Article in Press IJIMAI journal

Large Language Models for in Situ Knowledge Documentation and Access With Augmented Reality

Juan Izquierdo-Domenech, Jordi Linares-Pellicer, Isabel Ferri-Molla *

Valencian Research Institute for Artificial Intelligence (VRAIN), Universitat Politècnica de València (UPV), València (Spain)

Received 16 February 2023 | Accepted 22 June 2023 | Early Access 29 September 2023

Abstract

Augmented reality (AR) has become a powerful tool for assisting operators in complex environments, such as shop floors, laboratories, and industrial settings. By displaying synthetic visual elements anchored in real environments and providing information for specific tasks, AR helps to improve efficiency and accuracy. However, a common bottleneck in these environments is introducing all necessary information, which often requires predefined structured formats and needs more ability for multimodal and Natural Language (NL) interaction. This work proposes a new method for dynamically documenting complex environments using AR in a multimodal, non-structured, and interactive manner. Our method employs Large Language Models (LLMs) to allow experts to describe elements from the real environment in NL and select corresponding AR elements in a dynamic and iterative process. This enables a more natural and flexible way of introducing information, allowing experts to describe the environment in their own words rather than being constrained by a predetermined structure. Any operator can then ask about any aspect of the environment in NL to receive a response and visual guidance from the AR system, thus allowing for a more natural and flexible way of introducing and retrieving information. These capabilities ultimately improve the effectiveness and efficiency of tasks in complex environments.

Keywords

Augmented Reality, Deep Learning, Multimodal Interaction, Large Language Models, Transformers.

DOI: 10.9781/ijimai.2023.09.002

I. Introduction

Augmented Reality (AR) and its capability for superimposing synthetic elements on top of real environments has been, indeed, a key factor in the rise of Industry 4.0 [1]. There are numerous definitions of AR, with one of the most well-known being the one proposed by Azuma: *"AR is a system that supplements the real world with virtual (computer-generated) objects that appear to coexist in the same space as the real world"* [2]. Billinghurst also defines AR as an interactive experience in which real-world objects are enhanced by computer-generated perceptual information [3]; nonetheless, AR had been applied in the industry field even before such definitions [4]. The ability to enhance environments with this technology has been utilized in various industrial applications, including product design, process design and control, maintenance processes, and learning. Its benefits have been widely demonstrated. Examples include using AR to visualize and manipulate 3D models during the design process, to provide real-time guidance and instructions for Maintenance, Repair, and Overhaul (MRO) tasks, and to enhance training and education programs through interactive simulations and visualizations. Industry 4.0 lays on several pillars, such as the Industrial Internet of Things (IIoT), cloud computing, additive manufacturing, and AR; however, the latter is unique in that it focuses on the human factor [5]. On the other hand, shop floor operators have seen how their roles and required knowledge have been transformed to match completely different profiles, leading to a need for more skilled operators with advanced education in the use of technologies [6]–[9]. AR can serve as an assistive technology to support shop floor operators in these environments.

The evolution of Artificial Intelligence (AI) and its integration into the industry is one the most critical components behind what it has been defined as Industry 4.0. It can lead to a shift in the role of workers towards more value-added tasks, which can increase job satisfaction and improve overall productivity. By incorporating these technologies, manufacturers can create a more efficient and flexible workforce, ultimately leading to a better future of work in the manufacturing industry [10]. The current proposal is a step forward in achieving these objectives.

The proper training of operators is always the first challenge to be met to guarantee their subsequent ability to work effectively and efficiently. Apart from the emergence of new possibilities in this training, such as multimedia tools, Virtual Reality (VR) and AR, 'one-to-one' training is still very beneficial. Direct interaction is still a precious element in training in complex contexts, such as laboratories, control centers, and shop floors in the industry. However, after training, operators need immediate access to documentation that can solve new doubts or problems that may arise at any time. On these occasions, the presence of a specialized expert is very rare or

* Corresponding author.

E-mail addresses: juaizdom@upv.es (J. Izquierdo-Domenech), jlinares@dsic.upv.es (J. Linares-Pellicer), isfermol@upv.es (I. Ferri-Molla).

Please cite this article in press as:
J. Izquierdo-Domenech, J. Linares-Pellicer, I. Ferri-Molla. Large Language Models for in Situ Knowledge Documentation and Access With Augmented Reality, International Journal of Interactive Multimedia and Artificial Intelligence, (2023), http://dx.doi.org/10.9781/ijimai.2023.09.002

impossible. In contrast to initial training, it is unlikely that the expert or Subject Matter Expert (SME) will be available for the operator's day-to-day work [11].

It is, therefore, essential to develop solutions that enable the operator to access information quickly and efficiently in case of need. In this context, there is a need to know what means and interrogation mechanisms are available. An ideal solution would offer multiple interaction options, including operators' ability to ask questions and receive answers in Natural Language (NL), as discussed in [12]. It is necessary to consider technical documentation and expert knowledge to provide adequate answers. It is very interesting to offer not only the possibility to give answers in NL to the operator based on the technical documentation, but also the information provided by the experts; however, technical documentation and expert information are typically unstructured, presenting a significant challenge for creating operator assistance systems in complex environments. This often leads to the creation of time-consuming, *ad hoc* solutions for different environments, which can be overwhelming and cause an excessive workload, specially in environments with a high degree of diversity or work volume. Therefore, it is necessary to address the issue of unstructured technical documentation and expert information to create operator assistance systems that are efficient, effective, and sustainable in complex environments.

One of the latest technologies based on Deep Learning (DL), Large Language Models (LLMs), can aid in NL interaction and information retrieval by operators. The proposed system enables experts to train operators on the job, allowing for the system to serve as a knowledge source for subsequent consultations. This type of learning, known as in situ learning or Scenario Based Training (SBT), has been demonstrated to enhance knowledge acquisition and retention according to prior research [13].

One-on-one training, primarily SBT, is still the best way to provide knowledge of complex systems. Combining SBT with an automatic acquisition of information, adding multimodal elements, avoiding the need to structure or post-process the information, and making this knowledge available to the operators, is an element of great interest, and the main interest of this work.

Another issue when discussing complex environments, such as shop floors, is documentation access. The complexity of these environments tends to increase exponentially, as does the specialized knowledge and technology required by operators. Documentation about the different machines spread over a shop floor is critical for making them work and learning and maintenance processes. However, traditional forms of documentation, such as paper manuals, can be cumbersome and difficult to use due to their lack of portability, the potential for inaccuracies, and interpretation issues. As Ventura highlights, these issues can make it difficult to effectively utilize this type of information [14]. To address these challenges, several alternatives using AR technology have been proposed. For example, AR can be used to display machine-specific documentation on a user device in real-time as they work, allowing for easier reference and reducing the risk of errors due to outdated or incomplete information [8], [15], [16]. In the context of Industry 4.0, the need for such accessible and reliable information becomes even more pressing as the demands for decentralized, accurate, modular, and fast access to information become increasingly important [17]. By utilizing AR technology, it may be possible to improve the accessibility and usability of documentation in complex environments such as shop floors, ultimately leading to increased efficiency and productivity.

AI and its subsets, such as Machine Learning (ML) and DL, also play an indispensable role in the industry 4.0 field. The capabilities to develop solutions that range from Computer Vision (CV), Natural Language Processing (NLP), and finding patterns hidden in vast amounts of data are being applied in tasks such as predictive maintenance [18], process automation [19], or security enhancement [20]. Some architectures that enable developing applications that tackle these kinds of tasks are Convolutional Neural Networks (CNNs) for CV and Transformers for NLP. CNNs are a type of neural network architecture typically used for image classification, while Transformers have revolutionized NLP tasks by allowing for attention-based mechanisms to capture semantic dependencies between words. Transformers are behind the current LLMs and are mainly used for tasks such as language translation, text summarisation, sentiment analysis, question answering (QA), and language modeling [21]. AI tools are used to analyze large amounts of data, automate repetitive tasks, and improve decision-making processes, leading to increased efficiency, cost savings, and improved customer experiences. AI-powered systems are also being used to monitor, predict, and prevent potential equipment failures and downtime, reducing maintenance costs and increasing overall productivity.

This study aims to evaluate the effectiveness of using multimodal interaction and AR to enrich complex environments with additional information from a variety of sources (e.g., technical documentation, experience-based knowledge) in an unstructured format and to assess the feasibility of novice workshop operators accessing this information multimodally by anchoring it in the environment through AR.

The main contributions of this work are:

1. The ability to incorporate the knowledge and experience of SMEs in a flexible format and to continually update it through an iterative process,
2. The collection of information in multiple formats, anchored in the physical space and using NL, to reduce the need for access to technical documentation and SMEs,
3. Reducing the time between the emergence of a doubt and receiving a response.

This paper is structured as follows: In section II, numerous examples of AR and AI being applied to industrial settings are enumerated and described to emphasize this research's novelties. Then, in section III, the principal user roles and technology used for this research are explained. Section IV presents and evaluates the results from the experiment. Finally, in section V, the conclusions are described, highlighting the significance of the findings.

II. Background and Context

Technological advancements have led to the integration of AR and AI into various industries in recent years. These technologies have the potential to revolutionize the way shop floor operators and SMEs interact with and perceive the environment around them. This section will explore some of the current state-of-the-art applications of AR and AI in the industry, highlighting their potential impact and future possibilities.

A. AR in Industry

As a rapidly emerging technology, AR is increasingly being adopted across various industrial sectors, providing plenty of potential applications that can improve efficiency, productivity, and safety. There is a considerable amount of AR applications in the industry, and some examples are step-by-step guides [22], manufacturing [4], [23], [24], design [25], [26] and evaluation [27], [28]. Wang et al. [29] highlights in their literature review the need for research in several aspects when applying AR to the industry field, such as knowledge representation and contextual awareness. AR can provide many benefits in product design, allowing for faster and more collaborative tasks [30]–[34]. Process design and control is another field of interest, as indicated by Elia et al. [35], and several applications and systems have been developed [36], [37], bringing to attention the benefits of using AR in this field of application. Regarding maintenance, much interest has been put into solving challenging problems, such as reducing the Mean Time To Repair (MTTR) [38], reducing the cost of having SMEs on site [39], guiding in bad viewing conditions [40], or focusing on the operators' safety [41]. There has also been research about using AR for accessing information, such as the ARES framework, to adapt the information shown to the operator depending on several conditions, such as the time to perform a task [42]. This highlights the importance of the different roles and experiences on the shop floor and how the interface must show more or less information depending on these characteristics. Additionally, the findings indicate that using authoring tools by SMEs makes creating instructions more efficient and user-friendly. For example, Palmarini et al. developed the FARA authoring tool, which facilitates the creation of step-by-step AR animations for various procedures, such as maintenance, repair, and overhaul (MRO) [43].

B. AI in Industry

Equally important, AI is being applied in industrial fields rapidly, alongside AR in various domains. MRO, diagnosis, and predictive maintenance are among the fields where AI has found widespread applications. Predicting a possible error in the system before it happens leverages AI to foresee potential system failures before they occur, thus enabling proactive maintenance and increased operational efficiency. Both Carvalho et al. [18], and Zonta et al. [44] perform a systematic literature review where several ML and DL models are being applied for predictive maintenance, demonstrating the increasing research interest in this field. Other applications focus on customer support, where chatbots and recommendation systems can help companies provide faster and more personalized support to clients. Casillo et al. develop a chatbot framework for real-time assistance and efficient and personalized training [45]. Pattern recognition and prediction are, by nature, key applications of ML and DL algorithms. Detecting patterns in vast amounts of data might help data scientists obtain valuable insights, for example, to predict changes in product demand. Moroff et al. evaluate several ML and DL models such as Random Forest, XGBoost, Long-term short-term memory (LSTM) networks, and a multilayer perceptron (MLP), among others, as forecasting models [46]. Finally, AI models are being increasingly applied in the field of automation. Operators' previous tasks can be automatized intelligently, such as optimizing production processes and improving customer support. Maschler and Weyrich highlight, in their literature review, several studies in fields such as anomaly detection, time series prediction, fault diagnosis and prognostics, quality management, and computer vision [47].

C. Synergy Between AR and AI in Industry

As a complementary element to AR, AI opens new synergy possibilities. In the field of information access, Chidambaram develops a solution utilizing AR and the YOLO foundation model [48] to generate instructions that differentiate between novice users and SMEs [49]. As described by Standford, a foundation model means to *"Train one model on a huge amount of data and adapt it to many applications"*, or in other words, it is a model that has been pre-trained and provides various features that can be utilized for transfer learning or fine-tuning to fit specific requirements [50]. Examples of foundation models are YOLO for object detection, Stable Diffusion for image generation [51] or GPT for text generation [52]. Our previous research has focused on developing tools that guide and enhance the safety of shop-floor operators using AR and AI [12]; however, the present proposal in this work takes a closer look at the other side of the equation, the SMEs and how to use AR and AI to enrich the environment with information in a comfortable manner. Little research has been done regarding the use of these two technologies in enhancing unstructured information management and access, and this work proposes an approach to fill this gap.

D. Documentation Management and Access

With its human-centric view, the advent of Industry 5.0 [53], [54] brings about significant challenges in the realm of documentation and information access, owing to various factors such as decentralization, virtualization, and modularity [8]. This highlights the need for more effective methods for managing documentation. In light of the need for information to be easily accessible, updatable, and translatable, paper-based documentation is becoming obsolete. Further research must be conducted in this area, as several authors have emphasized [14]–[16]. This study seeks to address a key challenge in shop floor operations by exploring novel strategies to enhance information accessibility. The ultimate aim of this proposal is to leverage the knowledge and expertise of SMEs to create dynamic environments, enhancing them with knowledge anchors into spatial 3D real environments to improve efficiency and profitability.

E. Information Retrieval and Mental Decay

In accordance with the discussion presented in section I, the most optimal way to gain expertise in an industrial setting is to perform SBT and personalized tutelage with an SME; however, one of the most critical issues associated with this process is the maintenance of the acquired knowledge, particularly its tendency to deteriorate over time.

Mental decay, also known as knowledge decay, is a passive process in which the knowledge and skills of a person gradually decline over time if not actively reinforced. Studies have shown that mental decay can occur even when an individual is exposed to new information, with decay increasing as the time between exposure and retrieval increases [55]. Numerous studies have been conducted on knowledge retention and information retrieval in industrial settings in recent years. One such study by Adesope et al. found that repeated exposure to information leads to better knowledge retention compared to solitary exposure [56]. This finding is supported by other studies, such as the work of Karpicke and Roediger, who showed that retrieval practice can enhance long-term retention of information [57]. In addition, research has also been conducted on the impact of aging on knowledge retention and retrieval. For example, Bissig and Lustig found that older adults experience greater difficulty retrieving information from long-term memory compared to younger adults [58]. This finding has important implications for industrial settings, as the aging workforce is becoming increasingly prevalent and might be a focus of interest in future research [59].

In industrial environments, knowledge about machines and elements on the shop floor is often distributed through multiple documents and SMEs. Hence, having a reliable source for accessing and retrieving this information is essential. Information access is of paramount importance in industrial settings as it plays a critical role in ensuring the efficient performance of operations. Understanding how the knowledge provided to operators fades over time becomes increasingly important. It is essential to note that the operators involved in the experiments were only given the task to perform with prior knowledge. As explained in section IV, operators were subjected to repeated exposures of the same information because this can lead to better knowledge retention compared to a solitary exposure [60],[61].

The proposed tool aims to fulfill this gap of mental decay, thus providing access to technical and SME information at all times.

III. Proposed System

One of the significant challenges faced by shop floor operators in industrial settings is knowledge retrieval from the environment, as previously discussed in section I and II. It has been discussed that having an SME and utilizing SBT may be the ideal solution, but not always feasible in practice. Furthermore, mental decay adds to the difficulty as the shop floor operator may not always retain all of the information taught. Although AR applications have been proposed as a means of providing additional information in a context-aware system boosted by AI systems; accessing information naturally when technical documentation and SMEs are the only sources of information remains a challenge. This section presents a detailed description of the proposed system, highlighting the key roles of the SME, the shop floor operator, and their interactions with the system. This research aims to address the gap in the literature and justify the main contributions outlined in section I.

A. SME: Context Enrichment With Information

The SME is an expert who has acquired extensive knowledge in a particular field or topic; however, disseminating their knowledge and its contribution to the field remains challenging. While the possibility to ask the SME in case there is doubt exists, it may only sometimes be feasible, as the constant presence of an SME in the work environment may not be practical [11], [39]. Indeed, AR systems can reduce the cost of having SMEs on-site, but the challenge remains in effectively transferring the SME knowledge to the worksite. To bridge the gap in knowledge transfer to the site, this study proposes an architecture that considers the SMEs roles as a "Knowledge Transfer Experts".

The SMEs are responsible for utilizing the system to introduce "pills" of knowledge across the environment. In this research paper, a "pill" refers to a small unit of knowledge that can be added to a system. The term is chosen for its memorable connotation and aligns with the concept of intentional knowledge management. It is important to highlight that the presented architecture implementation relies on the fact that the environment needs to be previously scanned, a common feature in current SLAM-based AR solutions. Upon entering the environment, the SME can interact with their surroundings using touch interaction. This way, the SME can add specific "pills" of information to any element they find interesting to enrich, regardless of whether they are machines, control panels, or any other element of interest in complex environments. This information "pills" will be used by the system with two purposes:

1. To retrieve a specific "pill" linked to a specific position in the environment as-is,
2. For obtaining answers to specific questions.

The present study depicts a specialized tool that supports SMEs in contributing to a digitized environment. The tool facilitates data input through the means of either voice recognition or written text. Using ray-casting techniques, alongside touch interaction in AR, enables SMEs to pinpoint and enrich specific features of the 3D scanned mesh from the virtual environment. The interaction in the system is performed using touch input that is implemented differently depending on the final device used. Specific AR devices can trigger the touch action with hand-specific controls or even by using hands, while for mobile devices like tablets, touching the screen at the desired object triggers the touch action. In both cases, the interaction is implemented using ray-casting, which calculates a line or ray from the touch 2D coordinates and with the direction derived from the camera frustum. Then, the ray intersections are checked, and the object selected is the one that is closest to the user in 3D coordinates. The AR library maintains a congruent mapping of spatial coordinates between the physical and virtual worlds, resulting in accurately identifying elements in the 3D virtual space. A visualization of the SMEs task in the environment is displayed in Fig. 1.

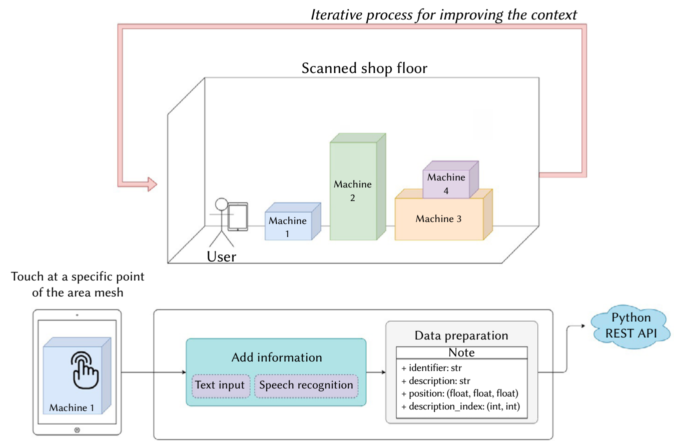


Fig. 1. SMEs annotate scanned environments.

B. Shop Floor Operator: Information Retrieval

Once the site has been enriched with anchored "pills" of knowledge, it is time for the shop floor operator to utilize these resources, reducing the frequency of their need to seek clarification from the SME and technical documentation.

As depicted in Fig. 2, the presented application offers two alternatives for accessing such information in a multimodal manner:

Touch & Area Selection Since many anchors might be disseminated around the site, the shop floor operator has the option to select a rectangular area and retrieve all the "pills" within that selection, as shown in Lane C in Fig. 3. Applying the ray-casting methods as introduced in section III.A, the system can obtain the 3D virtual coordinates of a chosen location on the shop floor by utilizing touch interaction. The process involves the shop floor drawing an area of interest from which anchored "pills" can be retrieved. The approach leverages the same ray-casting techniques employed by SMEs in the aforementioned section, demonstrating the tool's versatility across multiple domains.

Speech Recognition for NL Queries While the *touch and area selection* method is proactive, the system also includes a more reactive approach to obtaining information. The shop floor operator can ask any query in NL, such as "What temperature does glycol evaporate at?". The system will then process the query and provide the answer (e.g., "360°") as well as the location within the work environment where the SME anchored the corresponding "pill" of knowledge. This is also shown in Lane B in Fig. 3.

C. System Implementation

For evaluating the proposed system, a mobile application has been developed using the Unity platform, which allows for developing applications for both Android and iOS devices. AR through the Vuforia library has been integrated for scene recognition (i.e., the scanned laboratory). For Speech Recognition on device, the Vosk toolkit has been used, which allows for real-time voice recognition in diverse languages. Regarding the server side, since the information must persist between application uses, the Python FastAPI framework has been utilized for generating the different endpoints using a RESTful API.

Fig. 3 briefly summarises the application's key features. Lane A highlights the capabilities of SMEs within the app. After detecting the environment, the SME can add pieces of information, either handwritten

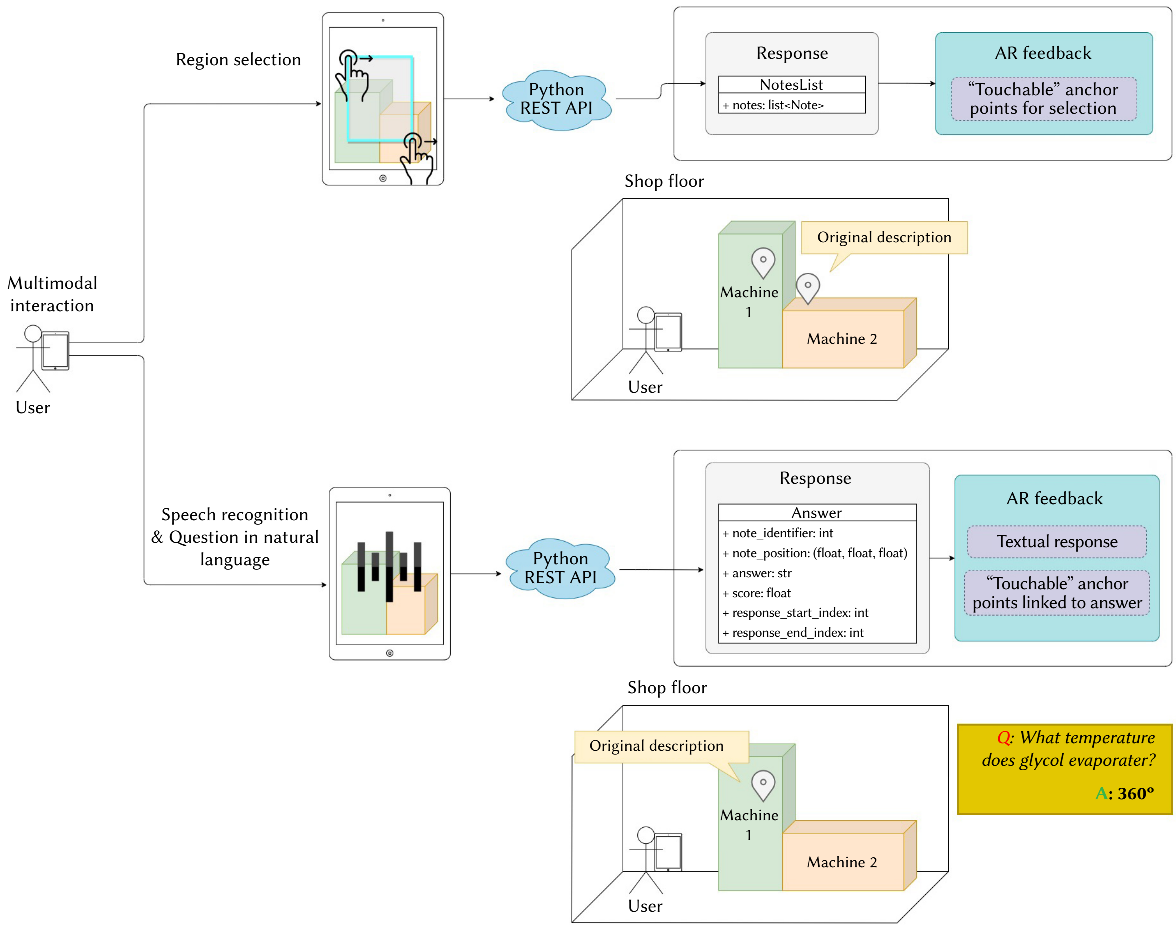


Fig. 2. Shop floor operator retrieves anchored information via NL query or area selection.

Fig. 3. Two shop floor roles: SMEs add information, operator retrieves it via NL/AR.

or by voice, at any point, using touch interaction. Lanes B and C outline the interaction from the shop floor perspective. In Lane B, the operator can use the application to ask, in NL, about anything regarding contextual information. This query is then sent to the transformer for processing. The specific transformer architecture is explained in detail in subsection D. The result is displayed not only as a text answer but also as a highlighted location in the environment where the SME added the information. Finally, Lane C briefly shows that the operator can retrieve any information in any environment by drawing a rectangle around the area of interest using touch interaction. The application will display all notes anchored by the SME within that area.

D. Consistency of the Information

Ensuring the consistency of textual information units when integrating them into a system is a crucial factor to consider. It is important to estimate whether a new block of textual information is contradictory, redundant, or provides new information content, particularly in a proposal where unstructured information is the fundamental input. The topic of consistency in LLMs is a recurring subject of study, as noted by Elazar et al. in their work on measuring LLMs [62]. In this context, consistency is treated systematically and comprehensively, with a focus on ensuring the correctness of the answers provided by the model prior to paraphrasing the input questions.

In this study, the main focus is on consistency, which involves ensuring that there are no redundancies or contradictions when introducing a new block of information related to a particular aspect of the environment. Redundant information can reduce the efficiency of the model, while contradictory information is even more critical. It is crucial to prevent SMEs from providing conflicting information about the same element, as this can cause problems when correcting subsequent information queries. Such conflicts may arise not only from an SME's error but also from speech-to-text conversion, for instance. To solve this problem, LLMs allow a fine-tuning process after pre-training to improve their behavior when faced with more specific problems. For this specific problem we use a Transformer architecture.

Transformer architecture is a neural network model composed of two key components: an encoder and a decoder. The encoder comprises multiple layers comprising two sub-layers: an attention mechanism and a fully connected feed-forward neural network. The decoder, on the other hand, follows a similar structure. However, in this case, each sub-layer comprises three sub-layers: the attention sub-layer, the feed-forward one, and a third sub-layer in which multi-head attention is applied to the output of the encoder. Using this mechanism makes it possible to have better results than with recurrent networks since it is possible to take into account the semantics of the input sentence more efficiently. In addition, the training process can be unsupervised; however, it is necessary to use significant amounts of unlabelled text. Due to the cost of training a transformer from scratch, not only in terms of time but also in terms of computational units and the amount of data needed to obtain a good performance, it is common to use pre-trained models. That is, using architectures that have already been trained with vast amounts of data. This allows the pre-trained transformers to have already learned most of the semantics of NL, so they can process and answer most of the questions or suggestions that the user asks in NL in a very flexible way. However, it is important to carefully select the dataset to fine-tune the model, as it is of balanced and representative data.

It is possible to find a wide variety of transformers with a wide range of capabilities in NL processing. Such as machine translation between language pairs, text summarisation, text-relevant QA, or conversational systems. Among the most commonly used transformer-type pre-training systems currently in use are DistilBERT [63], RoBERTa [64], Google's T5 [65], BLOOM [66], or GPT-3 [52]. Arroni et al. [67] provide a compelling example of the effective use of LLMs in their work on semantic classification.

In the case of this project, the GPT-JT model was used as resulting of the fine-tuning of the GPT-J model with UL2 training objectives [68], [69], achieving results similar to models of 175B parameters in many tasks, such as InstructGPT davinci v2, but with only 6B parameters [70]. GPT-JT has been trained in a decentralized way and allows its free download and use, as well as its installation and local use. The final LLM used in the final solution can be adapted to specific final requirements of the solution and available options. GPT-JT was a good commitment in evaluating the proposed solution, allowing good results in QA on technical documentation and a high degree of flexibility on few-short learning. Few-shot learning means that the model can be presented with one or several examples of the task to be solved to achieve higher degrees of accuracy in its responses.

With this purpose, the GPT-JT model has been tested to detect if a new piece of textual information is redundant or in contradiction with the previous information assigned to a specific element of the scene of the AR environment.

Although the main aim of this research was not to compare various models on the same instruct-based tasks, we set specific criteria for selecting the most appropriate model. Based on its Open Source availability, superior performance against instruct-based queries (as per the Hugging Face ranking), and ease of installation on local servers (6B parameters only), the GPT-JT model was chosen. The preliminary findings indicated that the chosen model classified the information consistently.

Table I shows a concrete example used to discriminate between "new", "contradictory", or "redundant" with few-shot learning, in this case, a single training example. Using only one example, it has been possible to verify the correctness of the classification of the new information block in all the tested examples. An exhaustive evaluation of this approach transcends the objectives of the present proposal, more typical of computational linguistics, requiring the creation of comprehensive and specific evaluation datasets which, in the best of cases, cannot guarantee their results in the face of domain problems.

The proposal in this paper focuses on using the few-shot learning consistency test of the transformer to detect redundancy or contradiction problems better and visually notify the SME of this possibility. When the SME is warned of a possible redundancy or contradiction, it can examine the text entered associated with an element and repeat in case of a possible error or reconfirm the correction of the new information element.

IV. Evaluation and Results

A. Experimental Setup

Experiments were conducted in a textile laboratory at Universitat Politècnica de València, Campus d'Alcoi. The section of the laboratory that was scanned for subsequent identification is a representative sample of an overall facility. It was selected because it contains a variety of equipment commonly used in textile manufacturing and provides a suitable environment for testing the developed system. The equipment used in the experiments includes machines for fabric dye testing, material cleaning, and emulsion homogenization, all essential for producing high-quality textile products. The laboratory's wide range of equipment and diverse capabilities make it an ideal environment for simulating potential scenarios, providing a representative setting for testing and evaluating various approaches and solutions. To evaluate the developed application, we selected an iPad Air device because of its compatibility with the system and development tools and its ease of use for the operators.

TABLE I. GPT-JT Tests Label Information As "New", "Redundant", Or "Contradictory" Based on Context

Context	Input	Output
Few-shot learning		
To turn on the machine switch on the red button. *To turn off the machine switch off the red button.*	*To turn on the machine it is necessary to switch on the red button.*	"redundant"
Same context.	*To turn on the machine it is necessary to switch on the blue button.*	"contradictory"
Same context.	*To pause the machine it is necessary to switch on the blue button.*	"new"
After few-shot learning...		
The emulsion is homogenized with an agitator. *One field of use is microcapsule emulsions or cosmetic creams.* *At the top of the panel is the button to raise and lower the agitator.* *In the central part of the panel are the buttons to turn on and off, and a wheel to control the number of Revolutions Per Minute (RPM).* *At the bottom, we find the motor and ignition indicators.*	*To lower the agitator, use the button on the top of the panel.*	"redundant"
Same context.	*The machine is called homogenizer.*	"contradictory"
Same context.	*We can find the buttons to turn off the machine at the bottom.*	"new"

B. Participants

As far as participants are concerned, a textile master teacher served as the SME for explaining and adding information. Two groups of participants were selected to evaluate the system. 30 participants were selected and distributed between groups A and B.

The 30 participants recruited for the study were all master's students in engineering, ranging in age from 22 to 28 years old, with an equal distribution of male and female participants. While all participants had previous experience with similar machines, none were familiar with the specific machine used in this study, making it a novel task for all participants.

While both groups were exposed to the explanation of the SME, during the system evaluation phase, group A had access to the documentation about the machines to be utilized and the SME. In contrast, group B had access to the developed application. The only restriction applied to group B was that they were instructed to use the application first for any questions about the machines. If the answer from the application was incorrect or lacked enough information, they were allowed to search in the technical documentation and ask the SME.

Both groups, A and B, were exposed to information about the environment and the machines they were going to use during the experiment. The information exposure took place simultaneously a week before the experiment for both groups. The information was presented by a SME, who, at the same time, introduced the information into the system. Three days before the experiment, both groups were again presented with the same information to boost retention.

C. Tasks

The participants were instructed to perform several interactions with three types of machines; fabric dye testing (Task 1), material cleaning with an ultrasonic machine (Task 2), and homogenizing emulsions (Task 3). Fig. 4 compares the completion time between groups A and B while performing the same tasks in seconds. To ensure objective and consistent measurements, the data was collected by a single external observer who followed standardized procedures throughout the data collection process.

The following standardized procedures were employed by the observer:

1. **Training and Familiarization**: The observer underwent comprehensive training to become familiar with the research goals, the tasks to be performed, and the specific machines involved. This training aimed to ensure that the observer had a thorough understanding of the procedures and requirements for accurate data collection.

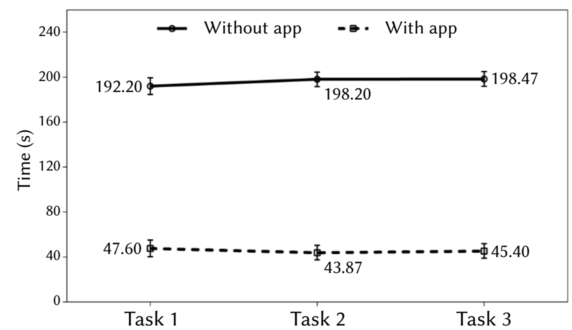


Fig. 4. Time comparison between groups A (No app) and B (With app).

2. **Non-Intrusive Observation**: The observer adopted a non-intrusive approach to minimize any potential influence on the participants' behavior or performance. The observer focused on discreetly observing the participants without interfering with their interactions or affecting the natural flow of the tasks. This approach aimed to ensure that the participants' actions were representative of their usual behavior.
3. **Data Recording**: The observer used the same data collection sheet to record relevant information during the observation process. This included capturing the start and end times of each task, any relevant observations or notes, and any additional contextual information that could be important for later analysis.

This approach allowed for precise time measurement and reduced the potential for biases or errors that could arise from multiple observers or inconsistent methods. It is clear that there is a significant reduction in time when specific information needs to be retrieved, thereby supporting the second and third main contributions of this research: anchoring information in the environment reduces the need for consulting technical documentation and SME, and the reduction in time between the emergence of a doubt and receiving a response.

D. Results

Table II shows the results of the two-factor ANOVA with repeated measures on one of them performed to determine if the effect of the group influences the task execution time. The results show that there is a statistically significant difference between the groups, regardless of the task ($p < 0{,}001$) such that the task execution time for workers using the app (45,62 seconds) was significantly lower than for workers not using the app (196,29 seconds). No differences were observed between tasks ($p = 0.964$), and the group and task interaction were not significant ($p = 0{,}781$). The term "group and task interaction" refers to

TABLE II. Comparison of Time Between Groups A and B, Along With the Corresponding P-values

	Task			Effects tests		
	1	2	3	Group	Tasks	Group*Task
	Mean (*Sd*)	*Mean* (*Sd*)	*Mean* (*Sd*)	***F*** **(g.l.);** ***p*****-value** (η^2)	***F*** **(g.l.);** ***p*****-value** (η^2)	***F*** **(g.l.);** ***p*****-value** (η^2)
Time (seconds)				F (1;28) = 1.545,22; p <0,001 (0,982	F ((2;56) = 0,04; p = 0,964 (0,001)	F ((2;56) = 0,25; p = 0,781 (0,009)
Group A	192, 20 (34, 32)	198, 20 (31, 26)	198, 47 (32, 45)			
Group B	47,60 (20, 28)	43, 87 (15, 24)	45, 40 (13, 62)			

TABLE III. Likert Questionnaire

	Min-Max	Mean (Sd)
I found the AR app easy to use for accessing information about the machines.	3-5	4 (0,85)
The information provided by the AR app was accurate and reliable.	2-5	3,27 (0,8)
The AR app helped me perform my tasks more efficiently.	3-5	4,6 (0,63)
The information provided by the AR app was useful and relevant to my needs.	4-5	4,8 (0,41)
Information was provided by the AR app quickly when I requested it.	3-5	4,07 (0,88)
I did not encounter errors or issues while using the AR app.	2-5	3,47 (0,74)
I did not need to refer to technical documentation/technical operator in addition to the AR app to find the information I needed.	3-5	3,93 (0,7)
The information provided by the AR app was helpful in completing my tasks.	4-5	4,73 (0,46)
The AR app made it easier to access information compared to other methods I have used in the past. I would	4-6	4,47 (0,64)
be willing to use the AR app on a regular basis as part of my workday.	4-5	4,33 (0,72)

the relationship between the different groups of participants (Groups A and B) and the specific tasks they performed. In this context, a non-significant interaction suggests that the effect of the group on task execution time was consistent across all tasks. In other words, the app usage had a similar effect regardless of the task performed.

Additionally, a Likert questionnaire of 5 points was delivered to the participants of group B (see table III) to measure their perceptions towards the combination of AR and anchored information retrieval in a multimodal manner and the NL interaction complement in the AR system. The questionnaire had ten questions, with values that ranged between 1 (Strongly disagree) and 5 (Strongly agree). The results support the notion that the participants had a favorable view towards integrating anchored information and its retrieval by AR systems. This indicates that this approach could potentially lead to improved outcomes in future studies and practical applications.

V. Conclusions

The roles of SMEs and shop floor operators are essential in Industry 4.0, but even more in the future advent of Industry 5.0. AR and AI techniques are being applied to improve the efficiency and effectiveness of these roles. However, while the use of AR and AI techniques is receiving much attention, only some studies have investigated the value of SMEs as a source of information. The unstructured nature of this information makes it challenging to manage and integrate with technical documentation. In this study, we developed and evaluated a system to extract information from SMEs that can be integrated with technical documentation. We used state-of-the-art AI architectures such as Transformers and LLMs to perform useful tasks such as QA and multimodal interaction on AR systems. Our results demonstrate the potential of integrating SME information with technical documentation to reduce the time it takes for operators to access relevant information. It is worth noting that Industry 5.0 is a human-centric approach to the industry that emphasizes the value added by people in the manufacturing process. While Industry 4.0 focuses on using advanced technology to automate and optimize production, Industry 5.0 recognizes the importance of operator comfort and satisfaction in the workplace. This approach considers the physical and emotional well-being of the workforce, as well as their creativity, problem-solving abilities, and interpersonal skills. By combining the strengths of both human workers and technology, Industry 5.0 aims to create a harmonious and efficient work environment that benefits all stakeholders.

This study highlights the importance of SMEs' knowledge for improving shop floor operations; however, there is still room for improvement in automating the extraction process and maintaining the accuracy and relevance of the information. Future research could focus on developing more advanced NLP techniques to better extract and organize SMEs' knowledge while ensuring that the information remains up-to-date and reliable.

Although the Vosk toolkit was used for the system implementation for speech recognition, in the presence of noisy or industrial environments, speech-to-text accuracy can be significantly improved by employing the latest Open Source models, such as Whisper [71].

Another area for future research is integrating SME knowledge with technical documentation. It would be beneficial to investigate how different types of information can be presented in a way that is easy to access and use for operators. Additionally, there is potential for integrating AR and AI techniques with SME knowledge to further enhance the efficiency and effectiveness of shop floor operations. For example, by implementing automatic information retrieval methods using object detection models, thus allowing operators to access relevant information while exploring the shop floor quickly.

As Industry 5.0 emphasizes the human-centric nature of manufacturing, it is essential to explore ways to improve operator comfort and satisfaction in the workplace. Future studies could investigate using AR and VR technologies to create more engaging and interactive training materials or wearable technologies to monitor and improve operator well-being.

Appendix

This appendix illustrates the different requests the application can make to the server, shown in Fig. 5. It outlines the possible requests the SME and the shop floor operator can make to the server.

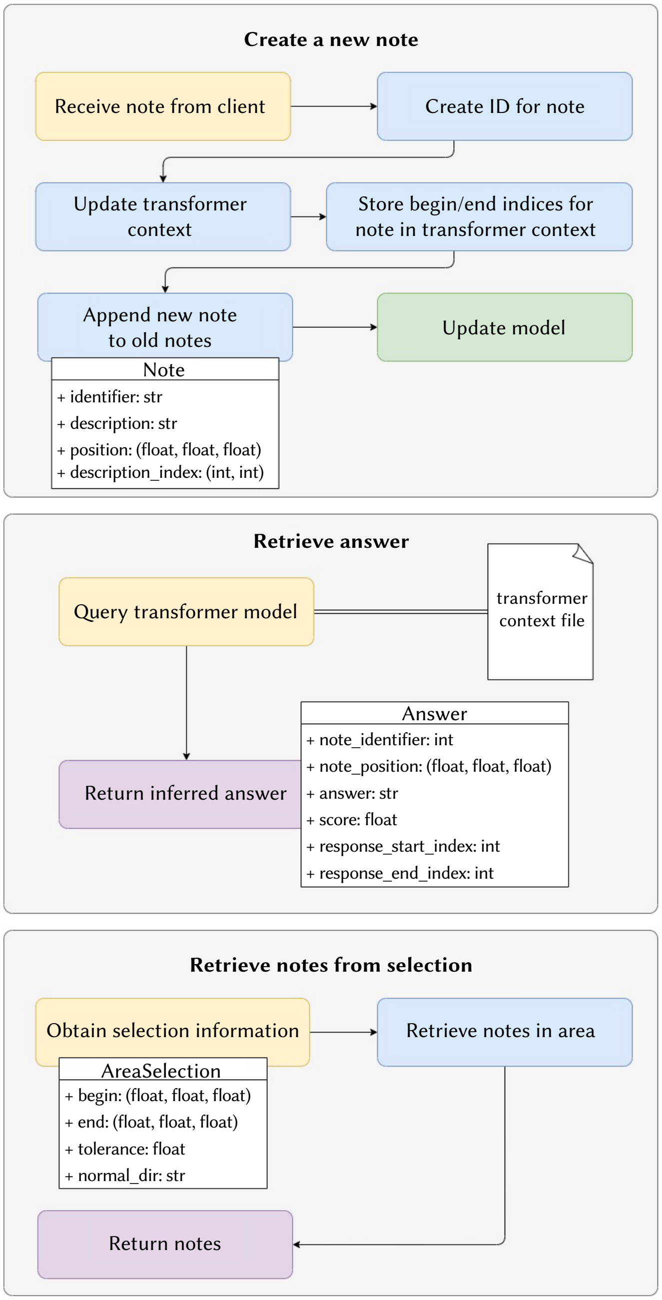


Fig. 5. Rest API calls.

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] R. Azuma, Y. Baillot, R. Behringer, S. Feiner, S. Julier, B. MacIntyre, "Recent advances in augmented reality," *IEEE Computer Graphics and Applications*, vol. 21, no. 6, pp. 34–47, 2001, doi: 10.1109/38.963459.

[3] M. Billinghurst, A. Clark, G. Lee, "A survey of augmented reality," *Foundations and Trends in Human- Computer Interaction*, vol. 8, no. 2-3, pp. 73–272, 2015, doi: 10.1561/1100000049.

[4] T. Caudell, D. Mizell, "Augmented reality: an application of heads-up display technology to manual manufacturing processes," in *Proceedings of the Twenty- Fifth Hawaii International Conference on System Sciences*, 1992, pp. 659–669. doi: 10.1109/HICSS.1992.183317.

[5] 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*, vol. 61, pp. 658–659, 10 2021, doi: 10.1016/j.jmsy.2021.05.006.

[6] S. Jaschke, "Mobile learning applications for technical vocational and engineering education: The use of competence snippets in laboratory courses and industry 4.0," in *Proceedings of 2014 International Conference on Interactive Collaborative Learning, ICL 2014*, 1 2014, pp. 605–608, Institute of Electrical and Electronics Engineers Inc. doi: 10.1109/ICL.2014.7017840.

[7] E. Marino, L. Barbieri, B. Colacino, A. K. Fleri, F. Bruno, "An Augmented Reality inspection tool to support workers in Industry 4.0 environments," *Computers in Industry*, vol. 127, 5 2021, doi: 10.1016/j.compind.2021.103412.

[8] 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*, vol. 56, no. March 2018, pp. 276–286, 2019, doi: 10.1016/j.rcim.2018.10.001.

[9] T. Masood, J. Egger, "Augmented reality in support of Industry 4.0—Implementation challenges and success factors," *Robotics and Computer-Integrated Manufacturing*, vol. 58, pp. 181–195, 8 2019, doi: 10.1016/j.rcim.2019.02.003.

[10] X. Xu, Y. Lu, B. Vogel-Heuser, L. Wang, "Industry 4.0 and Industry 5.0—Inception, conception and perception," *Journal of Manufacturing Systems*, vol. 61, pp. 530–535, 10 2021, doi: 10.1016/j.jmsy.2021.10.006.

[11] 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*, vol. 25, pp. 154–160, 2013, doi: 10.1016/j.procs.2013.11.019.

[12] J. Izquierdo-Domenech, J. Linares-Pellicer, J. Orta- Lopez, "Towards achieving a high degree of situational awareness and multimodal interaction with AR and semantic AI in industrial applications," *Multimedia Tools and Applications*, 9 2022, doi: 10.1007/s11042-022-13803-1.

[13] J. Lave, E. Wenger, *Situated learning: Legitimate peripheral participation*. Cambridge university press, 1991. doi: 10.1017/CBO9780511815355.

[14] C. A. Ventura, "Why switch from paper to electronic manuals?," in *Proceedings of the ACM conference on Document processing systems*, 2000, pp. 111–116. doi: 10.1145/62506.62525.

[15] F. Quint, F. Loch, "Using smart glasses to document maintenance processes," *Mensch und Computer 2015–Workshopband*, pp. 203–208, 2015, doi: 10.1515/9783110443905-030.

[16] C. Kollatsch, P. Klimant, "Efficient integration process of production data into Augmented Reality based maintenance of machine tools," *Production Engineering*, vol. 15, pp. 311–319, 6 2021, doi: 10.1007/s11740-021-01026-6.

[17] M. Hermann, T. Pentek, B. Otto, "Design principles for industrie 4.0 scenarios," in *Proceedings of the Annual Hawaii International Conference on System Sciences*, vol. 2016-March, 3 2016, pp. 3928–3937, IEEE Computer Society. doi: 10.1109/HICSS.2016.488.

[18] T. P. Carvalho, F. A. Soares, R. Vita, R. d. P. Francisco, J. P. Basto, S. G. Alcalá, "A systematic literature review of machine learning methods applied to predictive maintenance," *Computers and Industrial Engineering*, vol. 137, 11 2019, doi: 10.1016/j.cie.2019.106024.

[19] J. Ribeiro, R. Lima, T. Eckhardt, S. Paiva, "Robotic Process Automation and Artificial Intelligence in J. Sääski, T. Salonen, M. Hakkarainen, S. Siltanen, C. Woodward, J. Lempiäinen, "Integration of design and assembly using augmented reality," in *Micro- Assembly Technologies and Applications: IFIP TC5 WG5. 5 Fourth International Precision Assembly Seminar (IPAS'2008)*, Chamonix, France, 2008, pp. 295–404, Springer. doi: 10.1007/978-0-387-77405-3_39.

[20] T. Salonen, J. Sääski, C. Woodward, O. Korkalo, I. Marstio, K. Rainio, "Data pipeline from CAD to AR based assembly instructions," in *Proceedings of the ASME/AFM World Conference on Innovative Virtual Reality 2009, WINVR2009*, 2009, pp. 165–168. doi: 10.1115/WINVR2009-705.

[21] M. Fiorentino, G. Monno, A. E. Uva, "Tangible digital master for product lifecycle management in augmented reality," *International Journal on Interactive Design and Manufacturing*, vol. 3, no. 2, pp. 121–129, 2009, doi: 10.1007/s12008-009-0062-z.

[22] M. Fiorentino, R. Radkowski, C. Stritzke, A. E. Uva, G. Monno, "Design review of CAD assemblies using bimanual natural interface," *International Journal on Interactive Design and Manufacturing*, vol. 7, pp. 249– 260, 11 2013, doi: 10.1007/s12008-012-0179-3.

[23] L. Hou, X. Wang, L. Bernold, P. E. D. Love, "Using Animated Augmented Reality to Cognitively Guide Assembly," *Journal of Computing in Civil Engineering*, vol. 27, pp. 439–451, 9 2013, doi: 10.1061/(asce)cp.1943-5487.0000184.

[24] L. Hou, X. Wang, M. Truijens, "Using Augmented Reality to Facilitate Piping Assembly: An Experiment-Based Evaluation," *Journal of Computing in Civil Engineering*, vol. 29, 1 2015, doi: 10.1061/(ASCE) CP.1943-5487.0000344.

[25] X. Wang, S. K. Ong, A. Y. Nee, "A comprehensive survey of augmented reality assembly research," *Advances in Manufacturing*, vol. 4, pp. 1–22, 3 2016, doi: 10.1007/s40436-015-0131-4.

[26] Industry 4.0 - A Literature review," in *Procedia Computer Science*, vol. 181, 2021, pp. 51–58, Elsevier B.V. doi: 10.1016/j.procs.2021.01.104.

[27] A. Bécue, I. Praça, J. Gama, "Artificial intelligence, cyber-threats and Industry 4.0: challenges and opportunities," *Artificial Intelligence Review*, vol. 54, pp. 3849–3886, 6 2021, doi: 10.1007/s10462-020-09942-2.

[28] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, I. Polosukhin, "Attention Is All You Need," in *Advances in neural information processing systems*, vol. 30, 2017, pp. 6000– 6010. doi: 10.48550/arXiv.1706.03762.

[29] 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*, vol. 98, pp. 68–79, 6 2018, doi: 10.1016/j.compind.2018.02.001.

[30] D. K. Baroroh, C. H. Chu, L. Wang, "Systematic literature review on augmented reality in smart manufacturing: Collaboration between human and computational intelligence," *Journal of Manufacturing Systems*, vol. 61, pp. 696–711, 10 2021, doi: 10.1016/j.jmsy.2020.10.017.

[31] P. Wang, X. Bai, M. Billinghurst, S. Zhang, X. Zhang, S. Wang, W. He, Y. Yan, H. Ji, "AR/MR Remote Collaboration on Physical Tasks: A Review," *Robotics and Computer-Integrated Manufacturing*, vol. 72, 12 2021, doi: 10.1016/j.rcim.2020.102071.

[32] M. Sereno, X. Wang, L. Besancon, M. J. McGuffin, T. Isenberg, "Collaborative Work in Augmented Reality: A Survey," *IEEE Transactions on Visualization and Computer Graphics*, vol. 28, pp. 2530–2549, 6 2022, doi: 10.1109/TVCG.2020.3032761.

[33] B. Marques, S. Silva, J. Alves, T. Araujo, P. Dias, B. S. Santos, "A Conceptual Model and Taxonomy for Collaborative Augmented Reality," *IEEE Transactions on Visualization and Computer Graphics*, vol. 28, pp. 5113–5133, 12 2022, doi: 10.1109/TVCG.2021.3101545.

[34] B. Marques, S. Silva, J. Alves, A. Rocha, P. Dias, B. S. Santos, "Remote collaboration in maintenance contexts using augmented reality: insights from a participatory process," *International Journal on Interactive Design and Manufacturing*, vol. 16, pp. 419–438, 3 2022, doi: 10.1007/s12008-021-00798-6.

[35] V. Elia, M. G. Gnoni, A. Lanzilotto, "Evaluating the application of augmented reality devices in manufacturing from a process point of view: An AHP based model," *Expert Systems with Applications*, vol. 63, pp. 187–197, 11 2016, doi: 10.1016/j.eswa.2016.07.006.

[36] M. L. Yuan, S. K. Ong, A. Y. Nee, "Augmented reality for assembly guidance using a virtual interactive tool," *International Journal of Production Research*, vol. 46, pp. 1745–1767, 4 2008, doi: 10.1080/00207540600972935.

[37] S. K. Ong, Z. B. Wang, "Augmented assembly technologies based on 3D bare-hand interaction," *CIRP Annals - Manufacturing Technology*, vol. 60, no. 1, pp. 1– 4, 2011, doi: 10.1016/j.cirp.2011.03.001.

[38] D. Mourtzis, V. Siatras, J. Angelopoulos, "Real-time remote maintenance support based on augmented reality (AR)," *Applied Sciences (Switzerland)*, vol. 10, 3 2020, doi: 10.3390/app10051855.

[39] A. Gilchrist, "Introducing Industry 4.0," in *Industry 4.0*, Springer, 2016, ch. 13, pp. 195–215, doi: 10.1007/978-1-4842-2047-4.

[40] 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*, vol. 86, pp. 2033–2038, 10 2011, doi: 10.1016/j.fusengdes.2010.12.082.

[41] D. Tatić, B. Tešić, "The application of augmented reality technologies for the improvement of occupational safety in an industrial environment," *Computers in Industry*, vol. 85, pp. 1–10, 2 2017, doi: 10.1016/j.compind.2016.11.004.

[42] A. Syberfeldt, O. Danielsson, M. Holm, L. Wang, "Dynamic Operator Instructions Based on Augmented Reality and Rule-based Expert Systems," *Procedia CIRP*, vol. 41, pp. 346–351, 2016, doi: 10.1016/j.procir.2015.12.113.

[43] R. Palmarini, I. Fernández, D. Amo, D. Ariansyah, S. Khan, J. A. Erkoyuncu, R. Roy, "Fast Augmented Reality Authoring: Fast Creation of AR step-by-step Procedures for Maintenance Operations," *IEEE Access*, 2022, doi: 10.1109/ACCESS.2017.DOI.

[44] 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*, vol. 150, 12 2020, doi: 10.1016/j.cie.2020.106889.

[45] M. Casillo, F. Colace, L. Fabbri, M. Lombardi, A. Romano, D. Santaniello, "Chatbot in industry 4.0: An approach for training new employees," in *Proceedings of 2020 IEEE International Conference on Teaching, Assessment, and Learning for Engineering, TALE 2020*, 12 2020, pp. 371–376, Institute of Electrical and Electronics Engineers Inc. doi: 10.1109/TALE48869.2020.9368339.

[46] N. U. Moroff, E. Kurt, J. Kamphues, "Machine Learning and Statistics: A Study for assessing innovative Demand Forecasting Models," *Procedia Computer Science*, vol. 180, pp. 40–49, 2021, doi: 10.1016/j.procs.2021.01.127.

[47] B. Maschler, M. Weyrich, "Deep Transfer Learning for Industrial Automation: A Review and Discussion of New Techniques for Data-Driven Machine Learning," *IEEE Industrial Electronics Magazine*, vol. 15, pp. 65–75, 6 2021, doi: 10.1109/MIE.2020.3034884.

[48] J. Redmon, A. Farhadi, "YOLOv3: An Incremental Improvement," *arXiv preprint arXiv:1804.02767*, 2018, doi: 10.48550/arXiv.1804.02767.

[49] S. Chidambaram, H. Huang, F. He, X. Qian, A. M. Villanueva, T. S. Redick, W. Stuerzlinger, K. Ramani, "ProcessAR: An augmented reality-based tool to create in-situ procedural 2D/3D AR Instructions," in *DIS 2021 - Proceedings of the 2021 ACM Designing Interactive Systems Conference: Nowhere and Everywhere*, 6 2021, pp. 234–249, Association for Computing Machinery, Inc. doi: 10.1145/3461778.3462126.

[50] 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*, 8 2021, doi: 10.48550/arXiv.2108.07258.

[51] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, B. Ommer, "High-Resolution Image Synthesis with Latent Diffusion Models," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 10684–10695. doi: 10.48550/arXiv.2112.10752.

[52] A. Radford, K. Narasimhan, T. Salimans, I. Sutskever, "Improving Language Understanding by Generative Pre-Training," *OpenAI*, 2018.

[53] J. M. Rožanec, I. Novalija, P. Zajec, K. Kenda, H. Tavakoli Ghinani, S. Suh, E. Veliou, D. Papamartzivanos, T. Giannetsos, S. A. Menesidou, R. Alonso, N. Cauli, A. Meloni, D. R. Recupero, D. Kyriazis, G. Sofianidis, S. Theodoropoulos, B. Fortuna, D. Mladenić, J. Soldatos, "Human-centric artificial intelligence architecture for industry 5.0 applications," *International Journal of Production Research*, pp. 1–26, 11 2022, doi: 10.1080/00207543.2022.2138611.

[54] A. Akundi, D. Euresti, S. Luna, W. Ankobiah, A. Lopes, I. Edinbarough, "State of Industry 5.0—Analysis and Identification of Current Research Trends," *Applied System Innovation*, vol. 5, 2 2022, doi: 10.3390/asi5010027.

[55] O. Hardt, K. Nader, L. Nadel, "Decay happens: The role of active forgetting in memory," *Trends in Cognitive Sciences*, vol. 17, pp. 111–120, 3 2013, doi: 10.1016/j.tics.2013.01.001.

[56] O. O. Adesope, D. A. Trevisan, N. Sundararajan, "Rethinking the Use of Tests: A Meta-Analysis of Practice Testing," *Review of Educational Research*, vol. 87, pp. 659–701, 6 2017, doi: 10.3102/0034654316689306.

[57] J. D. Karpicke, H. L. Roediger, "The critical importance of retrieval for learning," *Science*, vol. 319, pp. 966–968, 2 2008, doi: 10.1126/science.1152408.

[58] D. Bissig, C. Lustig, "Who benefits from memory training?," *Psychological Science*, vol. 18, pp. 720–726, 8 2007, doi: 10.1111/j.1467-9280.2007.01966.x.

[59] M. Wolf, M. Kleindienst, C. Ramsauer, C. Zierler, E. Winter, "Current and future industrial challenges: demographic change and measures for elderly workers in industry 4.0," *Annals of the Faculty of Engineering Hunedoara*, vol. 16, no. 1, pp. 67–76, 2018.

[60] N. J. Cepeda, H. Pashler, E. Vul, J. T. Wixted, D. Rohrer, "Distributed practice in verbal recall tasks: A review and quantitative synthesis," *Psychological Bulletin*, vol. 132, pp. 354–380, 5 2006, doi: 10.1037/0033-2909.132.3.354.

[61] S. K. Carpenter, N. J. Cepeda, D. Rohrer, S. H. Kang, H. Pashler, "Using Spacing to Enhance Diverse Forms of Learning: Review of Recent Research and Implications for Instruction," *Educational Psychology Review*, vol. 24, pp. 369–378, 9 2012, doi: 10.1007/s10648-012-9205-z.

[62] Y. Elazar, N. Kassner, S. Ravfogel, A. Ravichander, E. Hovy, H. Schütze, Y. Goldberg, "Measuring and Improving Consistency in Pretrained Language Models," *Transactions of the Association for Computational Linguistics*, vol. 9, pp. 1012–1031, 12 2021, doi: 10.1162/tacl_a_00410.

[63] V. Sanh, L. Debut, J. Chaumond, T. Wolf, "DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter," *arXiv preprint arXiv:1910.01108*, 10 2019, doi: 10.48550/arXiv.1910.01108.

[64] 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*, 7 2019, doi: 10.48550/arXiv.1907.11692.

[65] 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*, vol. 21, pp. 5485–5551, 10 2020, doi: 10.48550/arXiv.1910.10683.

[66] T. L. Scao, A. Fan, C. Akiki, E. Pavlick, S. Ilić, D. Hesslow, "BLOOM: A 176B-Parameter Open-Access Multilingual Language Model," *arXiv preprint arXiv:2211.05100*, 11 2022, doi: 10.48550/arXiv.2211.05100.

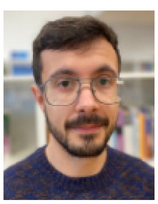
[67] S. Arroni, Y. Galán, X. Guzmán-Guzmán, E. R. Nuñez-Valdez, A. Gómez, "Sentiment Analysis and Classification of Hotel Opinions in Twitter With the Transformer Architecture," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 8, no. 1, p. 53, 2023, doi: 10.9781/ijimai.2023.02.005.

[68] Y. Tay, M. Dehghani, V. Q. Tran, X. Garcia, J. Wei, X. Wang, H. W. Chung, D. Bahri, T. Schuster, H. S. Zheng, D. Zhou, N. Houlsby, D. Metzler, "UL2: Unifying Language Learning Paradigms," *arXiv preprint arXiv:2205.05131*, 5 2022, doi: 10.48550/arXiv.2205.05131.

[69] Y. Tay, J. Wei, H. W. Chung, V. Q. Tran, D. R. So, S. Shakeri, X. Garcia, H. S. Zheng, J. Rao, Chowdhery, D. Zhou, D. Metzler, S. Petrov, N. Houlsby, Q. V. Le, M. Dehghani, "Transcending Scaling Laws with 0.1% Extra Compute," *arXiv preprint arXiv:2210.11399*, 10 2022, doi: 10.48550/arXiv.2210.11399.

[70] Together, "GPT-JT," 2022. [Online]. Available: https://huggingface.co/togethercomputer/GPT-JT-6B-v1.

[71] A. Radford, J. W. Kim, T. Xu, G. Brockman, C. McLeavey, I. Sutskever, "Robust Speech Recognition via Large-Scale Weak Supervision," *arXiv preprint arXiv:2212.04356*, 12 2022.



Juan Izquierdo-Domenech

Juan Jesús Izquierdo Doménech is an Adjunct Professor of Computer Science in Universitat Politècnica de València. He received his Bachelor's degree in Computer Science Engineering from Universitat Politècnica de València (UPV, Spain) and holds a Master's degree in Multimedia Applications from Universitat Oberta de Catalunya (UOC, Spain). He is currently performing his Ph.D. studies in UPV in the field of Human-Computer Interaction, Mixed Reality, and Artificial Intelligence.

Jordi Linares-Pellicer

Jordi Linares Pellicer is an Associate Professor at Universitat Politècnica de València (UPV, Spain), where he leads the VertexLit research group at the Valencian Research Institute for Artificial Intelligence (VRAIN). He received his Ph.D. in Computer Science from UPV and holds a Master's degree in Artificial Intelligence from Universidad Internacional de La Rioja (UNIR, Spain).

Isabel Ferri-Molla

Isabel Ferri Mollá is currently pursuing her Master's Degree in Artificial Intelligence, Pattern Recognition, and Digital Imaging at Universitat Politècnica de València (UPV, Spain). She received her Bachelor's Degree in Computer Science Engineering from UPV in 2022. Her research interests include areas of artificial intelligence, augmented reality, and human-computer interaction.