Environment awareness, multimodal interaction, and intelligent assistance in industrial augmented reality solutions with deep learning

Juan Izquierdo-Domenech[1*], Jordi Linares-Pellicer[1†] and Isabel Ferri-Molla[1†]

[1*]Valencian Research Institute for Artificial Intelligence (VRAIN), Universitat Politècnica de València, Camí de Vera, s/n, València, 46022, València, Spain.

*Corresponding author(s). E-mail(s): juaizdom@upv.es; Contributing authors: jlinares@dsic.upv.es; isfermol@epsa.upv.es;
†These authors contributed equally to this work.

Abstract

Augmented reality is increasingly used in various fields, especially industrial applications. Although augmented reality devices' characteristics and technological benefits are still evolving, augmented reality's clear advantages in facilitating mechanical tasks and improving operator performance have made it popular. In industrial settings, the human factor remains irreplaceable, but the evolution of artificial intelligence has allowed any activity on the shop floor to be given new semantic possibilities. Through a semantic layer, it is possible to interpret and validate the environment, provide multimodal interaction, and analyze and evaluate information to detect anomalies or risky situations. Deep learning has opened up new possibilities for existing augmented reality solutions, such as visual interpretation of the environment, natural language understanding for problem-solving, or automatic anomaly detection. This new intelligent layer minimizes unnecessary interactions with the environment, validates the operator's actions, and increases comfort, safety, and focus, making them more efficient in high cognitive level tasks. This work presents a general architecture based on a Semantic layer that relies on augmented reality systems and validates its advantages in a real industrial setting. Overall, integrating artificial intelligence and augmented reality solutions in industrial settings offers significant potential for improving productivity, safety, and worker satisfaction.

Keywords: Augmented reality, Multimodal, Deep learning, Transformers, Convolutional neural networks

1 Introduction

Industry 4.0 aims to enhance industrial production efficiency, speed, quality, optimization, and resilience by adopting new technologies [1, 2]. The Industrial Internet of Things (IIoT), additive manufacturing, Artificial Intelligence (AI), cybersecurity, and Augmented Reality (AR) play vital roles in this transformation, with AR particularly emphasizing the human factor [3].

Despite significant automation efforts in Industry 4.0, some tasks on the shop floor still require a human-centered approach and cannot be fully automated [4, 5]. To address this, adopting a Lean philosophy is recommended, where small changes are introduced and evaluated [6]. AR has gained popularity in the industrial field due to its potential to improve various aspects, including assembly, maintenance, training, and waste reduction [5, 7–11]. Since AR strongly focuses on the human factor in the industry, its implementation can benefit operators and task performance.

Accessing documentation can be complex on the shop floor, especially in non- or partially automated industries where technical documents may exist in paper format [12]. The lack of documentation standardization further complicates information retrieval, increasing the mental load on operators, especially in critical situations [13]. Operators may sometimes require support from Subject Matter Experts (SMEs), which may only be feasible occasionally due to cost or availability [14]. Operator attentiveness can also lead to task resolution errors [15]. Traditional problem-solving approaches may not be the most effective strategy for learning new concepts or procedures; instead, reducing cognitive load is crucial for learning [16]. Stress can impact cognition and knowledge acquisition, but controlled exposure can facilitate cognitive function [17]. This study aims to develop a system to reduce cognitive stress for operators facing unfamiliar tasks in a shop floor setting.

Furthermore, the concept of Operator 4.0 aims to establish reliable and interactive relationships between humans and machines, empowering "smart operators" with cutting-edge gadgets and novel skills to exploit Industry 4.0 technologies [18, 19]. Peruzzini et al. highlight these technologies' potential to alleviate the cognitive burden on operators [20]. Additionally, the emerging notion of Operator 5.0 aims to create more intuitive, symbiotic, human-centered, and cognitively supportive computing environments to enhance human adaptation capabilities, productivity, and mental well-being [21]. The integration of technologies like AR and AI is driving the emergence of ”softbots” as virtual systems in computing environments to automate tasks, offer conversation-like interactions, exhibit system intelligence, autonomy, proactivity, and process automation [22, 23].

AI tools such as Machine Learning (ML) and Deep Learning (DL) can complement existing systems to reduce cognitive load, improve context information, and enhance shop floor safety. In cases where older machines lack machine-generated data, visual information becomes crucial. Techniques like Convolutional Neural Networks (CNNs) enable the interpretation of visual cues, such as reading pressure gauge values. Algorithms like K-means [24] and Support Vector Machine (SVM) [25] can check for anomalies and highlight them through AR. Transformers [26] enable operators to ask questions in Natural Language (NL) and receive responses linked to AR systems through visual cues.

The main objective of this study is to enhance current AR solutions by adding cognitive capabilities through DL and foundation models [27]. Recent industry interest has been observed in AR solutions with cognitive capabilities, aiming to extend current AR solutions with a semantic layer, demonstrated in studies [28–31]. These studies illustrate the benefits of integrating environment awareness, multimodal interaction, and cognitive assistance into AR solutions, crucial for extracting information in fast-paced and evolving production environments [32]. This work develops and evaluates a system allowing users to interact multimodally, including NL Question Answering (QA) and contextualized visual indications in the AR environment, enhancing operator comfort and reducing the risk of errors. Integrating AR with AI technologies also impacts aspects such as comfort, security, focus, and knowledge acquisition [33].

The main contributions of this research are:

- To enhance the capabilities of existing AR systems by adding semantic skills to comprehend and interpret the environment,
- To make a profit from ML models to gain insights, such as anomaly detection or information retrieval with transformers, in industrial settings,
- To reduce the cognitive load from operators, improve task performance, and foster technology adoption.

This article is structured as follows: In section 2, the state-of-the-art research related to this work is presented. Then, in section 3, the different components of the proposed system are explained in detail. After that, in section 4, the technologies and their characteristics for implementing the evaluated system are explained. In section 5, a specific case study is performed to evaluate the system's validity, and the results are discussed in section 6. Finally, the conclusions are drawn in section 8.

2 Related work

2.1 AI in the Industrial Field

AI, especially ML, has significantly improved various aspects of the industrial field, including waste reduction, error prevention, enhanced quality, risk prevention, and faster learning [34]. In complex scenarios where traditional analytical methods struggle to relate variables and outcomes, ML proves to be a critical tool [35]. For instance, the "You Only Look Once" (YOLO) algorithm has been utilized by Zamora et al. to detect real-time assembly task actions [36]. ML and DL techniques excel in anomaly detection, outperforming traditional methods [37]. Architectures such as CNNs, RNNs, LSTMs, and AEs are popular for detecting anomalies in unstructured data like images, videos, audio, and time series [38]. Various studies have applied these techniques to detect anomalies in videos and sequential data [39, 40], as well as appearance-based anomalies in videos using pre-trained ResNet-50 models [41]. Additionally, GANs have been used to detect anomalies in image datasets and intrusion networks [42]. Traditional ML algorithms, especially unsupervised training models, have also shown strong performance in anomaly detection [43]. ML and DL are increasingly essential in monitoring and controlling industrial processes, demonstrating their relevance in AR industrial environments [44, 45].

2.2 AR in Industrial Tasks

AR technology has demonstrated great potential in the industrial field by simplifying complex tasks for operators. However, the complexity of the task and the visual cues utilized should be considered to avoid overloading the operator and diluting their focus of attention [46]. Several authors have studied the development and evaluation of AR applications in various industrial tasks, including step-by-step guides [47], product design [48–50], process control [51, 52], maintenance and security [53–60], and operator training [61–63]. A review of the current literature conducted by Bottani et al. in 2019 indicates that the most researched fields regarding AR in the industry are assembly, maintenance, and training [64]. It is essential to consider not only the complexity of the task but also the complexity added by AR to the interface.

2.3 Voice-directed Interfaces and NL Interaction

Voice-directed interfaces are gaining popularity in industrial settings due to their hands-free control and benefits for operators, particularly those with functional diversity [65]. In scenarios where complex machinery requires information from multiple sources, including paper documents, the ability for operators to interact through NL questions proves advantageous [66]. Examples of chatbots being used to train operators and improve usability in industrial settings, as well as their application in Maintenance, Repair, and Overhaul (MRO) tasks, have been documented in the literature [67, 68].

2.4 Cloud Computing Systems for Remote Assistance

Current investments in communication and Cloud Computing systems have spurred the development of remote assistance systems, such as TeamViewer Assist AR [69] and Vuforia Chalk [70], facilitating remote communication between operators and experts. While these AR technologies have been applied to maintenance tasks [53], challenges persist, notably the cost and availability of online Subject Matter Experts (SMEs).

In conclusion, AR, voice-directed interfaces, AI, and Cloud Computing have greatly enhanced various industrial aspects, simplifying tasks, reducing errors, and expediting learning. However, it's crucial to consider task complexity and the potential for AR interfaces to overwhelm operators. ML and DL techniques, especially in anomaly detection, outperform traditional methods. Despite challenges, AR, voice interfaces, and AI hold substantial potential in the industrial field, necessitating further research and development to unlock their capabilities. Our research aims to leverage modern DL techniques and AI foundation models to enhance information access in complex environments.

3 System architecture overview

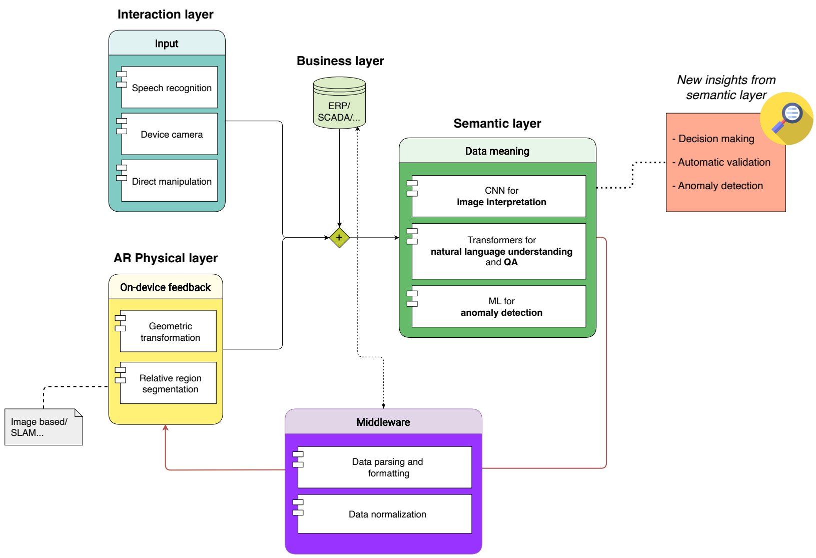


Fig. 1 System architecture layers.

This study proposes an architecture combining AR, ML, and DL techniques to support operators' tasks on a production shop floor. The architecture is designed to enable the creation of applications that offer intelligent assistance to the operator through multimodal interaction, potentially replacing or enhancing the need for an SME, even when using teleoperated AR systems.

The system consists of four communicating layers, as shown in Fig. 1. The central axis is the AR layer, facilitating integration and creating synergy beyond individual components. Further details about each layer are provided below.

3.1 Interaction layer

The Interaction layer provides various modes of operator interaction. Three modules enable multimodal interaction, as depicted in Fig. 1:

- **Speech recognition**: The system recognizes and responds to the operator's NL queries, providing prompt answers based on available technical information, reducing query time significantly, and avoiding the need for consulting SMEs or searching technical documentation.
- **Device camera**: Using a camera is a fundamental element in current AR solutions, available on mobile devices and AR Head-mounted displays (HMD). It supplies the possibility of adding visual cues to the operator and facilitates the validation of performed tasks using AI-based elements described later.
- **Direct manipulation**: This interaction style relates real-world actions to device-based actions [71]. The operator interacts through touch screens of mobile devices and hand gestures in specific AR HMDs. The operator can touch the digital representation of environment elements, such as buttons or AR indicators, to retrieve additional information or perform specific tasks.

The system's multimodality allows users to interact with it in three ways and provides feedback through AR cues or textual information, as explained later.

3.2 AR Physical layer

The AR Physical layer performs crucial tasks, including displaying synthetic content overlaid on the device's camera feedback, determining the operator's location and relative position in the industrial environment, and extracting areas of interest for later analysis. It locates analog controls, like pressure gauges, in the camera's captured image (Relative region segmentation in Fig. 1). Geometric transformation is applied to remove the perspective between the camera position and the control (Geometric transformation in Fig. 1), reducing input image variability for more effective CNN training (i.e., the machine view remains orthogonal to the device camera). Reference marks, such as printed images on machines or specific places on the shop floor, can guide the operator in the real environment. Moreover, the layer provides indications of possible anomalies, and answers operator queries about technical documentation.

Simultaneous Localization and Mapping (SLAM) creates a 3D map of the environment to determine the user's 3D location, allowing real-time combination of virtual objects and indoor tracking. The traditional A^* algorithm can

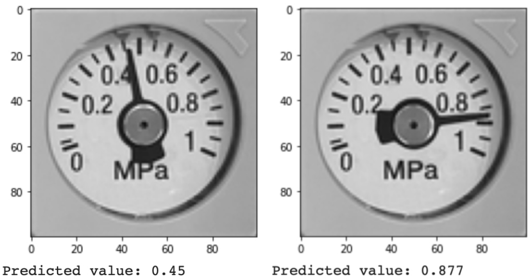


Fig. 2 Reading the pressure value with a CNN in non-sensorized machines.

guide the operator on the shop floor, while marker-based tracking and SLAM are employed to determine the operator's position.

The evaluated application utilizes AR technology to enable operators to perform complex tasks. Various tools and technologies are available for developing AR applications, including ARCore for Android [72], ARKit for iOS [73], and AR Foundation [74]. Libraries and SDKs like ARToolkit [75] or Vuforia streamline the integration of AR functionality, reducing programming complexity and supporting cross-platform development. Vuforia SDK is used and tested on an iPad device for this research.

3.3 Business layer

The Business layer plays a pivotal role in our architecture by seamlessly integrating shop floor operational systems with higher-level decision-making processes. It leverages Enterprise Resource Planning (ERP) and Supervisory Control And Data Acquisition (SCADA) systems, enabling efficient resource management and process supervision. Combining data from ERP, SCADA, and the semantic layer through CNNs enhances our system's capabilities, particularly in tasks like anomaly detection.

This layer acts as a bridge between operational and business components, utilizing ERP and SCADA data alongside DL from the semantic layer to provide comprehensive insights for informed decision-making. Our architecture promotes a holistic understanding of the industrial environment, improving operational efficiency and supporting intelligent decisions. The intermediate middleware layer serves as an interface, facilitating seamless communication between the AR Physical layer and the Business layer. Its primary function is to abstract and enable the retrieval of information from sensorized machines, enabling system scalability and flexibility.

3.4 Semantic layer

DL has enabled the development of CNN architectures for image input, utilizing convolutional layers to extract significant features and perform classifications or regressions, such as in Fig. 2, where a CNN can predict the pressure value from the image of a pressure gauge. Analyzing that image through CNNs is crucial when a machine is not sensorized (i.e., it is not connected to a SCADA system or does not have a PLC), thus being the only source of information the system can extract data from. Nevertheless, it's important to note that this visual data should not be perceived as a substitute for the structured data derived from PLCs and SCADA systems, as these systems excel in tasks related to automation, control, and historical data analysis. Instead, the integration of both structured and visual data sources emerges as a strategic approach, affording a comprehensive perspective of industrial operations that, in turn, enhances decision-making capabilities and facilitates process optimization.

On the other hand, ML allows the use of supervised models like SVM or Isolation Forest, and unsupervised models like K-means, to identify anomalies or outliers in feature vectors, such as values or states from machines or production lines. Supervised models require normal and anomalous feature vectors, while unsupervised models detect anomalies based on deviations from primary clusters in the training data.

The Semantic layer extracts meaning from the operator/device's environment and interactions, integrating three input types (See Fig. 1) to enhance system accuracy.

- CNNs classify discrete and analog controls to predict their values,
- ML algorithms detect anomalies in combinations of interpreted values,
- Transformers are used to answer NL questions from the operator in the context of a specific machine, process, or task.

Regarding the last point, the transformer-based answers significantly improve problem resolution for the operator. Combining SMEs' descriptions with existing technical documentation determines the transformer's context. The responses are transmitted to the middleware layer, where they are stored for subsequent retrieval by the AR Physical layer. This retrieval process allows for the presentation of new information within the operator's field of view. The proposed architecture considers displaying information such as the next step or anomalous information next to a control. Fig. 4 shows an example of this functionality.

3.4.1 Transformers for QA

In complex industrial tasks, AR systems provide visual cues to assist operators. By adding a Semantic layer (explained in Sec. 3.4) that utilizes CNNs to extract information and context, operators can validate actions, detect anomalies, and simplify interactions. The Semantic layer also enables automatic progression to the next step based on confirmed actions.

During MRO tasks, operators may encounter questions that require answers found in technical documentation and procedure manuals. These inquiries should be resolved quickly to avoid risks or costs resulting from delays. The operators may not have the necessary knowledge, leading them to consult either the technical documentation or a remote expert. One of the most promising neural network architectures, transformers, can address this issue.

Transformers can help interpret technical documentation and procedure manuals, even when written in NL, enabling operators to find the answers they need promptly.

Transformers outperform recurrent networks by incorporating attention mechanisms and acquiring language semantics through extensive text training. They possess high-level capabilities in NL processing, including text classification, chatbots, text summarization, and language translation. Popular transformer architectures include RoBERTa [76], DistilBERT [77], and Google's T5 [78]. In QA tasks, transformers can be specialized to provide context-based answers to NL questions. Training can be performed on large datasets like SQuAD (Stanford Question Answering Dataset) [79], or pre-trained transformers can be used for their ability to understand various question formulations in NL. Fine-tuning, as detailed in section 5, further enhances their accuracy, particularly when considering real industrial contexts.

The need for knowledge may arise during MRO tasks, with or without AR guidance. This study evaluates the availability of tools such as QA based on technical documentation and SME knowledge to analyze whether it improves task efficiency in complex environments such as shop floors.

3.4.2 Connecting transformers with AR for multimodal interaction

Transformers have proven to be an effective method for answering technical documentation questions in NL, offering valuable benefits such as streamlining the search process for specific problems and reducing cognitive load. In the evaluated application, the transformer's responses are integrated with AR technology, linking NL answers to physical locations and specific elements in the spatial mapping (see Fig. 4). This integration aims to provide precise indications of the elements of interest in the operator's physical environment alongside the NL response to a query (Fig. 3).

The proposed method for linking questions and AR responses involves preprocessing the technical documentation and utilizing existing 3D scanning technology to map the environment of interest. To achieve this, we add extra text to the technical documentation to identify elements of interest in the 3D scanned mesh unequivocally. The relevant elements in the industrial environment, such as panels, controls, indicators, and actuators, are identified and assigned unique identifiers in the 3D mesh of the shop floor area.

The connection between the identified elements and the technical documentation is established through two steps:

- Additional sentences are incorporated into the context, indicating the specific identifiers assigned to the elements of interest in the physical space. For instance, Fig. 6 illustrates four marked nozzles (0001, 0002, 0003, and 0004) of an Extruder machine, and their corresponding identifiers are listed in lines 26 to 28 of Listing 1.

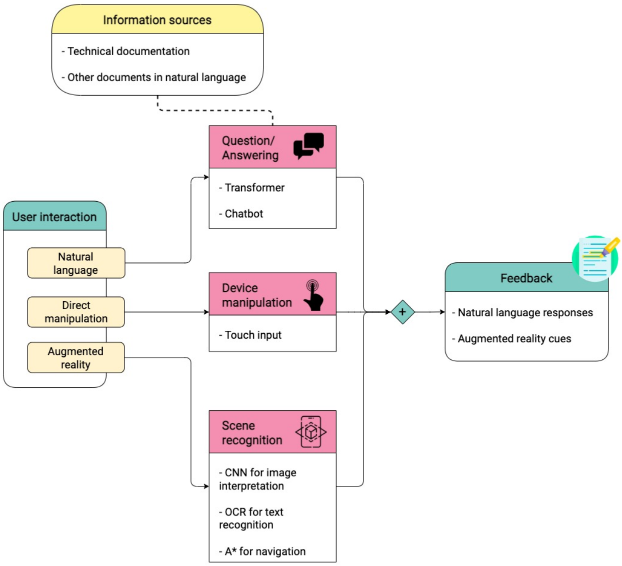


Fig. 3 Users can interact multimodally to obtain AR and NL feedback.

- In cases where the text does not clearly describe the element in question, the previous sentences are extended to provide the necessary context, enabling the transformer to obtain the required answer when asking a question.

The interaction between the operator, the transformer, and the AR system is depicted as follows (Fig. 4): The operator asks a question in NL, and the system forwards it to the transformer, which predicts and returns the answer. However, in certain cases, additional information is necessary to pinpoint the element referred to by the operator in the AR space. In such situations, a second internal question is sent to the transformer to obtain the identifier of the element of interest. Therefore, the application receives a response comprising two elements: (1) the answer provided by the transformer and (2) the ID of the element of interest associated with a specific location in the AR environment.

Explicit statements are added to identify specific controls in subsequent questions (see Listing 1, lines 26-28) to ensure that the transformer can accurately relate the response to a physical control in the real environment.

It is worth noting that the transformer used in this study belongs to the *Extractive Open QA* family. The answers are generated solely from the context and do not introduce new content. This approach is recommended when precise answers are required, based solely on the available technical documentation.

4 System implementation

The proposed architecture, its main elements, multimodal capabilities, and synergies are described, along with a real-world implementation and evaluation. This study analyzes the usage of two machines, an Extruder, and an

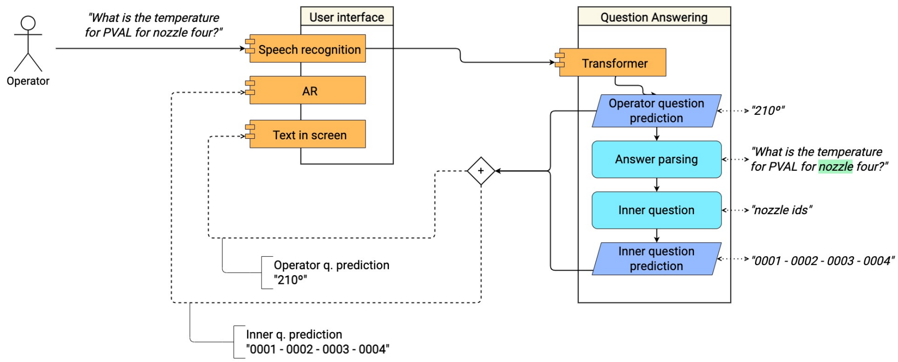


Fig. 4 Linking answers in NL to AR cues.

Injector, for creating filaments and molds from a polymer mixture. The process requires various machine parameters, such as nozzle temperatures, which vary based on the polymer type. The client-server architecture for implementing, validating, and verifying the proposed system is detailed in section 4.1 and 4.2, respectively.

4.1 Client-side application

The primary application that the operator interacts with has been developed using the Unity engine, allowing for the development of interactive, cross-platform apps for desktop and mobile devices. The app was tested on an iPad Air (4th generation). Since the system relies on AR for spatial location and operator guidance, a tool that facilitates AR application development in Unity was required. Vuforia was chosen for environment scanning since it offers prior 3D spatial mapping of the workspace and allows AR elements placement during application runtime. The initial scanning procedure is necessary as the system's primary objective is to guide the user during various tasks and processes. The result of the 3D scanned environment using an iPhone 13 with a LiDAR sensor can be seen in Fig. 5. While alternative methods like 3D laser scans exist for environment scanning, their high cost led us to choose Vuforia Area targets as the preferred technology. After the scan process, this tool allows for environment detection and tracking, so placing AR elements on the environment is possible. Additionally, a marker placed close to the control of interest aids in interpreting the CNN by eliminating the perspective of the image taken from the control. Regarding operator guidance on the shop floor, the Unity native's implementation of the A^* algorithm is utilized.

To distribute and optimize the Semantic layer functionality between the client and the server, the client performs local inferences from CNNs trained using the Open Neural Network Exchange standard (ONNX) [80]. The proposed system uses the open-source Keras library [81] to develop the architecture and train the model, which is subsequently exported to ONNX for inference on the mobile device, enabling the device to perform classification and regression tasks, which interpret images semantically.

As explained in Sec. 3.4.2, labeling the elements of interest (e.g., buttons, machines, nozzles) is required for the system to interpret the transformer's responses and identify the relevant elements in the environment. Fig. 6 displays the labels used for the temperature nozzles of the Extruder.

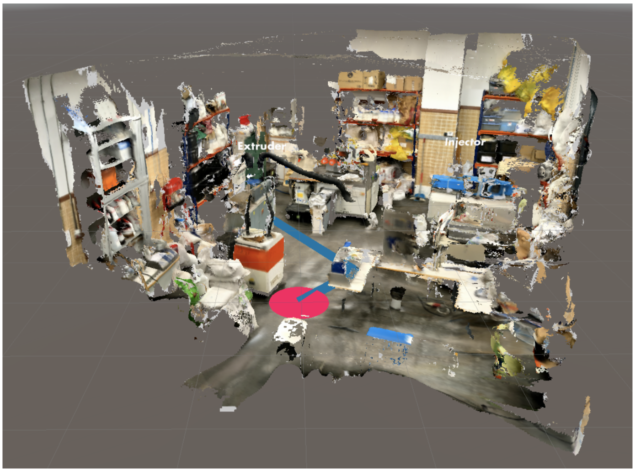


Fig. 5 Scanned mesh (Unity).

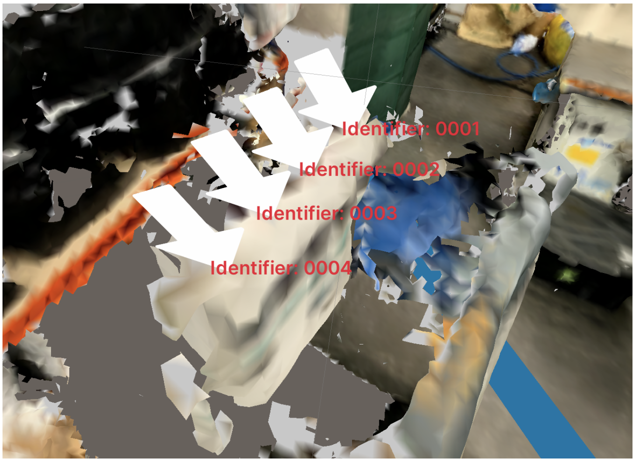


Fig. 6 Link between a physical element and transformer identifier (Unity).

4.2 Server-side application

To improve the system's overall efficiency, tasks involving complex DL models, such as anomaly detection models or transformers for QA, are distributed and solved on a server instead of the operator's device. This approach offloads the computational burden, leading to faster response times and reduced battery consumption.

For the evaluated implementation, the FastAPI framework for Python was used to run a transformer on the server. Specifically, the transformer was 80% 1x4 Block Sparse BERT-Large (uncased) Fine Tuned on SQuADv1.1 [82] due to its high prediction accuracy for the questions asked. Further insights into the decision-making process and analysis of the results are described in section 6.

4.2.1 Transformer selection

To establish a relationship between the transformer's response (Semantic layer) and the visual AR cue (AR Physical layer), controls are linked with IDs, such as 0001 or 0002, as shown in Listing 1. This linking allows triggering a second question to the transformer if a keyword like "nozzle" or "funnel" is detected in the operator's query. The result from the second question returns the control identifier and activates the cue in AR, as depicted in Fig. 4.

Four state-of-the-art QA transformers were evaluated to provide NL interaction, all available at Hugging Face.

1. Intel/bert-large-uncased-squadv1.1-sparse-80-1x4-block-pruneofa
2. bert-large-uncased-whole-word-masking-finetuned-squad
3. csarron/bert-base-uncased-squad-v1
4. csarron/mobilebert-uncased-squad-v2

Fourteen questions related to the context were performed to select the QA transformer for this study. The average scores and their results were calculated (refer to Table 1 and Table 2). The chosen model was Intel/bert-large-uncased-squadv1.1-sparse-80-1x4-block-pruneofa. It is important to note that the validity of the answers depends on the technical documentation and specific question formulation. The score value alone should not be the only factor in choosing a transformer. The possibility of using an Ensemble model to choose the highest-scoring model for a specific question should also be considered. It is worth mentioning that the study's focus is not to compare transformer models but instead to properly integrate their capabilities into an AR environment.

4.3 System operation

The presented architecture's benefits over a traditional approach have been evaluated using two applications. The first application utilizes AR, Transformers for NLP, and CNN for action validation, while the second only provides AR cues for operator guidance. The developed applications allow preparing new materials using different polymers from a list of possible processes. The AR Physical layer detects the current environment to ensure the operator is in the work area. The AR system guides the operator to the first machine using an AR route and a mini-map. Once in front of the machine, a sequence of AR cues shows the operator the actions to be performed if the process includes handling different elements. An example of the flow of use of the developed application is shown in Fig. 7. Again, the complete system, implemented in the first application, allows for asking questions in NL, thus generating responses through AR cues and textual information, and is also capable of interpreting analog values, such as pressure gauges.

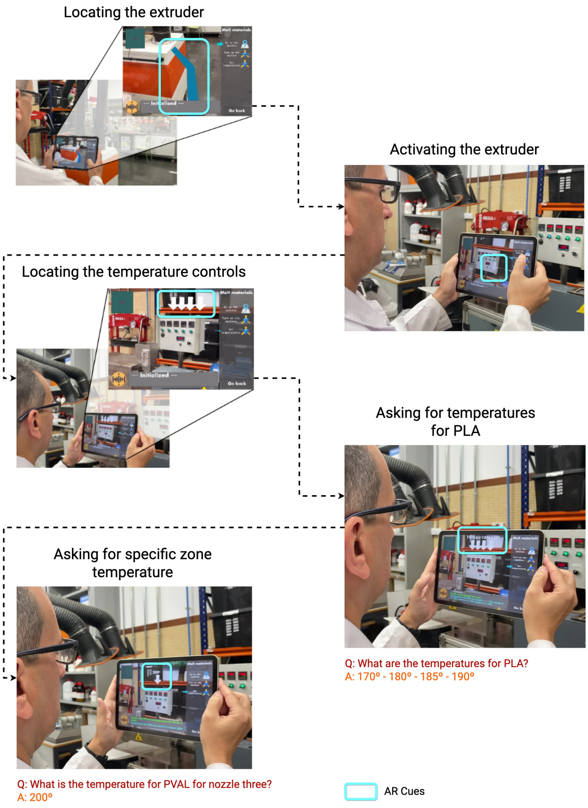


Fig. 7 Extruder workflow using the application, setting the nozzle temperatures. Visual validation, AR guiding, and voice interaction.

4.4 Implementation guidance

To generalize the benefits of the proposed architecture and facilitate its incorporation into various industrial environments, this research evaluates a specific implementation in a particular environment. The following common aspects should be considered:

1. **Technical documentation compilation** Collecting technical documentation from various sources is necessary for NL questions. The data is usually in PDF files or paper documents, but SMEs may have additional information acquired from experience. Specific tools are required to extract this data efficiently.
2. **Operatable elements** Define the machines the architecture needs to understand and operate. Automating the process of determining operable machines and controls would streamline development.
3. **AI model definition** The architectures need to be trained with several models for each control and its variations to achieve better results since elements such as gauges or buttons can significantly vary, even in the same machine. Simple CNN architectures have shown excellent results, especially with the preliminary perspective removal stage, requiring only a few pictures.
4. **Environment scanning** Enable the operator to move freely within the shop floor, necessitating 3D environment tracking. While 3D scanning and recognition tools are recommended, 2D markers from specific machines could serve as an alternative. Current AR systems generally offer these possibilities with specific authoring tools.
5. **Task definition** To highlight the AR cues linked to specific tasks, a good knowledge of what processes and tasks the operator will perform is necessary. The application should be flexible enough to manage these abstractions and highlight related elements to specific operator questions.
6. **Context and environment linkage** The architecture's advantage is the strategy used to link responses from the transformer and AR elements, as described in section 3.4.2. However, the implementation may differ for different cases.

5 System evaluation

A suitable environment was selected to evaluate the proposed system, focusing on tasks involving creating new materials using polymers mixed with various materials at different temperatures. These tasks require the use of multiple machines located in different areas. The system evaluation was conducted using the following elements:

- Three groups of people were evaluated (see Sec. 5.1), a group without AR, a group with traditional AR, and a group with the proposed system (i.e., AR and Semantic layer).
- Three work-related questions were posed to operators to measure the resolution time. Two groups of operators were tested, one with physical documentation and the other with the transformer integrated into the proposed AR system.

- A Likert questionnaire was used for qualitative analysis to evaluate two groups of operators, one with traditional AR and the other with the proposed semantic AR.

The chosen evaluation criteria follow similar studies in the literature to obtain qualitative and quantitative measures for assessing the suitability of the proposed AR system.

5.1 Participants

The proposed system was evaluated through the developed application by three groups of operators, each consisting of six operators with experience on the shop floor. Although the operators were familiar with industrial processes, they were not acquainted with the tasks evaluated, allowing for assessing potential application improvements.

Group A: The first group used an application that incorporated semantic AR through CNNs and transformers to complete the designated tasks. Additionally, this group could ask questions in NL when performing a specific task, providing text and AR cues (if possible).

Group B: The second group had access to an application that only used AR to display visual cues, which is the current standard in AR applications used in the industry. They also had access to technical documentation in a traditional electronic format (i.e., PDF files).

Group C: The third group did not have access to the application or any AR support. They were given access to the same source of technical documentation as group B and a list of tasks to be performed.

5.2 Task description

The system evaluation involved the operators performing eight tasks, comprising a complete process that required using two machines, the Extruder, and the Injector, described in section 4.

Task 1, Extruder Location: This task required locating the machine responsible for dissolving the selected material within the manufacturing plant. While groups A and B successfully used AR guidance, group C faced challenges due to limited information in the technical documentation.

Task 2, Turn on the Extruder: Upon locating the Extruder, the operators had to power it on. This step proved to be straightforward for all groups, involving a standard control and clear location visibility. Notably, group A benefited from the assistance of a CNN for automatic switch validation.

Task 3, Nozzles temperatures setup: Operators were tasked with establishing the temperatures of the four nozzles, which vary depending on the melting material. Group A utilized AR cues (e.g., arrows pointing at the nozzles) and NL interaction for easier execution (e.g., texts indicating the temperatures of the nozzles, as additional AR cues), while groups B and C faced difficulties relying solely on technical documentation.

Task 4, Extruder material insertion: In this step, operators introduced the compound into the Extruder's top funnel. AR cues provided additional assistance to groups A and B.

Task 5, Injector Location: Similar to Task 1, this step required locating the Injector machine. Groups A and B used AR cues, while group C encountered challenges without this assistance.

Task 6, Turn on the Injector: Activating the Injector involved using a standardized control, similar to task 2. Groups A and B used AR cues to find the control, while group C had minor difficulties finding the control location due to its concealed location.

Task 7, Injector material insertion: Operators needed to locate the Injector's funnel to insert the new mixture. No significant differences were observed among the three groups.

Task 8, Purge and injection: For this task, the operator had to use the Injector's Human Machine Interface (HMI) to perform purging and injection. Groups A and B received AR assistance on the HMI, while group C relied on technical documentation and encountered some challenges using the HMI menus.

To evaluate the possibility of asking in NL, operators from group A had to ask three questions to see the response from the system. These questions were:

Question 1: Asking for the temperature range of the nozzles for the Extruder for a specific material different from PLA.

Question 2: Query about a machine's location on the shop floor.

Question 3: Inquiry regarding the range of values on the pressure during the injection process.

In addition, a six-question questionnaire was administered to operators from groups A and B using a 5-point Likert scale (see Table 6). The questionnaire aims to compare and evaluate the use of standard AR applications versus adding a semantic layer on top of it with the capability of asking questions in NL. The results of the questionnaire can be found in section 6.

6 Analysis and results

Table B1 presents the task completion times for the three groups. Table B2 compiles the time required to answer three specific questions for groups A and C. As a reminder, operators from group A could use NL to ask questions to the selected transformer. Group B and group C operators had to search for answers in technical documentation.

A two-factor ANOVA test was conducted, with one factor using repeated measures, to assess whether the group impacts the task execution time. Table 3 displays the means and standard deviations of task execution times for each group (refer to Fig. B1 in appendix B for a comparative view). The test results indicate that there is a statistically significant difference between the groups, regardless of the task performed ($F(7, 105) = 288.39$; $p < 0.001$; $\eta^2 = 0.951$).

Table 1 Transformer predictions to questions (Intel/bert-large-uncased-squadv1.1-sparse-80-1x4-block-pruneofa).

Index	Question	Selected transformer prediction
0	*What is PLA?*	polylactide
1	*What are all the temperatures for melting PLA?*	170º - 180º - 185º - 190º
2	*What are the temperatures for PLA?*	170º - 180º - 185º - 190º
3	*What is the range of temperatures for melting PLA?*	170º - 180º - 185º - 190º
4	*What is the temperature for the first nozzle for PLA?*	170º
5	*What are the temperatures for PVAL?*	190º - 195º - 200º - 210º
6	*What is the temperature for PVAL for nozzle four?*	210º
7	*How do I start using the extruder?*	setting the nozzle controls to the desired temperature
8	*What is the next step after setting the temperatures?*	pour the materials into the funnel
9	*Where is the extruder's funnel?*	top of the machine
10	*Nozzle id's*	0001 - 0002 - 0003 - 0004
11	*nozzle id's*	0001 - 0002 - 0003 - 0004
12	*Activation button id*	0005
13	*funnel ID*	0006

Table 2 Scores for the 4 QA transformers tested.

Model index	Mean score
1	±68.46
2	±30
3	±53.38
4	±28

However, the interaction between group and task was also significant, suggesting that the execution time of a task depends on the group performing it ($F(14, 105)$ =8.30; $p < 0.001$; η^2 =0.525).

Tasks 2, 4, 6, and 7 show no statistically significant differences in execution times between the operator groups. Conversely, in tasks 1, 5, and 8, the execution times of operators in group C, who only used the documentation, were significantly longer than those of operators in groups A and B, who used the applications; no significant differences were found between groups A and B. For task 3, the execution time of operators in group C was significantly longer than that of operators in group A; no significant differences were observed between groups A and B or B and C.

Table 4 presents the p-values of the pairwise comparisons with Bonferroni correction to adjust the significance level of each pairwise comparison between the groups and tasks.

The response times to the questions posed were analyzed, and the results are presented in table 5 (see also Fig. B2 in appendix B). The analysis indicates a statistically significant difference between groups A and C concerning the three proposed questions. The response times of operators who worked with

Table 3 Descriptive execution times of tasks and statistical contrasts.

	Task							
	1 *Mean (Sd)*	**2** *Mean (Sd)*	**3** *Mean (Sd)*	**4** *Mean (Sd)*	**5** *Mean (Sd)*	**6** *Mean (Sd)*	**7** *Mean (Sd)*	**8** *Mean (Sd)*
Group								
Group A	18,67 (3,56)	5,50 (2,26)	50,67 (15,95)	23,00 (6,81)	19,67 (4,93)	12,00 (3,29)	6,83 (3,82)	89,83 (8,91)
Group B	18,50 (2,26)	4,67 (1,37)	64,17 (11,58)	15,83 (4,71)	18,83 (6,31)	10,50 (2,59)	6,83 (2,64)	85,50 (11,52)
Group C	37,83 (8,04)	5,67 (3,01)	77,83 (14,58)	16,83 (5,49)	53,17 (10,07)	14,83 (7,91)	7,67 (2,25)	126,00 (23,38)

Table 4 Two to two comparisons P-values (Bonferroni correction).

	Task							
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Group A Vs. Group B	1	1	0,358	0,141	1	1	1	1
Group A Vs. Group C	$<$ **0,001**	1	**0,014**	0,247	$<$ **0,001**	1	1	**0,004**
Group B Vs. Group C	$<$ **0,001**	1	0,345	1	$<$ **0,001**	0,501	1	**0,002**

the application were significantly lower than those of operators who performed tasks with documentation for all three questions ($p < 0,001$, $p < 0,001$, and $p < 0,001$, respectively). However, the interaction between the group and the task was not significant.

Table 5 Descriptive answer times and statistical contrasts.

	Question			**Tests of within-subjects effects**	
	1	**2**	**3**	**Group**	**Group*Question**
	Mean (Sd)	*Mean (Sd)*	*Mean (Sd)*	**F(g.l.);** $p - valor(\eta^2)$	**F(g.l.);** $p - valor(\eta^2)$
Group				$F(2;\ 20) = 3,93$; $p = 0,036(0,274)$	$F(2;\ 20) = 0,57$; $p = 0,572(0,054)$
Group A	19,67 (8,89)	10,67 (5,79)	15,17 (5,08)		
Group C	87,33 (12,31)	70,83 (19,45)	87,83 (21,74)		

Table 6 describes and compares the scores given to each statement based on the group of workers. The results indicate that operators from group A showed more significant agreement than those from group B for the questions *"It helped me solve technical questions seamlessly"* and *"I can access additional information effortlessly"*. These answers suggest that workers in group A found the system more helpful and easier to use than group B workers.

The results show that the proposed semantic layer reduces the time required for tasks of a more cognitive nature and provides greater comfort and security in the qualitative assessment of results. Consequently, group A shows better results and higher levels of satisfaction. The evaluation of the proposed solution confirms the significant advantages that can be obtained by complementing current AR systems in industrial environments with the proposed semantic and multimodal layers. These findings align with similar studies such as Rasmussen

Table 6 Descriptive and comparative scores to the questions raised about the performance of the tasks.

Question	Group, *median (IR)*		Mann-Whitney U test	
	Group A	Group B	U	*p*-value
I think AR is helpful	5 (5 - 5)	5 (4 - 5)	15	0,523
I feel I can finish the tasks faster	4,5 (4 - 5)	4 (4 - 5)	13,5	0,423
I felt more confident and safe using the application	5 (5 - 5)	5 (5 - 5)	15	0,317
It helped me solve technical questions seamlessly	5 (4 - 5)	2,5 (2 - 3)	1	**0,005**
I can focus better on the task	4,5 (4 - 5)	4 (4 - 5)	15	0,575
I can access additional information effortlessly	5 (5 - 5)	2 (2 - 2)	0	**0,002**

IR: interquartile range

et al. on workspace awareness using a multi-camera solution [28], Eversberg et al. on cognitively assisted AR through digital twins [29], Wang et al. on spatial cognition [30], and Zhang et al. on the benefits of multimodal interaction [31]. It is important to acknowledge that further investigation and analysis are necessary to address the open questions surrounding the semantic layer and to refine the system's capabilities in future research.

7 Discussion

The present study investigated the effectiveness of a proposed semantic layer in augmenting industrial workers' AR tasks. The analysis of the study's data revealed several significant findings. Firstly, the two-factor ANOVA test demonstrated a statistically significant difference in task execution times between the three groups (A, B, and C). Operators in group C, who relied solely on technical documentation, exhibited significantly longer task execution times than those in groups A and B, who used the proposed application-based semantic layer. Moreover, task completion times varied significantly across different tasks, suggesting that the semantic layer's impact depends on the task's nature. The response times to specific questions showed that operators in group A, using the semantic layer, had significantly lower response times than those in group C, who used only technical documentation. This finding indicates that the semantic layer can streamline obtaining information and answering queries, potentially improving workers' efficiency. The likert-scale evaluation indicated that operators in group A expressed higher levels of agreement with statements related to the proposed system's helpfulness, confidence, and ease of use. These results suggest that the semantic layer contributes positively to the overall user experience and satisfaction.

These outcomes imply that the semantic layer has the potential to enhance workers' cognitive capabilities and enable them to perform complex tasks more effectively. The advantages observed in tasks requiring more cognitive processing further emphasize the importance of addressing human-computer interaction challenges in AR systems. The seamless integration of contextual information and multimodal interaction in the proposed solution potentially alleviates cognitive load, improving task performance.

We acknowledge several limitations in this study:

1. The sample size for each group was relatively small, which may limit the generalizability of the results. Future studies with more extensive and diverse samples are necessary to validate and extend our findings.
2. While various tasks were considered, the chosen tasks may only partially represent some possible scenarios in industrial settings. Further investigation across a broader range of tasks would enhance the study's comprehensiveness.
3. The study was conducted in a controlled environment, and real-world industrial conditions may introduce additional complexities not fully accounted for in this research.

8 Conclusions

In the Industry 4.0 era, the operator plays a crucial role in incorporating the mechanisms that benefit from this revolution. This is becoming a fundamental element of the new Industry 5.0 definition that has, as one essential component, the human-centric approach and the value-added work of future operators. AR is an essential technology to enable this evolution toward an automated, efficient, and user-centric environment. As tasks performed by operators on the shop floor become increasingly complex, there is a growing need for qualified labor. However, remote assistance or access to an SME may only sometimes be available. In critical situations, the operator's cognitive load can increase stress levels, affect problem resolution, and compromise safety.

This study proposes a system consisting of four layers to assist operators in the Industry 4.0 era. The AR Physical layer is the primary interface between the operator and the shop floor. Multimodal interaction using NL, AR, and direct manipulation helps operators learn the system more quickly and interact with it more naturally. By incorporating ML and DL models, the system further assists operators in understanding the environment (e.g., using CNNs to get states and values of non-sensorized machines), enabling them to interact in NL and receive verbal and visual AR feedback.

The proposed system was implemented and evaluated with three groups of operators, and its real-life use cases have demonstrated significant benefits:

- **Enhanced productivity and reduced errors**: The first group used the proposed system with a semantic layer, and they experienced a substantial increase in task completion efficiency and a notable reduction in errors. This improvement showcases how AR technology, when integrated with semantic layers, can significantly enhance operator performance in complex industrial tasks.
- **Comparison with limited application**: The second group, using a limited application version without the semantic layer, struggled to match the efficiency and accuracy of the first group. This comparison underscores the

importance of incorporating semantic layers in AR solutions to improve cognitive tools and environmental awareness.

- **Importance of semantic layers**: In contrast, the third group, which relied solely on technical documentation, faced longer task completion times and a higher error rate. These findings further highlight the importance of integrating AR solutions with semantic layers, as they outperformed traditional documentation in assisting operators.

These findings highlight the importance of including semantic layers in AR solutions to enhance cognitive tools and environmental awareness, and are aligned with recent studies, such as the one by Eswaran et al. on augmented opportunities in the industry [83].

The proposed system effectively reduces task completion time and minimizes errors, ultimately contributing to improved operational efficiency and operator performance. As the Industry 4.0 revolution progresses, it is crucial to continue exploring and refining such AR systems to empower operators, mitigate cognitive load, and foster a safe and productive industrial setting. As Industry 4.0 and 5.0 advance, this research paves the way for future investigations in AR systems. There is a need for further refinement and exploration of novel interactions to enhance AR system efficiency for operator assistance. Integrating AI-driven decision support can empower operators to handle complex tasks effectively. Additionally, extending semantic layers within AR solutions shows promise in augmenting operators' cognitive tools and environmental awareness. These future research directions can significantly advance AR systems, creating a more user-centric, efficient, and safe industrial environment. These advancements hold the potential to revolutionize various industries, from manufacturing and logistics to healthcare and maintenance, providing tangible benefits to operators and organizations alike.

Declarations

Conflict of interest

The authors declare that they have no conflicts of interest to report regarding the present study.

Funding

The authors did not receive support from any organization for the submitted work.

Data Availability

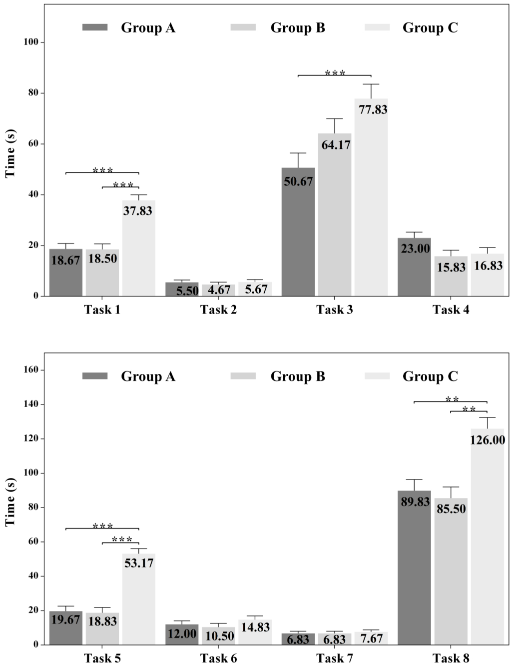
Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

Appendix A Transformer context

```python
context = """
Extruder information:
- PLA is Polylactide.
- For melting PLA, 170º - 180º - 185º - 190º.
- PLA nozzle one/first: 170º.
- PLA nozzle two/second: 180º.
- PLA nozzle three/third: 185º.
- PLA nozzle four/fourth: 190º.

- PVAL is Polivinylalcohol.
- For melting PVAL, 190º - 195º - 200º - 210º.
- PVAL nozzle one/first: 190º.
- PVAL nozzle two/second: 195º.
- PVAL nozzle three/third: 200º.
- PVAL nozzle four/fourth: 210º.

Extruder tasks:
1º Turn on the machine turning the activation button to the left.
2º Heat up the machine by using setting the nozzle controls to the desired
    temperature.
3º When the machine is heated, pour the materials into the funnel, found at
    the top of the machine.
4º The material will appear through the fourth nozzle.

-------------------------------------

Extruder IDs:
- Nozzle ID ( 0001 - 0002 - 0003 - 0004 ).
- Activation button ID ( 0005 ).
- Funnel ID ( 0006 ).
"""
```

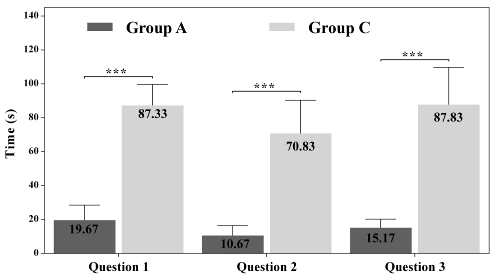
Listing 1 Transformer context based on technical documentation (Python language).

Appendix B System evaluation



$**\,p < 0.01, ***p < 0.001$

Fig. B1 Comparative of tasks, group A (Semantic AR) Vs. group B (Only AR) Vs. group C (Traditional).



$***p < 0.001$

Fig. B2 Comparative of answers (Group A Vs. Group C).

Table B1 Task time results (Group A Vs. Group B Vs. Group C).

	Group A						Group B						Group C						Description
	Op. 1	**Op. 2**	**Op. 3**	**Op. 4**	**Op. 5**	**Op. 6**	**Op. 1**	**Op. 2**	**Op. 3**	**Op. 4**	**Op. 5**	**Op. 6**	**Op. 1**	**Op. 2**	**Op. 3**	**Op. 4**	**Op. 5**	**Op. 6**	
Task 1	16s	21s	18s	24s	19s	14s	20s	17s	21s	15s	18s	20s	34s	45s	28s	32s	49s	39s	*Extruder locating*
Task 2	4s	4s	3s	9s	7s	6s	4s	7s	3s	5s	4s	5s	9s	3s	8s	2s	4s	8s	*Extruder activation*
Task 3	52s *	42s	81s *	38s	51s	40s	72s	64s	82s	57s	61s	49s	87s	98s	67s	57s	76s	82s	*Temperature setup*
Task 4	23s	21s	22s	28s	12s	32s	15s	19s	23s	10s	16s	12s	24s	14s	9s	21s	14s	19s	*Material insertion*
Task 5	20s	17s	17s	22s	28s	14s	24s	13s	17s	16s	29s	14s	54s	64s	43s	43s	66s	49s	*Injector locating*
Task 6	14s	8s	17s	12s	9s	12s	10s	7s	12s	13s	8s	13s	19s	7s	13s	17s	6s	27s	*Injector activation*
Task 7	6s	4s	14s	4s	5s	8s	5s	6s	5s	12s	7s	6s	7s	8s	6s	6s	7s	12s	*Material insertion*
Task 8	78s	85s	97s	87s	89s	103s	91s	101s	83s	92s	69s	77s	110s	143s	89s	131s	129s	154s	*Purge and injection*

Table B2 Question time results (Group A Vs. Group C).

	Group A						Group C						Question
	Op. 1	**Op. 2**	**Op. 3**	**Op. 4**	**Op. 5**	**Op. 6**	**Op. 1**	**Op. 2**	**Op. 3**	**Op. 4**	**Op. 5**	**Op. 6**	
Question 1	10s	35s *	14s	21s	23s	15s	98s	76s	89s	103s	87s	71s	*Temperature range*
Question 2	6s	9s	16s *	10s	4s	19s *	68s	45s	70s	59s	102s	81s	*Machine location*
Question 3	13s	17s	8s	13s	17s	23s *	56s	113s	83s	91s	110s	74s	*Pressure values*

* Ask again to transformer (bad answer to a first question)

References

[1] H. Kagermann, J. Helbig, A. Hellinger, W. Wahlster, *Recommendations for implementing the strategic initiative INDUSTRIE 4.0: Securing the future of German manufacturing industry; final report of the Industrie 4.0 Working Group* (Forschungsunion, 2013)

[2] L.D. Xu, E.L. Xu, L. Li, Industry 4.0: State of the art and future trends. International Journal of Production Research **56**(8), 2941–2962 (2018). https://doi.org/10.1080/00207543.2018.1444806

[3] C.H. Chu, L. Wang, S. Liu, Y. Zhang, M. Menozzi, Augmented reality in smart manufacturing: Enabling collaboration between humans and artificial intelligence. Journal of Manufacturing Systems **61**, 658–659 (2021). https://doi.org/10.1016/j.jmsy.2021.05.006

[4] B.V. Guerreiro, R.G. Lins, J. Sun, R. Schmitt, in *Advances in Manufacturing*, vol. 0 (Springer Heidelberg, 2018), pp. 161–170. https://doi.org/10.1007/978-3-319-68619-6{_}16

[5] J.M. Runji, Y.J. Lee, C.H. Chu, Systematic Literature Review on Augmented Reality-Based Maintenance Applications in Manufacturing Centered on Operator Needs. International Journal of Precision Engineering and Manufacturing-Green Technology (2022). https://doi.org/10.1007/s40684-022-00444-w

[6] J.P. Womack, D.T. Jones, D. Roos, A máquina que mudou o mundo. Campus: Rio de Janeiro (1992)

[7] X. Wang, S.K. Ong, A.Y. Nee, A comprehensive survey of augmented reality assembly research. Advances in Manufacturing **4**(1), 1–22 (2016). https://doi.org/10.1007/s40436-015-0131-4

[8] T. Zonta, C.A. da Costa, R. da Rosa Righi, M.J. de Lima, E.S. da Trindade, G.P. Li, Predictive maintenance in the Industry 4.0: A systematic literature review. Computers and Industrial Engineering **150** (2020). https://doi.org/10.1016/j.cie.2020.106889

[9] R. Palmarini, J.A. Erkoyuncu, R. Roy, H. Torabmostaedi, A systematic review of augmented reality applications in maintenance. Robotics and Computer-Integrated Manufacturing **49**, 215–228 (2018). https://doi.org/10.1016/j.rcim.2017.06.002

[10] S. Jaschke, in *Proceedings of 2014 International Conference on Interactive Collaborative Learning, ICL 2014* (Institute of Electrical and Electronics Engineers Inc., 2014), pp. 605–608. https://doi.org/10.1109/ICL.2014.7017840

[11] A.M. Huenerfauth, Mobile technology applications for manufacturing, reduction of muda (waste) and the effect on manufacturing economy and efficiency. International Journal of Interactive Mobile Technologies **8**(4), 20–23 (2014). https://doi.org/10.3991/ijim.v8i4.3797

[12] C. Kollatsch, P. Klimant, Efficient integration process of production data into Augmented Reality based maintenance of machine tools. Production Engineering **15**(3-4), 311–319 (2021). https://doi.org/10.1007/s11740-021-01026-6

[13] M. Gattullo, G.W. Scurati, M. Fiorentino, A.E. Uva, F. Ferrise, M. Bordegoni, Towards augmented reality manuals for industry 4.0: A methodology. Robotics and Computer-Integrated Manufacturing **56**(March 2018), 276–286 (2019). https://doi.org/10.1016/j.rcim.2018.10.001

[14] A. Gilchrist, in *Industry 4.0* (Springer, 2016), chap. 13, pp. 195–215. https://doi.org/10.1007/978-1-4842-2047-4

[15] R.W. Backs, K.A. Seljos, Metabolic and cardiorespiratory measures of mental effort: the effects of level of difficulty in a working memory task. International Journal of Psychophysiology **16**, 57–68 (1994)

[16] J. Sweller, Cognitive Load During Problem Solving: Effects on Learning. Cognitive Science **12**(2), 257–285 (1988). https://doi.org/10.1207/s15516709cog1202{_}4

[17] C. Sandi, Stress and cognition. Wiley Interdisciplinary Reviews: Cognitive Science **4**(3), 245–261 (2013). https://doi.org/10.1002/wcs.1222

[18] D. Romero, J. Stahre, M. Taisch. The Operator 4.0: Towards socially sustainable factories of the future (2020). https://doi.org/10.1016/j.cie.2019.106128

[19] D. Romero, J. Stahre, T. Wuest, O. Noran, P. Bernus, A. Fast-Berglund, D. Gorecky, in *Proceedings of the international conference on computers and industrial engineering (CIE46)* (Tianjin, China, 2016), pp. 29–31. URL https://www.researchgate.net/publication/309609488

[20] M. Peruzzini, F. Grandi, M. Pellicciari, Exploring the potential of Operator 4.0 interface and monitoring. Computers and Industrial Engineering **139** (2020). https://doi.org/10.1016/j.cie.2018.12.047

[21] L.P. Zambiasi, R.J. Rabelo, S.P. Zambiasi, R. Lizot, in *IFIP International Conference on Advances in Production Management Systems* (2022), pp. 494–502. URL https://link.springer.com/bookseries/6102

[22] D. Romero, J. Stahre, Towards the Resilient Operator 5.0: The Future of Work in Smart Resilient Manufacturing Systems. Procedia CIRP **104**, 1089–1094 (2021). https://doi.org/10.1016/j.procir.2021.11.183

[23] R.J. Rabelo, D. Romero, S. Popov Zambiasi, in *IFIP International Conference on Advances in Production Management Systems (APMS)* (2018), pp. 456–464. https://doi.org/10.1007/978-3-319-99707-0{_}57. URL https://inria.hal.science/hal-02177873

[24] G.H. Ball, D.J. Hall, ISODATA, a novel method of data analysis and pattern classification. Tech. rep., Stanford research inst Menlo Park CA (1965)

[25] C. Cortes, V. Vapnik, L. Saitta, Support-Vector Networks Editor. Machine Learning **20**, 273–297 (1995)

[26] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, Ł. Kaiser, I. Polosukhin, in *Advances in neural information processing systems*, vol. 30 (2017), pp. 6000–6010. https://doi.org/10.48550/arXiv.1706.03762

[27] R. Bommasani, D.A. Hudson, E. Adeli, R. Altman, S. Arora, S. von Arx, On the Opportunities and Risks of Foundation Models. arXiv preprint arXiv:2108.07258 (2021). https://doi.org/10.48550/arXiv.2108.07258. URL http://arxiv.org/abs/2108.07258

[28] T. Rasmussen, T. Feuchtner, W. Huang, K. Grønbæk, Supporting workspace awareness in remote assistance through a flexible multi-camera system and Augmented Reality awareness cues. Journal of Visual Communication and Image Representation p. 103655 (2022). https://doi.org/10.1016/J.JVCIR.2022.103655

[29] L. Eversberg, P. Ebrahimi, M. Pape, J. Lambrecht, A cognitive assistance system with augmented reality for manual repair tasks with high variability based on the digital twin. Manufacturing Letters **34**, 49–52 (2022). https://doi.org/10.1016/J.MFGLET.2022.09.003

[30] Z. Wang, X. Bai, S. Zhang, M. Billinghurst, W. He, Y. Wang, D. Han, G. Chen, J. Li, The role of user-centered AR instruction in improving novice spatial cognition in a high-precision procedural task. Advanced Engineering Informatics **47** (2021). https://doi.org/10.1016/j.aei.2021.101250

[31] J. Zhang, S. Wang, W. He, J. Li, Z. Cao, B. Wei, Projected augmented reality assembly assistance system supporting multi-modal interaction. The International Journal of Advanced Manufacturing Technology 2022 123:3 **123**(3), 1353–1367 (2022). https://doi.org/10.1007/

S00170-022-10113-6

[32] P.C.y. Sheu, *Semantic computing* (Wiley Online Library, 2010)

[33] C.K. Sahu, C. Young, R. Rai, Artificial intelligence (AI) in augmented reality (AR)-assisted manufacturing applications: a review. International Journal of Production Research **59**(16), 4903–4959 (2021). https://doi.org/10.1080/00207543.2020.1859636

[34] M. Bertolini, D. Mezzogori, M. Neroni, F. Zammori, Machine Learning for industrial applications: A comprehensive literature review. Expert Systems with Applications **175**, 114,820 (2021). https://doi.org/10.1016/J.ESWA.2021.114820

[35] H. Esen, F. Ozgen, M. Esen, A. Sengur, Artificial neural network and wavelet neural network approaches for modelling of a solar air heater. Expert Systems with Applications **36**(8), 11,240–11,248 (2009). https://doi.org/10.1016/J.ESWA.2009.02.073

[36] M.A. Zamora-Hernández, J.A. Castro-Vargas, J. Azorin-Lopez, J. Garcia-Rodriguez, Deep learning-based visual control assistant for assembly in Industry 4.0. Computers in Industry **131** (2021). https://doi.org/10.1016/j.compind.2021.103485

[37] A.Y. Javaid, Q. Niyaz, W. Sun, M. Alam, A deep learning approach for network intrusion detection system. EAI International Conference on Bio-inspired Information and Communications Technologies (BICT) (2015). https://doi.org/10.4108/eai.3-12-2015.2262516

[38] R. Chalapathy, S. Chawla, Deep Learning for Anomaly Detection: A Survey. arXiv preprint arXiv:1901.03407 pp. 1–50 (2019). https://doi.org/10.48550/arXiv.1901.03407

[39] R.T. Ionescu, F.S. Khan, M.I. Georgescu, L. Shao, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2019), pp. 7842–7851

[40] W. Lu, Y. Cheng, C. Xiao, S. Chang, S. Huang, B. Liang, T. Huang, Unsupervised Sequential Outlier Detection with Deep Architectures. IEEE Transactions on Image Processing **26**(9), 4321–4330 (2017). https://doi.org/10.1109/TIP.2017.2713048

[41] G. Pang, C. Yan, C. Shen, A. Van Den Hengel, X. Bai, in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (2020), pp. 12,173–12,182

[42] H. Zenati, C.S. Foo, B. Lecouat, G. Manek, V.R. Chandrasekhar, Efficient GAN-Based Anomaly Detection. arXiv preprint arXiv:1802.06222 (2018). https://doi.org/10.48550/arXiv.1802.06222

[43] V. Škvára, T. Pevný, V. Šmídl, Are generative deep models for novelty detection truly better? arXiv preprint arXiv:1807.05027 (2018). https://doi.org/10.48550/arXiv.1807.05027

[44] X. Song, P. Sun, S. Song, V. Stojanovic, Event-driven NN adaptive fixed-time control for nonlinear systems with guaranteed performance. Journal of the Franklin Institute **359**(9), 4138–4159 (2022). https://doi.org/10.1016/J.JFRANKLIN.2022.04.003

[45] R.B. Gopaluni, A. Tulsyan, B. Chachuat, B. Huang, J.M. Lee, F. Amjad, S.K. Damarla, J. Woo Kim, N.P. Lawrence, Modern Machine Learning Tools for Monitoring and Control of Industrial Processes: A Survey. IFAC-PapersOnLine **53**(2), 218–229 (2020). https://doi.org/10.1016/J.IFACOL.2020.12.126

[46] R. Radkowski, J. Herrema, J. Oliver, Augmented Reality-Based Manual Assembly Support With Visual Features for Different Degrees of Difficulty. International Journal of Human-Computer Interaction **31**(5), 337–349 (2015). https://doi.org/10.1080/10447318.2014.994194

[47] G.W. Scurati, M. Gattullo, M. Fiorentino, F. Ferrise, M. Bordegoni, A.E. Uva, Converting maintenance actions into standard symbols for Augmented Reality applications in Industry 4.0. Computers in Industry **98**, 68–79 (2018). https://doi.org/10.1016/j.compind.2018.02.001

[48] Y.P. Luh, J.B. Wang, J.W. Chang, S.Y. Chang, C.H. Chu, Augmented reality-based design customization of footwear for children. Journal of Intelligent Manufacturing **24**(5), 905–917 (2013). https://doi.org/10.1007/s10845-012-0642-9

[49] Y. Shen, S.K. Ong, A.Y. Nee, Augmented reality for collaborative product design and development. Design Studies **31**(2), 118–145 (2010). https://doi.org/10.1016/j.destud.2009.11.001

[50] S.K. Ong, Y. Shen, A mixed reality environment for collaborative product design and development. CIRP Annals - Manufacturing Technology **58**(1), 139–142 (2009). https://doi.org/10.1016/j.cirp.2009.03.020

[51] S.K. Ong, Z.B. Wang, Augmented assembly technologies based on 3D bare-hand interaction. CIRP Annals - Manufacturing Technology **60**(1), 1–4 (2011). https://doi.org/10.1016/j.cirp.2011.03.001

[52] M.L. Yuan, S.K. Ong, A.Y. Nee, Augmented reality for assembly guidance using a virtual interactive tool. International Journal of Production Research **46**(7), 1745–1767 (2008). https://doi.org/10.1080/00207540600972935

[53] D. Mourtzis, V. Siatras, J. Angelopoulos, Real-time remote maintenance support based on augmented reality (AR). Applied Sciences (Switzerland) **10**(5) (2020). https://doi.org/10.3390/app10051855

[54] D. Tatić, B. Tešić, The application of augmented reality technologies for the improvement of occupational safety in an industrial environment. Computers in Industry **85**, 1–10 (2017). https://doi.org/10.1016/j.compind.2016.11.004

[55] D.B. Espíndola, L. Fumagalli, M. Garetti, C.E. Pereira, S.S. Botelho, R. Ventura Henriques, A model-based approach for data integration to improve maintenance management by mixed reality. Computers in Industry **64**(4), 376–391 (2013). https://doi.org/10.1016/j.compind.2013.01.002

[56] L.E. Garza, G. Pantoja, P. Ramírez, H. Ramírez, N. Rodríguez, E. González, R. Quintal, J.A. Pérez, Augmented reality application for the maintenance of a flapper valve of a fuller-kynion type m pump. Procedia Computer Science **25**, 154–160 (2013). https://doi.org/10.1016/j.procs.2013.11.019

[57] S. Benbelkacem, M. Belhocine, A. Bellarbi, N. Zenati-Henda, M. Tadjine, Augmented reality for photovoltaic pumping systems maintenance tasks. Renewable Energy **55**, 428–437 (2013). https://doi.org/10.1016/j.renene.2012.12.043

[58] Z. Ziaei, A. Hahto, J. Mattila, M. Siuko, L. Semeraro, Real-time markerless Augmented Reality for Remote Handling system in bad viewing conditions. Fusion Engineering and Design **86**(9-11), 2033–2038 (2011). https://doi.org/10.1016/j.fusengdes.2010.12.082

[59] N. Zenati, N. Zerhouni, K. Achour, in *Proceedings of the IEEE International Conference on Industrial Technology*, vol. 2 (2004), pp. 848–852. https://doi.org/10.1109/icit.2004.1490185

[60] I. Barakonyi, T. Psik, D. Schmalstieg, in *ISMAR 2004: Proceedings of the Third IEEE and ACM International Symposium on Mixed and Augmented Reality* (2004), pp. 141–150. https://doi.org/10.1109/ISMAR.2004.11

[61] A. Monroy Reyes, O.O. Vergara Villegas, E. Miranda Bojórquez, V.G. Cruz Sánchez, M. Nandayapa, A mobile augmented reality system to

support machinery operations in scholar environments. Computer Applications in Engineering Education **24**(6), 967–981 (2016). https://doi.org/10.1002/cae.21772

[62] S. Webel, U. Bockholt, T. Engelke, N. Gavish, M. Olbrich, C. Preusche, An augmented reality training platform for assembly and maintenance skills. Robotics and Autonomous Systems **61**(4), 398–403 (2013). https://doi.org/10.1016/j.robot.2012.09.013

[63] F. De Crescenzio, M. Fantini, F. Persiani, L. Di Stefano, P. Azzari, S. Salti, Augmented reality for aircraft maintenance training and operations support. IEEE Computer Graphics and Applications **31**(1), 96–101 (2011). https://doi.org/10.1109/MCG.2011.4

[64] E. Bottani, G. Vignali, Augmented reality technology in the manufacturing industry: A review of the last decade. IISE Transactions **51**(3), 284–310 (2019). https://doi.org/10.1080/24725854.2018.1493244

[65] M. Baldauf, R. Bösch, C. Frei, F. Hautle, M. Jenny, in *MobileHCI 2018 - Beyond Mobile: The Next 20 Years - 20th International Conference on Human-Computer Interaction with Mobile Devices and Services, Conference Proceedings Adjunct* (Association for Computing Machinery, Inc, 2018), pp. 119–126. https://doi.org/10.1145/3236112.3236128

[66] E. Coli, N. Melluso, G. Fantoni, D. Mazzei, Towards Automatic building of Human-Machine Conversational System to support Maintenance Processes. arXiv preprint arXiv:2005.06517 (2020). https://doi.org/10.48550/arXiv.2005.06517

[67] M. Casillo, F. Colace, L. Fabbri, M. Lombardi, A. Romano, D. Santaniello, in *Proceedings of 2020 IEEE International Conference on Teaching, Assessment, and Learning for Engineering, TALE 2020* (Institute of Electrical and Electronics Engineers Inc., 2020), pp. 371–376. https://doi.org/10.1109/TALE48869.2020.9368339

[68] K. Mleczko, Chatbot as a Tool for Knowledge Sharing in the Maintenance and Repair Processes. Multidisciplinary Aspects of Production Engineering **4**(1), 499–508 (2021). https://doi.org/10.2478/mape-2021-0045

[69] TeamViewer. TeamViewer Assist AR (2021). URL www.teamviewer.com/es/realidad-aumentada

[70] PTC. Vuforia Chalk (2017). URL www.ptc.com/es/products/vuforia/vuforia-chalk

[71] B. Shneiderman, C. Plaisant, M.S. Cohen, S. Jacobs, N. Elmqvist, N. Diakopoulos, *Designing the user interface: strategies for effective*

human-computer interaction (Pearson, 2016)

[72] Google. ARCore (2018). URL developers.google.com/ar

[73] Apple. ARKit (2017). URL developer.apple.com/augmented-reality

[74] Unity. AR Foundation (2018). URL unity.com/es/unity/features/arfoundation

[75] H. Kato, M. Billinghurst, in *Proceedings - 2nd IEEE and ACM International Workshop on Augmented Reality, IWAR 1999* (Institute of Electrical and Electronics Engineers Inc., 1999), pp. 85–94. https://doi.org/10.1109/IWAR.1999.803809

[76] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, V. Stoyanov, RoBERTa: A Robustly Optimized BERT Pretraining Approach. arXiv preprint arXiv:1907.11692 (2019). https://doi.org/10.48550/arXiv.1907.11692

[77] V. Sanh, L. Debut, J. Chaumond, T. Wolf, DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. arXiv preprint arXiv:1910.01108 (2019). https://doi.org/10.48550/arXiv.1910.01108

[78] C. Raffel, N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, Y. Zhou, W. Li, P.J. Liu, Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. The Journal of Machine Learning Research **21**(1), 5485–5551 (2020). https://doi.org/10.48550/arXiv.1910.10683

[79] P. Rajpurkar, J. Zhang, K. Lopyrev, P. Liang, SQuAD: 100,000+ Questions for Machine Comprehension of Text. arXiv preprint arXiv:1606.05250 (2016). https://doi.org/10.48550/arXiv.1606.05250

[80] The Linux Foundation. ONNX: Open Neural Network Exchange (2017). URL github.com/onnx

[81] F. Chollet. Keras (2015). URL github.com/fchollet/keras

[82] O. Zafrir, A. Larey, G. Boudoukh, H. Shen, M. Wasserblat, Prune Once for All: Sparse Pre-Trained Language Models. arXiv preprint arXiv:2111.05754 (2021). https://doi.org/10.48550/arXiv.2111.05754

[83] M. Eswaran, A.K. Gulivindala, A.K. Inkulu, M. Raju Bahubalendruni, Augmented reality-based guidance in product assembly and maintenance/repair perspective: A state of the art review on challenges and opportunities. Expert Systems with Applications **213**, 118,983 (2023). https://doi.org/10.1016/J.ESWA.2022.118983