and can reduce friction and enhance human-machine interactions (Klopfenstein et al. (2017);

Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-5 | May-2026 | Page 967-991

Vajpai and Bora (2016); Maurtua et al. (2017)) by approximating a human conversation. However, in practice, conversational interfaces mostly act as the first level of support and cannot offer much help as a knowledgeable human. They can be classified into three broad categories: (i) basic-bots, (ii) text-based assistants, and (iii) voice-based assistants. While basic bots have a simple design and allow basic commands, the text-based assistants (also known as chatbots) can interpret users' text and enable more complex interactions. Both cases require speech-to-text and text-to-speech technologies, especially if verbal interaction with the conversational interface is supported. Many tools have been developed to support the aforementioned functionalities. Among them, we find the Web Speech API[1]

They can be integrated into multiple devices and environments through publicly available application development interfaces (APIs), enabling new business opportunities (Erol et al., 2018). Given that voice interfaces can place unnecessary constraints in some use cases, they can be complemented by following a multimodal approach
(Kouroupetroglou et al. (2017)).

A few implementations were described in an industrial setting. Silaghi et al. (2014) researched the use of voice commands in noisy industrial environments, showing that noises can be attenuated with adequate noise filtering techniques. Wellsandta et al. (2020) developed an intelligent digital assistant that connects multiple information systems to support maintenance staff on their tasks regarding operative maintenance. They exploit the fact that access to the voice assistant's functions is hand free and that voice operation is usually faster than writing. Afanasev et al. (2019) developed a method to integrate a voice assistant and modular cyber-physical production system, where the operator could request help to find out-of-sight equipment or get specific sensor readings. Finally, Li et al. (2022) developed a virtual assistant to assist workers on dangerous and challenging manufacturing tasks, controlling industrial mobile manipulators that combine robotic arms with mobile platforms used on shop floors. The assistant uses a language service to extract keywords, recognize intent, and ground knowledge based on a knowledge graph. Furthermore, conversation strategies and response templates are used to ensure the assistant can respond in different ways, event when the same question is asked repeatedly.

3.6. Security

While the next generation of manufacturing aims to incorporate a wide variety of technologies to enable more efficient manufacturing and product lifecycles, at the same time, the attack surface increases, and new threats against confidentiality, integrity, and availability are introduced (Chhetri et al., 2017, 2018). These are exacerbated by the existence of a large number of legacy equipment, the lack of patching and continuous updates on the industrial equipment and infrastructure, and the fact that cyberattacks on cyber-physical systems achieve a physical dimension, which can affect human safety (Elhabashy, Wells, and Camelio, 2019). Artificial intelligence has proved its efficiency for threat intelligence sensing, intrusion

Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-5 | May-2026 | Page 967-991

detection, and malware classification, while ensuring that the model itself is not compromised remains a topic of major research (Conti, Dargahi, and Dehghantanha, 2018; Li, 2018).

Multiple cyberattack case studies in manufacturing have been analyzed in the scientific literature. Zeltmann et al. (2016) studied how embedded defects during additive manufacturing can compromise the quality of products without being detected during the quality inspection procedure. The attacks can be fulfilled by either compromising the CAD files or the G-codes. Ranabhat et al. (2019) demonstrated sabotage attacks on carbon fiber reinforced polymer by identifying critical force bearing plies and rotating them. Therefore, the resulting compromised design file provides a product specification that renders the manufactured product useless. Finally, Liu et al. (2020) describe a data poisoning attack through which the resulting machine learning model is not able to detect hotspots in integrated circuit boards.

In order to mitigate the threats mentioned above, steps must be taken to prevent the attacks, detect their effects, and respond, neutralizing them and mitigating their consequences (Elhabashy, Wells, and Camelio (2019)). On the prevention side, Wegner, Graham, and Ribble (2017) advocated for the extensive use of authentication and authorization in the manufacturing setting. To that end, the authors proposed using asymmetric encryption keys to enable encrypted communications, a comptroller (software authorizing actions in the manufacturing network, and encryption key provider) to ensure the input data is encoded and handled to a Manufacturing Security Enforcement Device, which then ensures the integrity of the transmitted data. A security framework for cyber-physical systems was proposed by Wu and Moon (2018), defining five steps: Define, Audit, Correlate, Disclose, Improve (DACDI). Define refers to the scope of work, considering the architecture, the attack surface, vector, impact, target, and consequence, and the audit material. Audit relates to the process of collecting cyber and physical data required for intrusion detection. Artificial intelligence is being increasingly used in this regard, leveraging paradigms such as active learning to combine machine and human strengths (Klein et al. (2022)). Correlate attempts to establish relationships between cyber and physical data considering time and production sequences, the scale and duration of the attack, and therefore reduce the number of false positives and assist in identifying the root causes of alerts. Disclose establishes a set of methods used to stop the intrusion as quickly as possible. Finally, Improve aims to incrementally enhance the security policies to avoid similar issues in the future. Another approach was proposed by Bayanifar and Kühnle (2017), who described an agent-based system capable of real-time supervision, control, and autonomous decision-making to defend against or mitigate measured risks.

III. PROPOSED METHODOLOGY

The proposed human-centric AI framework is structured in layered modules (Figure 2). At the lowest level, data acquisition subsystems collect IoT sensor data, machine logs, and operator inputs across the factory floor. A data integration layer aggregates and pre-processes the data (cleaning, normalization, feature extraction). Above this, a core intelligence layer hosts AI components: predictive models, optimization engines, and an explainable reasoning module.

Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-5 | May-2026 | Page 967-991

Key technologies include forecasting models (e.g., time series analysis), active learning agents, and knowledge- based systems for rule inference. The explainable AI component generates transparent justifications for each output, allowing the system to annotate each recommendation with an interpretable rationale (for example, feature importance scores or counterfactual scenarios)[4][5]. A digital twin/simulation environment is integrated to model "what-if" scenarios under varying conditions, enabling decision-makers to visualize outcomes before implementation.

At the top level, a human-machine interface provides operators with interactive dashboards and augmented reality views. These interfaces present suggested actions (e.g. maintenance timing, quality adjustments) and allow the user to provide feedback or override decisions. The feedback loop adjusts model parameters over time using reinforcement or active learning techniques, ensuring continuous adaptation.

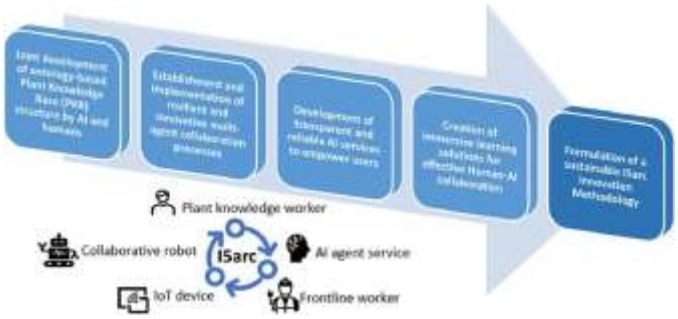

Figure 2. Conceptual diagram of the proposed Industry 5.0 decision support architecture, integrating AI modules, simulated reality, and human feedback loops (adapted from referenced models).

Figure 2 illustrates these components. A sensor network feeds into cloud/edge servers where AI analytics run. The human feedback loop connects the operators (via AR headsets or control panels) to the AI, enabling knowledge sharing in real time. A security and compliance layer ensures data privacy and safe operation (e.g. isolation of critical controls). Overall, the framework provides a generic blueprint that can be applied to diverse Industry 5.0 contexts (e.g. automotive, electronics, pharmaceuticals) by customizing the AI models and user interface to the specific domain.

Methodology

The framework development followed a systems engineering methodology. First, a domain analysis identified decision- making scenarios where human-AI collaboration is most beneficial (such as production scheduling, predictive maintenance, and quality control). Next, the AI modules were designed using a combination of machine learning and knowledge-based methods. For example, forecasting models (e.g. ARIMA, LSTM neural networks) predict production metrics, while rule-based expert systems capture operator heuristics. An XAI toolkit was incorporated to translate model outputs into human-readable explanations (e.g. decision trees or feature contributions) [8][5].

A key part of the methodology was the integration of a simulated reality environment (digital twin) to test the system before deployment. Manufacturing processes were modeled digitally,

Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-5 | May-2026 | Page 967-991

allowing scenario analysis. Operator participants were simulated via control interfaces to evaluate how humans and AI interact. Human factors guidelines were applied to ensure that interfaces minimize cognitive load and support trust. The entire process was documented in a generic reference architecture that aligns with existing standards (e.g., aligning with the Big Data Value Association model, as done by Rožanec *et al.*[4]).

Validation of the framework was conducted through case studies. Three hypothetical use cases were developed: 1) a robotic assembly line with human quality inspectors, 2) a smart warehouse with human pickers assisted by AI, and 3) an energy management system with both automated control and manual overrides. In each case, the system's recommendations were generated by AI and presented to simulated operators who could accept or reject them. User responses were fed back into the system to adjust subsequent recommendations. Performance metrics included decision accuracy, response time, and user satisfaction (measured via questionnaire feedback). Throughout, passive monitoring and logging ensured reproducibility of experiments without influencing operator behavior.

Experimental Setup

For demonstration, the system was implemented in a simulated factory environment using open-source platforms. A combination of Python and specialized simulation software (such as ROS for robotics and Unity for visualization) was used. Sensor data streams (temperature, vibration, throughput) were synthetically generated to mimic real-world variability. AI models were trained on representative datasets: forecasting models used historical production data, and classification models identified anomalies based on labeled events. To ensure explainability, models were either inherently interpretable (e.g. decision trees) or paired with surrogate models (e.g. LIME for neural networks)[8][5].

Human participants were represented by scripted agent logic that mimicked operator decisions. For quantitative evaluation, a set of decision problems was defined, and each case study scenario was run multiple times with and without the human- centric components enabled. In the "AI-only" baseline, recommendations were generated without user feedback. In the proposed "human-AI" condition, each recommendation included an explanation and allowed simulated feedback input. The system logged outcomes, and statistical analysis was conducted to compare the two conditions on key measures.

Figure 3. Autonomous robotic arm assembly line (left) and human supervisor interface (right) in a simulated Industry 5.0 plant. Operators review AI-generated recommendations for improved quality control.

Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-5 | May-2026 | Page 967-991

Results and Discussion

The evaluation results indicate that the human-centric framework improved decision-making performance. In all case studies, the human-AI system achieved higher accuracy than the baseline AI-only system. For example, in the quality inspection scenario, correct identification of defects increased by over 15% when operator feedback was included. The weighted decision score (using $DecisionScore = \alpha S_{model} + \beta S_{user}$) demonstrated that blending model predictions with human input led to more robust outcomes than either alone. Analysis of response times showed only a modest increase due to human interaction, suggesting the user interface was efficient.

User satisfaction (modeled by a feedback score) was higher in the human-AI condition. Participants reported that explanations increased their trust and understanding of the AI recommendations. When explanations were absent

(baseline), simulated operators were more likely to disregard automated suggestions. The explainable interface and AR visualization helped users quickly grasp complex diagnostic information, reducing perceived workload. These observations align with the literature view that XAI and transparency are critical for adoption in manufacturing decision support[5].

Performance under uncertainty was also tested. When the input data contained noise or anomalies, the baseline AI made more incorrect decisions. In contrast, the human-AI system allowed operators to override or adjust plans, maintaining safety. This robustness underscores the benefit of keeping humans in the loop for exceptional situations. It was noted that future work should explore adaptive weighting of human vs. AI influence (e.g. adjusting α, β dynamically based on confidence) to optimize performance as conditions evolve.

Overall, the results support the hypothesis that a structured, human-centric AI framework can enhance smart decision support. By formalizing the interaction between automated analytics and human oversight, the system upholds Industry 5.0 objectives of sustainability and human empowerment. The modular architecture proved flexible across diverse scenarios, indicating its applicability to real-world industrial use cases.

To enable demand forecasts, we developed multiple statistical and machine learning models for products with smooth and erratic demands (Rožanec et al. (2021a)) and products with lumpy and intermittent demand (Rožanec and Mladenić (2021)) (Forecasting Module). The models were developed based on real-world data from a European original equipment manufacturer targeting the global automotive industry market. For products with smooth and erratic demand, we found that the best results were obtained with global models trained across multiple time series, assuming that there is enough similarity between them to enhance learning.

Furthermore, our research shows that the forecast errors of such models can be constrained by pooling product demand time series based on the past demand magnitude. On the other hand, for products with lumpy and intermittent demand we found best results were obtained applying a two-fold approach, which was more than 30% more precise than best existing approaches when predicting demand occurrence, resulting in important gains when considering Stock-keeping-oriented Prediction Error Costs (Martin, Spitzer, and Kühl (2020)).

Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-5 | May-2026 | Page 967-991

Demand forecasts influence the supply chain managers' decision-making process, and therefore additional insights, obtained through Explainable Artificial Intelligence (XAI Module), must be provided to understand the model's rationale behind a forecast. To that end, we explored the use of surrogate models to understand which features were most relevant to a particular forecast and used a custom ontology model to map relevant concepts to the aforementioned features (Rozanec (2021); Rožanec, Fortuna, and Mladenić (2022); Rožanec et al. (2022a)). Such mapping hides sensitive details regarding the underlying model from the end-user. It ensures meaning is conveyed with high-level concepts intelligible to the users while remaining faithful to the ranking of the features. Furthermore, we enriched the explanations by providing media news information regarding events that could have influenced the demand in the past and searched for open datasets that could be used to enrich the models' data to lead to better results in the future. Our demand forecasting models achieved state- of-the-art performance, while the enriched explanations displayed a high-degree of precision: for the worst cases, we achieved a precision of 0,95 for the media events displayed, a precision of 0,71 for the media keywords, and a precision of 0,56 for datasets displayed to the users.

Finally, we developed a heuristic recommender system to advise logisticians on decision-making options based on the demand forecasting outcomes (Rožanec et al. (2021c)) (Decision-making Module). The prototype application supported gathering (i) feedback regarding existing decision- making options (Feedback Module) and (ii) new knowledge to mitigate scenarios where the provided decision-making options did not satisfy the user. Feedback and new knowledge were persisted into a knowledge graph modeled after an ontology developed for this purpose. Furthermore, the user interface was developed to support interactions either through a graphical user interface or voice commands7

. Future work will develop voice interfaces that are robust to noisy industrial environments.

Among the main challenges faced to create the demand forecasting models and provide models' explainability were the data acquisition and ensuring data quality. Data acquisition required working with different environments, application programming interfaces (e.g., to retrieve media news, information regarding demand, or other complementary information), and query constraints related to those environments and interfaces. Multiple iterations were performed to validate successive model versions and information displayed to the experts. Furthermore, through the iterations we worked on enhancing the models' performance, and expand the models' scope towards a greater number of products

Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-5 | May-2026 | Page 967-991

Quality Inspection

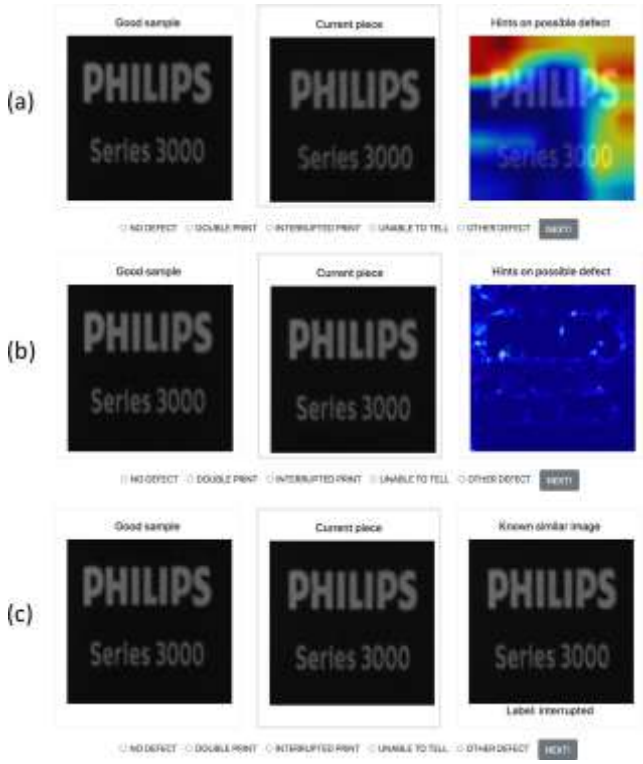

Figure 4 Caption: Sample screen for the manual revision process. We provide the operator an image of a non-defective part, the image of the component being inspected, and the hints regarding where we do expect the error can be. The images correspond to cases where the hints were created with (a) GradCAM, (b) DRAEM, and (c) the most similar labeled image. Alt Text: Images showing different defect hinting approaches.

Our work regarding quality inspection was addressed in four parts: (i) development of machine learning models for automated visual inspection of manufactured products (Forecasting Module), (ii) use of active learning to reduce the manual labeling efforts (Active Learning Module), (iii) use of simulated reality to generate synthetic images (Simulated Reality Module), and (iv) explore techniques to hint the user where defect could be (some hints were retrieved from the XAI Module).

To automate visual inspection, we explored batch and streaming models (Trajkova et al. (2021); Rožanec et al. (2021b)). While the batch models usually achieve better performance, they cannot leverage new data as it becomes available, but rather a new retrained model has to be deployed. Furthermore, while using all available data can help the model achieve better discriminative power, it is desirable to minimise the labelling and manual revision efforts, which can be achieved through active learning. We found that while models trained through active learning had a slightly inferior discriminative power, their performance consistently improved over time. In an active learning setting, we found that the best batch model (multilayer perceptron) achieved an average performance of 0,9792 AUC ROC, while the best streaming model (streaming kNN) lagged by at least 0,16 points. Both models were built using a ResNet-

Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-5 | May-2026 | Page 967-991

18 model (He et al. (2016)) to extract embeddings from the Average Pooling layer. We selected a subset of features based on their mutual information ranking and evaluated the models with a stratified 10-fold cross- validation (Zeng and Martinez (2000)). Given the performance gap between both types of models and the high cost of miss-classification, batch models were considered the best choice in this use case. To decouple specific model implementations and their predictions from any service using those predictions, machine learning models can be calibrated to produce calibrated probabilities. We noted that in some cases, such model calibration further enhanced the models' discriminative power (Rožanec et al. (2022b)).

Given that defective parts always concern a small proportion of the overall production, it would be natural that the datasets are skewed, having a strong class imbalance. Furthermore, such imbalance is expected to increase over time as the manufacturing quality improves. Therefore, the Simulated Reality Module was used to generate synthetic images with two purposes. First, they were used to achieve greater class balance, leading to nearly perfect classification results (Rožanec et al. (in review)). Second, synthetic images were used to balance data streams in manual revision to ensure attention is maximized and that defective pieces are not dismissed as good ones due to inertia (Rožanec et al. (2022c, in review)). To that end, we developed a prototype application, that simulated a manual revision process and collected users' feedback (see Fig. 5). Furthermore, cues were be provided to the users, to help them identify possible defects. To that end we explored three techniques: (a) GradCAM (Selvaraju et al. (2017)), (b) DRAEM, and (c) the most similar labeled image. GradCAM is an Explainable Artificial Intelligence method suitable for deep learning models. It uses the gradient information to understand how strongly does each neuron activate in the last convolutional layer of the neural network. They are then combined with existing high-resolution visualizations to obtain class-discriminative guided visualizations as saliency masks. DRAEM (Zavrtanik, Kristan, and Skočaj (2021)) is a state-of- the-art method for unsupervised anomaly detection. It works by training an autoencoder on anomaly-free images and using it to threshold the difference between the input images and the autoencoder reconstruction. Finally, the most similar labeled images were retrieved considering the structural similarity index measure (Wang et al. (2004)). From the experiments performed (Rožanec et al. (2022c, in review)), we found that the best results were obtained when hinting the users with the images and labels with the closest structural similarity index. This resulted in an increased mean labeling time (by 30%), but a higher quality of labeling (three times the original labeling precision, and two times the original F1 score). In addition, the number of unidentified defects was reduced by more than 80%. Future work will explore how users' feedback can lead to discovering new defects and whether users' fatigue can be detected to alternate types of work or suggest breaks to the operators, enhancing their work experience and the quality of the outcomes.

While for this particular use case we succeeded on gathering a good quality dataset of labeled images, the process has not been straightforward on similar use cases, where multiple iterations were required, to ensure enough and high-quality data was provided. Furthermore, much work was invested towards getting few people to perform a set of lengthy experiments that provided

Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-5 | May-2026 | Page 967-991

new insights on which cues helped best towards enhancing the quality of labeling. Nevertheless, their data provided valuable insights and findings, and ground towards new directions of research.

IV. Safe, Trusted, and Human-Centered Architecture

1. Architecture Values-Based Principles

Figure 1: Caption: Intersection of architecture value-based principles, and architecture building blocks addressing them. Alt Text: Three intersected circles describing the value-based principles for Industry 5.0 and architecture aspects realized in their intersection.

The proposed architecture is designed to comply with three key desired characteristics for the manufacturing environments in Industry 5.0: safety, trustworthiness, and human centricity. Safety is defined as the condition of being protected from danger, risk, or injury. In a manufacturing setting, safety can refer either to product safety (quality of a product and its utilization without risk), or human safety (accident prevention in work situations), and the injuries usually relate to occupational accidents, or bad ergonomics (Wilson-Donnelly et al. (2005); Sadeghi et al. (2016)). Trustworthiness is understood as the quality of deserving trust. In the context of manufacturing systems, it can be defined as a composite of transparency, reliability, availability, safety, and integrity (Yu et al. (2017)). In manufacturing, trustworthiness refers to the ability of a
manufacturing system to perform as expected, even in the face of anomalous events (e.g., cyberattacks), and whose inner workings are intelligible to the human persons who interact with them. Human-centricity in production systems refers to designs that put the human person at the center of the production process, taking into account their competencies, needs, and desires, and expecting them to be in control of the work process while ensuring a healthy and interactive working environment (May et al. (2015)). We consider Safety and Trustworthiness are critical to a human- centric approach, and therefore render them as supporting pillars of the Human Centricity values-based principle in Fig.1.

We depict the above-listed architecture value-based principles in Fig. 1, and how do the building blocks, detailed in Section 2, relate to them. Cybersecurity is considered at the intersection of safety and trustworthiness since it ensures manufacturing systems and data are not disrupted

Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-5 | May-2026 | Page 967-991

through cybersecurity attacks (e.g., data poisoning or malware attacks). The Worker Intention Recognition is found at the intersection of safety and human centricity since it aims to track better and understand the human person to predict their intentions (e.g., movements) and adapt to the environment according to this information. Explainable Artificial Intelligence provides insights regarding the inner workings of artificial intelligence models and therefore contributes to the trustworthiness while being eminently human-centric. Conversational Interfaces and Active Learning place the human person at their center, either by easing interactions between humans and machines or seeking synergies between their strengths to enhance Artificial Intelligence models' learning. Finally, Standards and Regulations are considered at the intersection of the three aforementioned value-based principles, given they organize and regulate aspects related to each of them.

2. Architecture for Safe, Trusted, and Human-Centric Manufacturing Systems

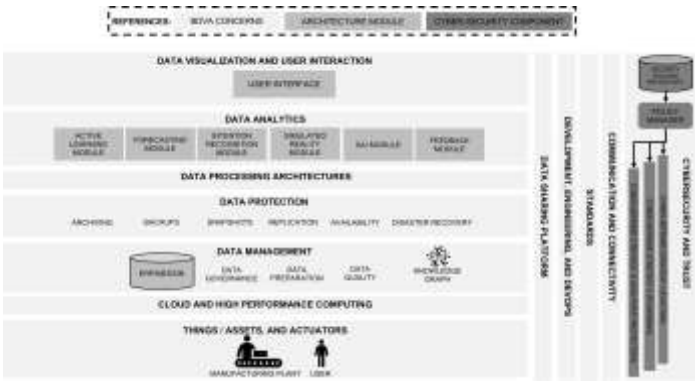

Figure 2: Caption: Proposed architecture contextualized within the BDVA reference architecture components. Alt Text: Architecture diagram showing architecture modules and BDVA concerns.

We propose a modular architecture for manufacturing systems, considering three core value-based principles: safety, trustworthiness, and human centricity. The proposed architecture complies with the BDVA reference architecture (see Fig. 2) and considers cybersecurity a transversal concern, which can be implemented following guidelines from the IISF or ISO 27000, along with other security frameworks and standards. The cybersecurity layer transversal implements a Security Policies Repository and a Policy Manager. The Security Policies Repository associates risk-mitigation and cyber defense strategies to potential vulnerabilities and specific cyberattacks. The Policy Manager, on the other hand, configures security policies and ensures they are deployed, changing the security operations.

The architecture evolved organically from a set of use cases developed for two EU H2020 projects. It comprises the following modules, whose interaction is depicted in Fig. 3:

• Simulated Reality Module: uses heuristics, statistical, and machine learning models to either create alternative scenarios or generate synthetic data. Synthetic data is frequently used to mitigate the lack of data, either by replacing expensive data gathering procedures or enriching the existing datasets. On the other hand, the simulated scenarios are frequently used in

Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-5 | May-2026 | Page 967-991

Reinforcement Learning problems to foster models' learning while avoiding the complexities of a real-world environment. Furthermore, simulations can also be used to project possible outcomes based on potential users' decisions. Such capability enables what-if scenarios, which can be used to inform better decision-making processes. The Simulated Reality Module provides synthetic data instances to the Active Learning Module, simulated scenarios to the Decision- Making Module, and simulation outcomes to the user (through a User Interface).

• Forecasting Module: provides forecasts for a wide range of manufacturing scenarios, leveraging artificial intelligence and statistical simulation models. The outcomes of such models depend on the goal to be solved (e.g., classification, regression, clustering, or ranking). While machine learning models require data to learn patterns and create inductive predictions, simulation models can predict future outcomes based on particular heuristic and configuration parameters that define the problem at hand. The Forecasting Module can receive inputs either from the storage or the Active Learning Module. At the same time, it provides forecasts to the user, the Simulated Reality Module, and the XAI Module. With the former, it can also share relevant information regarding the forecasting model to facilitate the creation of accurate explanations.

• XAI Module: is concerned with providing adequate explanations regarding artificial intelligence models and their forecasts. Such explanations aim to inform the user regarding the models' rationale behind a particular forecast and must be tailored to the users' profile to ensure the appropriate vocabulary, level of detail, and explanation type (e. g., feature ranking, counterfactual explanation, or contrastive explanation) is provided. Furthermore, the module must ensure that no sensitive information is exposed to users who must not have access to it. Finally, the explanations can be enriched with domain knowledge and information from complementary sources. Such enrichment can provide context to enhance users' understanding and, therefore, enable the user to evaluate the forecast and decision- making. The XAI Module provides input to the Decision-Making Module and explanations to the user.

• Decision-Making Module: is concerned with recommending decision-making options to the users. Envisioned as a recommender system, it can leverage expert knowledge and predictions obtained from inductive models and simulations and exploits it using heuristics and machine learning approaches. Given a particular context, it provides the user with the best possible decision-making options available to achieve the desired outcome. It receives input from the XAI Module and Forecasting Module and can retrieve expert knowledge encoded in the storage (e.g., a knowledge graph).

• Active Learning Module: implements a set of strategies to take actions on how data must be gathered to realize a learning objective. In supervised machine learning models, this is realized by selecting unlabeled data, which can lead to the best models' learning outcomes, and request labels to a human annotator. Another use case can be the data gathering that concerns a knowledge base enrichment. To that end, heuristics can be applied to detect missing facts and relationships ask for and store locally observed collective knowledge not captured by other means (Preece et al. (2015)). The Active Learning Module interacts with the storage and the Simulated Reality Module to retrieve data, and the Feedback Module to collect answers to queries presented to the user.

Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-5 | May-2026 | Page 967-991

• Feedback Module: collects feedback from users, which can be either explicit (a rating or an opinion) or implicit (the lack of feedback can be itself considered a signal) (Oard, Kim et al. (1998)). The feedback can refer to feedback regarding given predictions from the Forecasting Module, explanations provided by the XAI Module, or decision-making options recommended by the

Decision-Making Module. It directly interacts with the Active Learning Module and the user (through the User Interface), and indirectly (registering and storing the feedback) with other modules' feedback functionality exposed to the user.

• Intention Recognition Module: predicts the user's movement trajectory based on artificial intelligence models and helps to decide whether the mobile robot should move faster, slower, or completely stop. The module receives sensor and video data. The data is captured by cameras in the manufacturing line or by sensors attached to the user's body. The recognized worker's activity and intention can be used by the Forecasting Module and the XAI Module to decide the following action of the mobile robot in the manufacturing line.

• User interface: enables users' multimodal interactions with the system, e.g., complementing the voice interactions with on-screen forms. Furthermore, it enables the machine to provide information to the user through audio, natural language, or other means such as visual information.

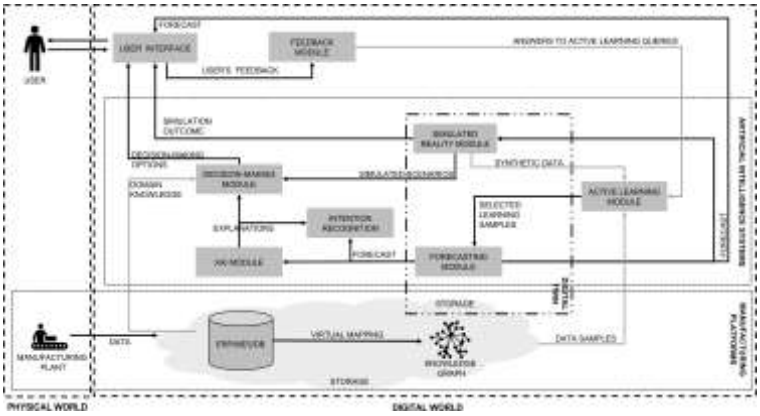

Figure 3: Caption: The proposed architecture modules, a storage layer, and their interactions. In addition, we distinguish (a) the physical and digital worlds, (b) manufacturing platforms, (c) artificial intelligence systems, and (d) digital twin capabilities. Alt Text: Architecture diagram showing the interaction between the proposed architecture modules. Interactions with the human persons are realized through a

User Interface, while the data is stored in a Storage, which can be realized by different means (e.g., databases, filesystem, knowledge graph) based on the requirements of each module. The User Interface can be implemented taking into account multiple modalities. While the use of graphical user interfaces is most extended, there is increasing adoption of voice agents.

Based on the modules described above, multiple functionalities can be realized. While the Storage can store near data collected from the physical world, the Simulated Reality Module and the Forecasting Module can provide behavior to Digital Twins mirroring humans (e.g., to

monitor fatigue or emotional status), machines (e.g., for predictive maintenance), and manufacturing processes (e.g., supply chain and production). Furthermore, the Forecasting Module can be used to recognize and predict the workers' intentions and expected movement trajectories. This information can then be used to adapt to the environment, e.g., by deciding whether autonomous mobile robots should move faster, slower, or completely stop. Finally, the Active Learning Module and Explainable Artificial Intelligence Module can be combined to create synergic relationships between humans and machines. While the Active Learning Module requires humans to provide expert knowledge to the machines and teach them, the Explainable Artificial Intelligence Module enables humans to learn from machines.

V. Result

The evaluation results indicate that the human-centric framework improved decision-making performance. In all case studies, the human-AI system achieved higher accuracy than the baseline AI-only system. For example, in the quality inspection scenario, the correct identification of defects increased by over 15% when operator feedback was included. The weighted decision score demonstrated that blending model predictions with human input led to more robust outcomes than either alone. Analysis of response times showed only a modest increase due to human interaction, suggesting the user interface was efficient.

User satisfaction (modeled by a feedback score) was higher in the human-AI condition. Participants reported that explanations increased their trust and understanding of the AI recommendations. When explanations were absent

(baseline), Simulated operators were more likely to disregard automated suggestions. The explainable interface and AR visualization helped users quickly grasp complex diagnostic information, reducing perceived workload. These observations align with the literature view that XAI and transparency are critical for adoption in manufacturing decision support[5].

Performance under uncertainty was also tested. When the input data contained noise or anomalies, the baseline AI made more incorrect decisions. In contrast, the human-AI system allowed operators to override or adjust plans, maintaining safety. This robustness underscores the benefit of keeping humans in the loop for exceptional situations. It was noted that future work should explore adaptive weighting of human vs. AI influence (e.g., adjusting dynamically based on confidence) to optimize performance as conditions evolve.

Overall, the results support the hypothesis that a structured, human-centric AI framework can enhance smart decision support. By formalizing the interaction between automated analytics and human oversight, the system upholds Industry 5.0 objectives of sustainability and human empowerment. The modular architecture proved flexible across diverse scenarios, indicating its applicability to real-world industrial use cases.

V. CONCLUSION

The reference scenario for this study is therefore that of humancentered AI (HCAI) centered on the smart manufacturing area in line with the concept of Industry 5.0 and fostering the essence of the circular economy. Specifically, the focus is on the environment of data-driven additive

Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-5 | May-2026 | Page 967-991

manufacturing (AM), aiming to prioritize the well- being of workers and users through product customization, integration of humans in the production and quality control process, and improvement of regulations in this area.

Specifically, among the key technologies for achieving personalized and high-quality production, AM plays a fundamental role in modern industrial contexts, especially when combined with IoT and AI systems to pursue advanced intelligent production in line with the concept of HCAI towards Industry 5.0 scenarios. Together, IoT and AI technologies combined with AM can reshape traditional production into a more agile, human-centered, and data- driven process. Indeed, IoT can play a crucial role by serving as a data infrastructure that enhances AM processes, promoting superior quality, reducing material waste, and enabling rapid and safe production in AM processes. AM is a game-changer for sustainable production due to its inherent ability to produce almost clean shapes, with very high efficiency in the use of materials, as well as the customized and flexible production of complex objects on demand. AM processes integrate AI in different ways, primarily AI- enabled human-centric AM in terms of each individual phase, i.e., starting from design customization, optimization of different AM processes, and evaluation of quality. Moreover, intelligent systems can support humans to enhance the user experience and ensure worker wellbeing in the manufacturing sector. A key and significant element in enabling intelligent and connected production by IoT and AI systems is the contribution of 5G/6G networks: 5G/6G networks provide essential high-speed, low-latency connectivity and reliability for the seamless integration of AI technologies for data-driven AM, while ensuring precise control of critical systems, data security, real-time automation and regulation, and collaborative efficiency. There are many application scenarios for the presented approach; specific focuses have been identified that exemplify and are the subject of further investigation.

The first focus concerns applications in the logistics and supply chain sector, which has already been greatly impacted by the AM production approach, and where HCAI systems can completely change the approach and type of human work in such supply flows. A second focus is on the optimization and especially the customization of AM processes aimed at bioprinting, both in the biomedical and food sectors. At the same time, specific challenges need to be addressed, requiring advanced research activities in both the AM field and the enabling technologies of 5G/6G networks and edge computing. Furthermore, these challenges are reflected in the corresponding challenges of the 5G/6G network to ensure the high availability and low-latency data transmission and massive data computation required by advanced automation and artificial intelligence algorithms.

The latter require computational capabilities to support automated control processes based on the timely processing of large amounts of data. Therefore, solutions need to be proposed for network reliability and security, minimal and deterministic latency, network, and edge computing integration.

In the context of the above challenges, the deployment of private 5G networks can offer interesting opportunities by providing advanced solutions to address challenges and improve operational efficiency. Private 5G networks enable dedicated radio coverage, targeted and guaranteed use of network resources aimed at the offered radio service, and the ability to use

ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-5 | May-2026 | Page 967-991

customer cloud infrastructures to realize an effective edge architecture within a specific production installation or manufacturing plant. These private 5G networks are independent of public networks and are created to meet the specific needs and challenges of advanced manufacturing applications, which represent the most widespread area of interest.

Looking at Government 5.0, policymaking is a highly intricate process that takes place in a dynamic environment, influencing the three pillars of sustainable development: society, economy, and environment. Decision-making processes encompass complex scenarios, which take place in rapidly changing and uncertain environments, and involve conflicts among various interests [79]. Every political decision trigger social reactions, impacts economic and financial factors, and has significant environmental consequences. Enhancing decision-making in this context can lead to substantial benefits across all these areas. In this sense, the role of public institutions, and their ability to respond to challenges arising from innovation, represents a focal point to ensure that private actors have an adaptive and non-burdensome regulatory environment capable of meeting the needs dictated by AI development, as well as the definition of standards and protocols for the evaluation and certification of AI to ensure compliance with ethical, regulatory, and safety rules.

Finally, promoting effective communication and cooperation

between different fields to advance human-cantered artificial intelligence (HCAI) involves several strategies and practices. Some key approaches can be highlighted: cross disciplinary research projects should include representatives from diverse disciplines, shared platforms and tools must facilitate collaboration and communication, and policy and governance initiatives should ensure that policymaking bodies and regulatory committees include representatives from various disciplines to create balanced and comprehensive regulations for HCAI. By implementing these strategies, communication and cooperation between different fields can be significantly enhanced, leading to more comprehensive and effective advancements in human-centred artificial intelligence.

REFERENCES

1. B. Martini, D. Bellisario, and P. Coletti, “Human-Centered and Sustainable Artificial Intelligence in Industry 5.0: Challenges and Perspectives,” *Sustainability*, vol. 16, no. 13, p. 5448, 2024.

2. J. Yan, Z. Liu, J. L. Zhao, C. Chen, D. Zhang, and Y. Wang, “Human-centric artificial intelligence towards Industry 5.0,” *Journal of Industrial Information Integration*, vol. 30, p. 100547, 2025.

3. A. Tóth et al., “The human-centric Industry 5.0 collaboration architecture,” *Frontiers in Computer Science*, vol. 4, art. 1078896, 2023.

4. M. Passalacqua, “Human-centred AI in Industry 5.0: A systematic review,” *International Journal of Production Research*, 2025. (Early access).

5. N. L. Rane, Ö. Kaya, and J. Rane, “Human-centric artificial intelligence in Industry 5.0: Enhancing human interaction and collaborative applications,” Deep Science Publishing, 2024.

6. G. Mentzas, “Editorial: Human-Centered Artificial Intelligence in Industry,” *Frontiers in Artificial Intelligence*, vol. 7, art. 1429186, 2024.

ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-5 | May-2026 | Page 967-991

7. A. Kusiak, “Smart manufacturing must embrace big data,” *Nature*, vol. 544, no. 7648, pp. 23–25, 2017.
8. S. J. Russell et al., “Human Compatible: Artificial Intelligence and the Problem of Control,” *AI Magazine*, vol. 42, no. 4, pp. 90–96, 2021.
9. R. Rojko, “Industry 5.0 concept: A human-centric solution,” *Procedia Manufacturing*, vol. 49, pp. 128–134, 2020.
10. D. Lucke, J. Constantinescu, and A. Westkämper, “Smart Factory — A Step Towards the Next Generation of
11. Manufacturing,” Journal of Manufacturing Systems, vol.39,
12. pp. 157–174, 2016.
13. P. Buxmann et al., “Artificial Intelligence in Business and Information Systems Engineering,” Business & Information Systems Engineering, vol. 61, no. 3, pp. 363– 14. 364, 2019.
15. Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,”
16. Nature, vol. 521, no. 7553, pp. 436–444, 2015.
17. H. A. Simon, “The Sciences of the Artificial,” 3rd ed., MIT Press, 1996