Augmented Reality-Based Training System Using Multimodal Language Model for Context-Aware Guidance and Activity Recognition in Complex Machine Operations

Waseem Ahmed and Qingjin Peng *

Posted Date: 26 January 2026

doi: 10.20944/preprints202601.1913.v1

Keywords: augmented reality; large language models; multimodal large language models; iterative design; prompt structure; coordinate measuring machine; operation training

 doi:10.20944/preprints202601.1913.v1

Article

Augmented Reality-Based Training System Using Multimodal Language Model for Context-Aware Guidance and Activity Recognition in Complex Machine Operations

Waseem Ahmed and Qingjin Peng *

Department of Mechanical Engineering, University of Manitoba, Winnipeg, MB, R3T 2N2, Canada

* Correspondence: qingjin.peng@umanitoba.ca

Abstract

Augmented Reality (AR) and Large Language Models (LLMs) have made significant advances across many fields, opening new possibilities, particularly in complex machine operations. In complex operations, non-expert users often struggle to perform high-precision tasks and require constant supervision to execute tasks correctly. This paper proposes a novel AR-MLLM-based training system that integrates AR, multimodal large language models (MLLM), and prompt engineering to interpret real-time machine feedback and user activity. It converts extensive technical text into structured, step-by-step commands. The system uses a prompt structure developed through an iterative design method and refined across multiple machine operation scenarios, enabling GPT-5 to generate task-specific contextual digital overlays directly on the physical machines. A case study with participants was conducted to assess the effectiveness and usability of the AR-MLLM system in Coordinate Measuring Machine (CMM) operation training. The experimental results demonstrate high accuracy in task recognition and feature measurements. The data further show reduced time and user workload during task execution with the proposed AR-MLLM system. The proposed system not only provides real-time guidance and enhances efficiency in CMM operation training but also demonstrates the potential of the AR-MLLM design framework for broader industrial applications.

Keywords: augmented reality; large language models; multimodal large language models; iterative design; prompt structure; coordinate measuring machine; operation training

1. Introduction

Augmented reality (AR)-based systems have emerged as promising solutions for industrial training [1]. AR has significantly improved training efficiency by providing real-time guidance, helping operators learn new skills quickly, and reducing training time and material waste [2]. In the context of complex machine operations, AR training systems enable users to visualize step-by-step guidance in their actual work environments.

Despite these advancements, most AR-based training solutions remain ineffective. In these systems, the training flow assumes that users follow instructions exactly in the same order [3]. However, in actual operations, users frequently rely on technical manuals or written instructions to perform the procedures. These documents are often lengthy, densely worded, and not always intuitive, particularly for novices who are unfamiliar with machine interfaces. Even when the text is understood, execution in a real environment is difficult because the manuals lack concrete visual guidance or the spatial context of each step. For example, a manual might state, "set the tool offset before starting the process," without indicating where the offset controls are located on the machine or how to confirm that the task has been performed correctly. Similarly, another issue arises when trainees encounter a Human-Machine Interface (HMI) that displays errors or machine feedback,

 doi:10.20944/preprints202601.1913.v1

further confusing users in determining the correct action to take [4]. Thus, extracting operational steps from raw textual data and integrating them with live machine feedback in training systems to generate actionable guidance remains a complex challenge [5].

Integration of LLMs in AR opens new possibilities, particularly in domains where procedures are complex and tasks dynamically vary [6–8]. However, existing implementations cannot infer machine states and recognize real-time user activities to provide feedback when users deviate from the expected procedure or sequence. Therefore, we propose an AR-based training system that can read dynamic manual instructions in real time and incorporate visual input in an AR environment to interpret both machine states and user activities in complex machine operations.

In this paper, we introduce an innovative augmented reality (AR)-based training system that integrates a multimodal large language model (MLLM) with prompt engineering, spatial anchoring and AR content placement to facilitate context-aware procedural guidance. In this system, we utilize ChatGPT-5 (vision-enabled configuration) to interpret real-time feedback from machine displays and recognize user actions. It further transforms extensive procedural texts into stepwise prompts and displays them using AR overlays and 3D guidance. This approach not only improves training efficiency but also provides expert assistance to users in complex operations through multimodal reasoning capability.

The remainder of this paper is organized as follows. Section 2 presents a comprehensive literature review that critically analyzes existing research on AR, LLM, VLM, and MLLM in operational planning. It also highlights the key developments and gaps in the current literature. Section 3 introduces our proposed system, including the design and integration of the MLLM in AR. Section 4 details the implementation of the proposed system. Section 5 presents the operation of the proposed system. Section 6 outlines the method used to evaluate the system's performance. Section 7 reports and discusses the results of the study. Finally, Section 8 concludes the paper and outlines directions for future research.

2. Related Work

2.1. Augmented Reality in Industrial Training

Augmented reality (AR) has been widely reported to improve industrial training and assistance by placing instructions in the worker's field of view, reducing time-to-competence and errors in Industry 4.0 scenarios [9,10]. Several studies have explored the potential of AR across various industries [11]. AR was mostly used in the gaming and entertainment industries, however, in recent years, AR technologies have become increasingly popular in professional training and industrial settings [12]. Unlike other training methods such as virtual reality (VR), AR overlays digital information over the user's real-world view, allowing users to interact with both physical and virtual aspects simultaneously [13,14]. AR has significantly improved production efficiency by providing real-time guidance [15], helping workers learn skills quickly, and reducing training time and material waste [2]. For instance, an AR-assisted guidance system for the assembly of avionics equipment was developed to provide dynamic assembly instructions overlaid onto a real-world environment using real-time pose tracking [16]. This method increased worker performance and reduced errors in assembly tasks. The advantages of AR have been further evaluated in material-forming and machining processes, where AR training resulted in faster completion times and fewer errors than traditional video or paper-based manuals [2]. AR also offers significant advantages in railway maintenance training and task execution by providing realistic and engaging content [17]. Similarly, in the tool change process, AR reduced cognitive load by overlaying visual cues and instructions on the real-world environment. This step-by-step guidance of AR enables users to perform tasks easily and efficiently, improving their overall effectiveness [18]. AR is particularly effective for initial task exposure because it reduces cognitive load compared to other technologies [19–21].

Despite these advancements, existing AR applications predominantly focus on predefined training content and assume an ideal process flow. However, these systems often struggle to keep

 doi:10.20944/preprints202601.1913.v1

pace with the nature of real-world operations, where the machine status and user activity vary in real time, creating a bottleneck for their widespread implementation. This gap highlights the opportunity to leverage AR for complex machine operations by integrating adaptive, context-aware mechanisms that interpret live conditions and provide real-time, task-specific guidance.

2.2. Integration of MLLM

For many years, industry operations planning, maintenance work, and training have relied heavily on static manuals and the experience of skilled workers. Although this approach has served its purpose, it often slows work, especially on tasks that require precision or pose safety risks. The integration of Artificial Intelligence (AI), particularly LLMs, has introduced a transformative paradigm for industrial operations [22]. Recent literature underscores this shift, highlighting that LLMs fundamentally reshape task interpretation, execution, and planning across complex workflows [23]. This transformation is exemplified by the "RoboGPT" system, which employs LLMs to generate automated step-by-step instructions for robot-based assembly tasks in construction [6]. These models can optimize assembly sequence planning by decomposing tasks into logical task steps. Beyond assembly sequences, LLMs, such as ChatGPT, have been utilized to convert textual instructions into physical steps for tasks such as electric panel maintenance. The system employs optical character recognition (OCR) to process text instructions that are then sent to the ChatGPT server. The server processes the input and transforms complex instructions into simple, sequenced operational commands. These commands are subsequently sent back to the AR system to invoke virtual objects and display relevant prompts on the physical machine [24]. Another study demonstrated that ChatGPT could successfully generate the underlying code for Web-based Augmented Reality (Web-AR) applications from natural language prompts, suggesting that LLMs can automate parts of the development workflow, thereby making AR technology more accessible to individuals without specialized programming skills [25].

Despite these significant advancements, LLMs remain limited in their understanding of physical tasks, which often require visual perception beyond language. Visual inputs provide essential context about the environment, such as the machine state and user actions, which cannot be reliably inferred from text alone. To address this limitation, recent work has leveraged multimodal large language models (MLLMs) that jointly process vision and language for more efficient scene understanding [26]. For instance, Text to Automated General-purpose Guidance in AR (TAGGAR) utilizes LVLMs to create general-purpose AR task guidance without requiring expertise or complex computer-aided design [27]. This system integrates GPT-4V to process natural language instructions and images, generating visual guidance and accurately anchoring visuals in the physical environment using "GroundingDINO," an object-detection model that locates targets and renders appropriate AR visuals. In another study, a similar vision-language model, GPT-4V, was used to provide cognitive assistance and feedback to users. This system employed a reality encoder to capture audio, images, and 3D space data and convert them into inputs for Multimodal Large Language Models (MLLMs). These inputs were processed using GPT-4V and Ferret for reasoning and object detection. Subsequently, a reality decoder converts the MLLM output into AR overlays and anchors them in real-world environments [8].

2.3. Prompt Engineering

The above discussions show that MLLMs can interpret responses in both textual and visual contexts. This is because these models are based on the transformer architecture and focus only on the instructions and parts that matter most in the context. This enables them to understand complicated instructions and generate appropriate responses [7,28]. However, when these models are used without additional input or external tools, such as structured prompts or domain-specific data, they produce inaccurate or incomplete outputs. Some studies have pointed out that existing MLLM models face challenges in interpreting correctly and struggle to apply domain rules that are obvious when deployed for general-purpose tasks [29]. This limitation makes it difficult to rely solely

 doi:10.20944/preprints202601.1913.v1

on MLLM outputs in real-world environments, particularly in industrial operations that require precision, safety, and consistency.

To overcome these issues, researchers in AI and engineering have explored various methods to make MLLMs more reliable and produce more accurate responses. Two main strategies have emerged in this regard. The first is fine-tuning LLMs on domain-specific tasks using resources such as equipment manuals, training datasets, and maintenance logs. This method has shown an increase in accuracy because the modals become familiar with the domain's workflow. However, fine-tuning can improve task relevance and accuracy, particularly in specialized industrial domains. However, it requires large, high-quality datasets, considerable training time, and ongoing updates whenever the domain changes, which is not always practical in industry. Additionally, it reduces the flexibility of the model in specific contexts [30,31].

The second increasingly popular strategy is to use prompts rather than change the model itself. This approach can guide the model step by step without modifying its core architecture. Research shows that even small adjustments in a prompt can greatly improve a model's reasoning ability and reduce errors [32]. This is further confirmed by recent comparative studies that show that prompt tuning outperforms fine-tuning, particularly in tasks that require flexibility and contextual inference [33].

Studies on maintenance, construction, and AR-based training have shown that without clear prompts, multimodal models may focus on the wrong objects, misunderstand a user's intention, or misinterpret visual details that humans would instantly recognize [27]. Thus, various prompt techniques, such as asking the model to think step-by-step, assigning it a specific role, or defining the specific output format, have been designed and have shown significant improvements in many tasks [34–38]. Another advantage of the prompt technique is that it is simple to apply, requires fewer resources than fine-tuning, and can be reused in different domains. These benefits make prompt engineering an attractive choice for real-world operations, particularly in industrial environments, where efficiency and adaptability are essential [39].

In conclusion, multimodal LLMs have shown great potential and are becoming increasingly common; however, to the best of our knowledge, existing studies have not yet deeply explored their integration for complex machine operations. The current system primarily focuses on general-purpose applications and lacks implementation for inferring machine states and recognizing real-time user activities. Moreover, industrial environments often contain many overlapping objects and cluttered backgrounds, all of which can confuse MLLMs. Thus, a structured prompt design is necessary to ensure rapid deployment and help MLLMs deliver predictable, safe, and reliable outputs in real-world industrial workflows [7].

3. System Design

3.1. Overview

The proposed system is designed to assist users in performing complex machine operations in real time by combining visual recognition and large-language models. It is connected to a vision-language model (ChatGPT-5) and deployed on the Microsoft HoloLens 2. The proposed method is built to interpret both instructions and dynamic machine states, providing real-time feedback in the user's physical environment.

3.2. System Architecture

This AR-based training system is developed using the Unity game engine (6000.0.29f1) [40], which serves as the primary component for integrating AR with AI assistance. To maintain precise alignment between virtual instructions and content on the physical machine, we use local spatial anchoring with Microsoft's Mixed Reality OpenXR and AR Foundation plugins. The core of the system is the multimodal LLM (ChatGPT-5), which is integrated into the training via the Azure OpenAI service, as shown in Figure 1.

 doi:10.20944/preprints202601.1913.v1

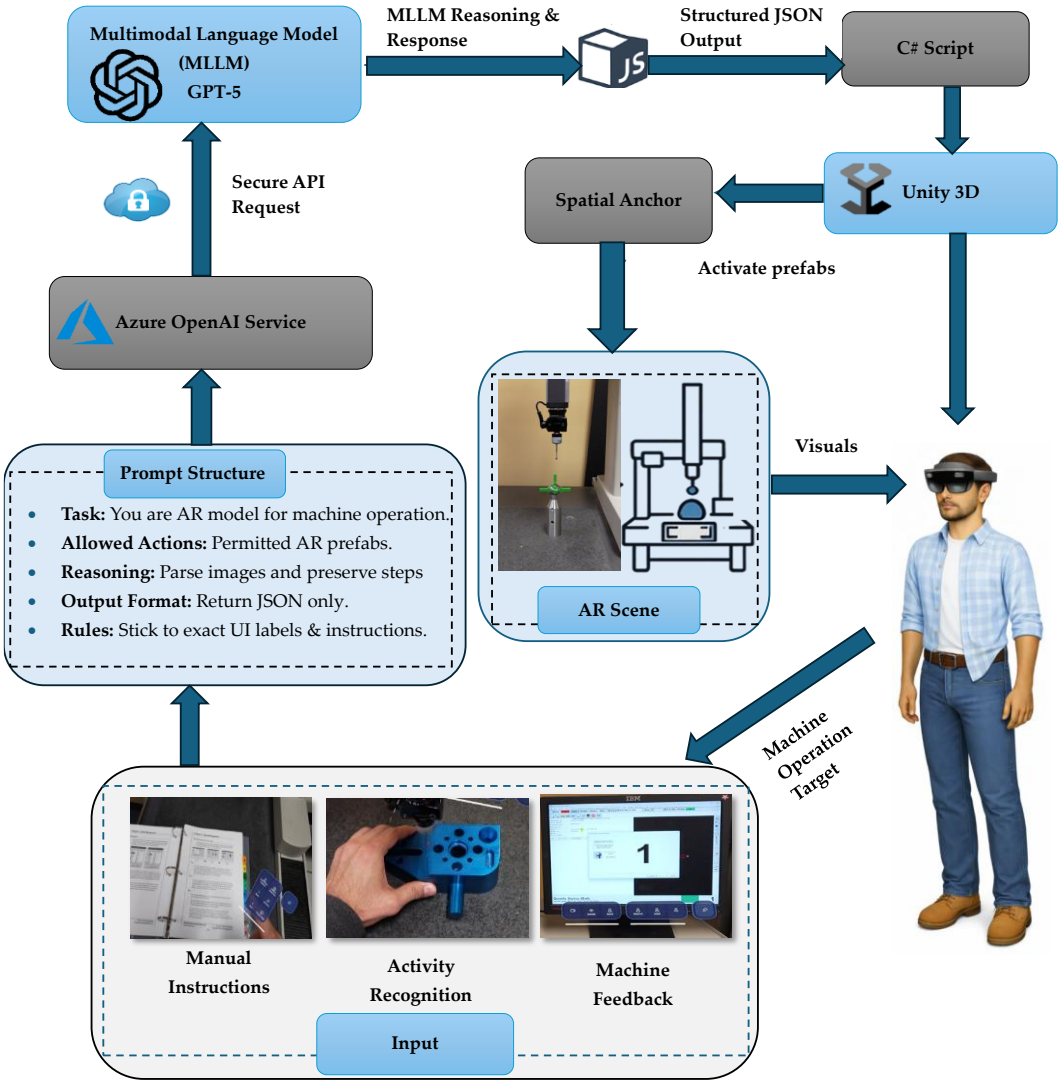


Figure 1. System architecture.

Communication between the HoloLens application and Azure OpenAI is handled through a C# API implemented in Microsoft Visual Studio 2022, which sends multimodal requests and receives structured responses, which are then parsed and rendered as AR guidance in the Unity application. At the start of training, the user captures an image using the HoloLens camera [41] to recognize the current activity and sends an input to the ChatGPT-5 model via a secure API request [6,8,24,25,28]. Once the input is received by the ChatGPT- 5 model, it interprets the image using prompt design and performs multimodal reasoning to recognize the nature of the task. Upon receiving a response from the GPT online server, the output is returned in structured JavaScript Object Notation (JSON) format. The JSON is parsed in the Unity application using C#, which allows the AR application to activate the AR interface. If the input is purely textual, such as a manual instruction, the structured prompt with the extracted content asks GPT to convert it into clear, stepwise instructions. Similarly, if the input recognizes the user activity and machine's HMI feedback, ChatGPT-5 interprets the image and triggers the corresponding anchor. Finally, the output is rendered in the user's field of view using the spatial anchor method. This system pipeline continuously infers both the machine state and operator activity from a live scene. Whenever the user encounters a new on-screen instruction, ChatGPT -5 recaptures and reprocesses the frame. This process ensures that visual cues and actions are synchronized during training. Moreover, for real-time interaction and user engagement, the training system incorporates the Microsoft HoloLens 2 AR headset. This device facilitates intuitive user

 doi:10.20944/preprints202601.1913.v1

interactions and allows trainees to engage with virtual instructional content and manipulate digital elements in an AR environment.

4. System Implementation

4.1. AR-MLLM Workflow

In the proposed system, a multimodal interference pipeline is implemented for complex machine AR-based training. This includes local heuristic validation to filter the low-quality captures, followed by semantic validation to receive accurate output from the MLLM as illustrated in Figure 2. This dual approach ensures only clear visual data is processed, which results in deterministic spatial anchor activation.

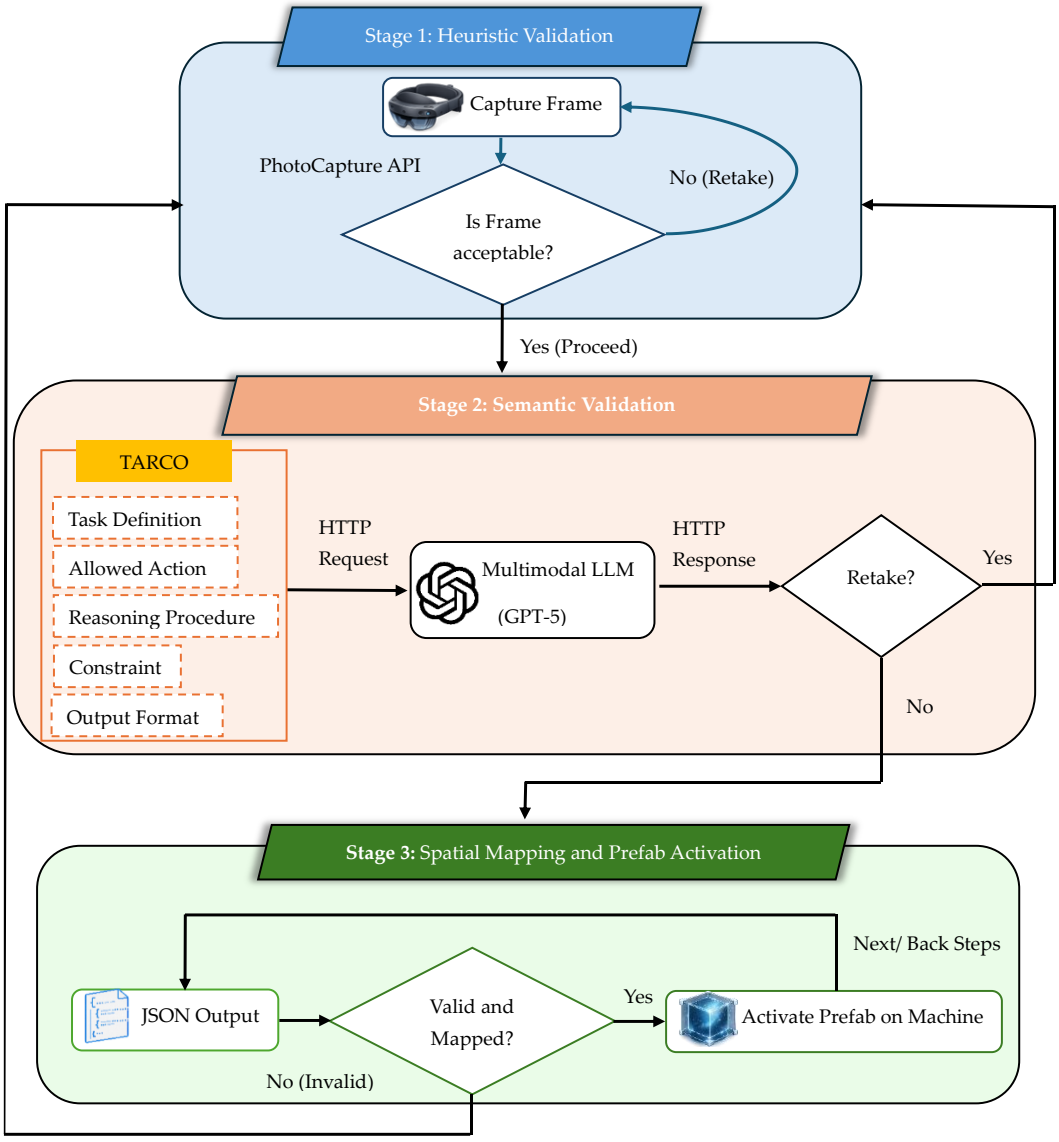


Figure 2. AR-MLLM workflow.

4.1.1. Stage 1: Heuristic validation

Industrial lighting variations and hardware-level luminance constraints are identified as major challenges that can adversely impact the reliability of high-precision tasks in AR environments [42]. To mitigate this issue, the AR–MLLM workflow initiates when a user captures a high-definition frame through HoloLens 2 PhotoCapture API in the AR environment. The system immediately executes the

 doi:10.20944/preprints202601.1913.v1

heuristic validation and calculates the mean luminance (L_{mean}) of the frame by averaging the weighted RGB components of all n pixels as defined in Equation (1),

$$L_{mean} = \frac{1}{n}\sum_{i=1}^{n}(0.2126R_i + 0.7152G_i + 0.07226B_i) \quad (1)$$

where Ri, Gi and Bi represent the red, green and blue colors for the ith pixel. The coefficients for these components are derived from the ITU-R BT.709-6 standard to calculate the relative luminance, ensuring the visual data is suitable for semantic interpretation [43]. The system accepts the frame only if it satisfies the heuristic quality criteria.

$$L_{min} < L_{mean} < L_{max} \quad (2)$$

Frames falling outside this threshold range indicate poor lighting or exposure and are rejected locally to avoid wasting API tokens and reduce LLM costs. A local recursive retake loop then triggers the system to capture a new frame before proceeding.

4.1.2. Stage 2: Semantic Validation

Once the frame is confirmed as valid during the local validation stage, it is forwarded to the MLLM via an HTTP request. To achieve an accurate interpretation of the physical environment, we employ two prompting strategies: Technical Instruction Extraction, which transforms manual text into the structured command output and User Activity and Machine Feedback Recognition, which interprets the user action and machine display feedback.

Moreover, to reduce the model search's space and ambiguity in output, the visual input is processed through a contextual prompt structure consisting of the TARCO components.

$$\Phi = (T, A, R, C, O), \quad (3)$$

where T is the task definition, A represents allowed actions, R is the reasoning procedure, C denotes constraints and rules, and O is the output format. The final output(z) is defined as a function of visual input (I) and prompt structure(Φ).

$$Z = f_{MLLM}(I, \Phi), \quad (4)$$

If the model perceives the visual input but lacks sufficient feature clarity for a high-confidence inference, it issues a RETAKE command to re-initialize the cycle until a deterministic state is reached.

4.1.3. Stage 3: Spatial Mapping and Prefab Activation

The final stage performs a semantic-to-spatial mapping. The AR system parses the MLLM output and attempts to match each label with its corresponding prefab on the machine. When the mapping is successful, the corresponding prefab is activated in the AR scene, guiding the user to the next step. The user can then move forward or revisit previous steps, creating a continuous, interactive loop throughout the training process. If the output does not conform to the expected schema, or if the user wants to proceed to the next operation, they trigger a correction loop and recapture an image to ensure the response is valid either for execution or for the next operation.

Figure 3 shows when the system detects that a captured frame is blurred or low-quality during operation, it automatically re-captures the image and repeats this process until a clear image is obtained. After the Unity application confirms that the image meets the required quality, it sends the image to ChatGPT along with a structured prompt for visual interpretation. This quality control step prevents the AR-MLLM from generating responses based on unreliable visual input and avoids triggering an incorrect prefab in the AR scene.

 doi:10.20944/preprints202601.1913.v1

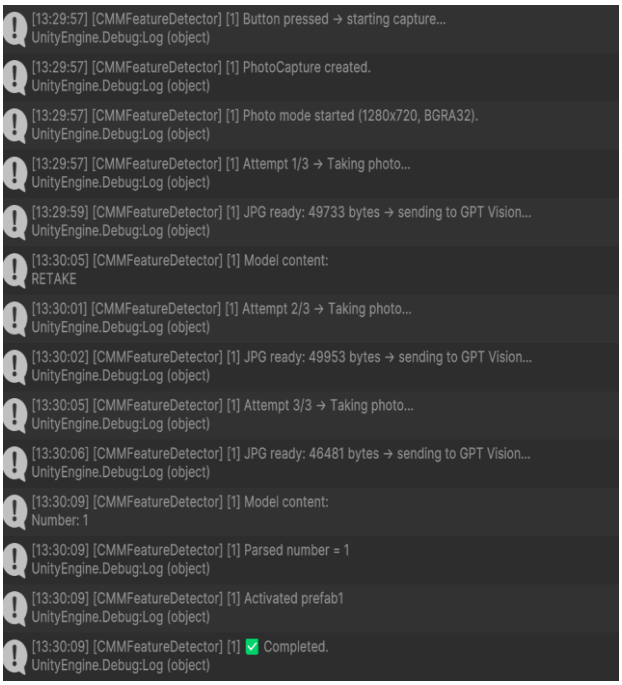


Figure 3. Frame capturing to prefab activation using MLLM in Unity.

4.2. TARCO Prompt Structure

MLLMs are primarily trained on general-purpose datasets and lack the domain-specific contextual knowledge required to accurately interpret complex machine operations. In addition, technical instructions and machine feedback often contain symbolic characters, logical dependencies, and repetitive terminology, making it difficult for models to generate appropriate responses when this information is processed directly.

Therefore, to ensure that the model output is actionable and unambiguous, the TARCO prompt structure is employed as a secondary input instruction to guide the model and enable it to produce outputs that are clear, structured, and directly usable within the AR training workflow [44]. This framework consists of following five core components.

- Task Definition (T): This specifies the role of the MLLM model and instructs it to extract the required information. It defines the model's objective so that all its responses remain aligned with the specific machine.
- Allowed Action (A): This part restricts the model to a predefined set of allowable actions and provides a strict JSON schema that must be followed, preventing models from producing their own actions that do not exist in the AR system.
- Reasoning Procedure (R): It guides the model to follow the execution order as defined in the technical instruction and ignore any shortcut steps from the technical instruction.
- Constraints and Rules (C): These rules instruct the model to use only those commands that match the technical instructions and prevent others that do not exist. It also instructs the model not to rephrase or invent labels on its own.
- Output Format (O): This part guides the model to convert the output to the required format without additional explanation. This strict format ensures that the output is executed reliably.

Table 1 presents the prompt structures used in this study for CMM operation. These prompts were developed and refined through an iterative testing process across multiple machine operation scenarios. They enable the system to convert lengthy and complex instructions into executable actions that can be processed in real time, while also supporting the recognition of user activities. Furthermore, the prompts are adaptable to various types of input, improving the model's ability to generate accurate, contextually relevant outputs.

 doi:10.20944/preprints202601.1913.v1

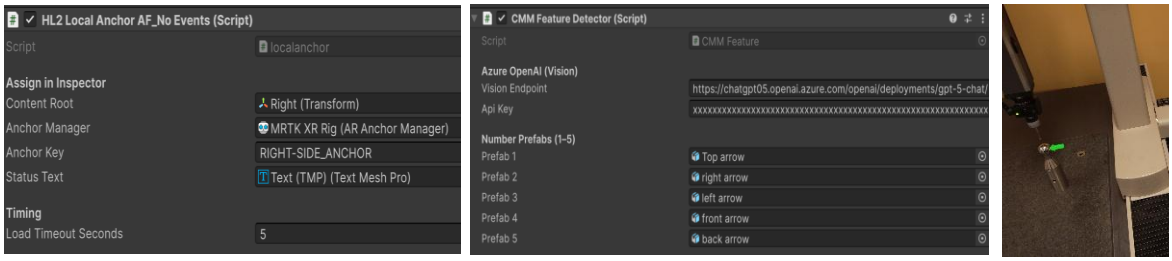
Table 1. Prompt structure for understanding technical instructions and user activity recognition.

Components	Technical Instruction Prompt	User Activity Recognition and Machine Feedback Prompt
Task Definition	1. You are an AR command extractor for a Coordinate Measuring Machine (CMM). The machine includes components such as the probe head, stylus, workpiece, and measurement features (i.e., circles, cylinders, spheres). 2. Your task is to convert the technical instructions into a single JSON object and ignore all keyboard shortcuts in the manual instructions.	1. You are an AR model for a Coordinate Measuring Machine (CMM) operation. 2. Your task is to detect machine feedback displayed in the image. 3. Identify the type of feature being measured by the CMM (circle, cylinder, or sphere) and calculate its diameter and/or length in inches.
Allowed Action	1. Translate the step-by-step instructions into AR UI actions using ONLY manual/touch interactions. 2. Follow the strict JSON schema: json { “commands”: [{ “action”: “open_menu”, “menu_path”: [“<Top>“, “<Sub>“] }, { “action”: “click”, “target”: “<ButtonLabel>“ }, { “action”: “select”, “target”: “<OptionLabel>“, “group”: “<ControlNameOptional>“ }] }. 3. For menu notation in instruction such as “Menu: A⇒B”, follow this schema: json{“action”:”open_menu”, “menu_path”: [“A”,”B”] }.	1. When the CMM machine displays numbered instructions for operation, return only: Number: <digit>. 2. When the user sees the workpiece, follow the required output schema: text Diameter: <value or N/A> Length: <value or N/A> Conclusion: <Circle>.
Reasoning Procedure	1. Preserve the exact step sequence as written in the technical manual. 2. Map textual instructions directly to the pre-defined AR UI actions.	1. Examine stylus–surface contact, curvature, edges, and feature geometry. 2. Infer the feature type from the visible contact pattern and measurement context.
Constraints and Rules	1. Use only exact UI labels from the technical instructions (i.e., “Qualify”, “Stylus Manager”, “Ball”, “Measure”, “Circle”, “Cylinder”, “OK”, “ID”). 2. Do not invent or rephrase UI labels. Labels must match the real interface exactly. 3. Never produce actions that do not exist in the instruction set.	1. If the diameter cannot be determined, return N/A for diameter. 2. Ensure the chosen feature type matches visible CMM contact points and shape cues.
Output Format	1. The output must contain only JSON objects, with no explanations, notes, or additional text.	1. Output must contain only the required format: Number: <digit>) or 2. Diameter/Length/feature (if analyzing a feature dimension). 3. No additional explanations.

 doi:10.20944/preprints202601.1913.v1

4.3. Spatial Anchoring

To maintain the alignment of virtual prefabs on a physical machine, a local spatial anchor is used. An anchor key is assigned to each virtual prefab, as shown in Figure 4(a), allowing the AR anchor manager to save and reload anchors and maintain the content fixed at the same position in the real world. After spatially anchoring the prefabs, a vision-to-action detector component is applied, as shown in Figure 4(b), which sends the captured images to ChatGPT-5 via the Azure OpenAI endpoint. Based on the model output, the system determines which prefab should be triggered and activated in the AR scene, as shown in Figure 4(c). Together, these steps facilitate seamless synchronization between the virtual and physical machine states, which is essential for delivering a realistic training experience.

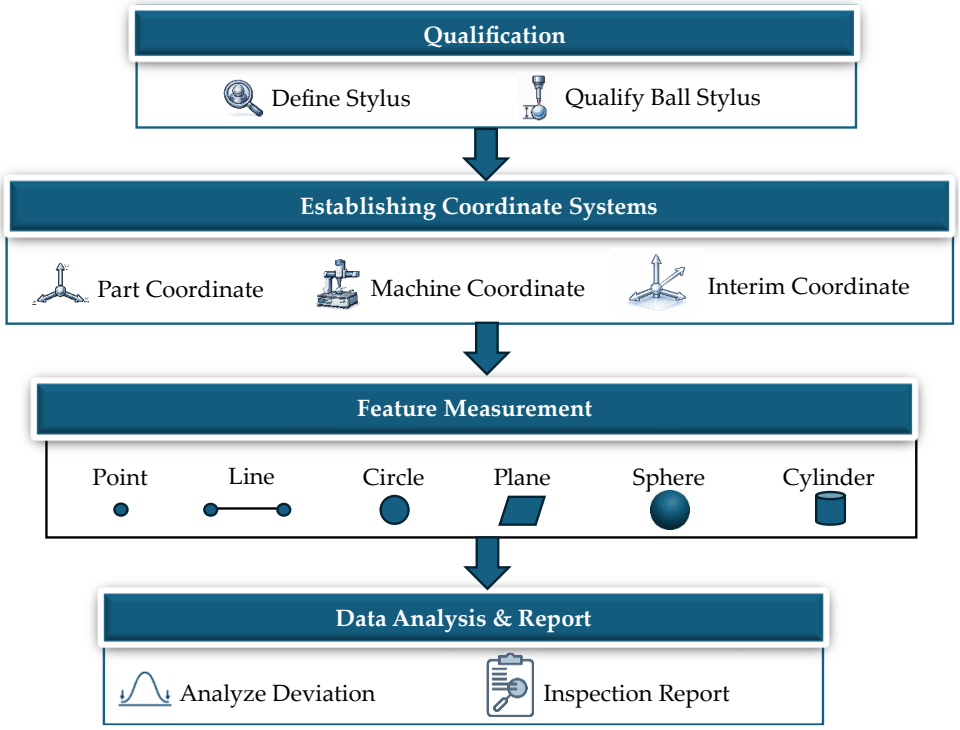


(**a**) (**b**) (**c**)

Figure 4. Anchor implementation: (**a**) Spatial anchor setup in Unity; (**b**) Vision to Action component; (**c**) Prefab activation in AR environment.

5. AR-MLLM CMM Training

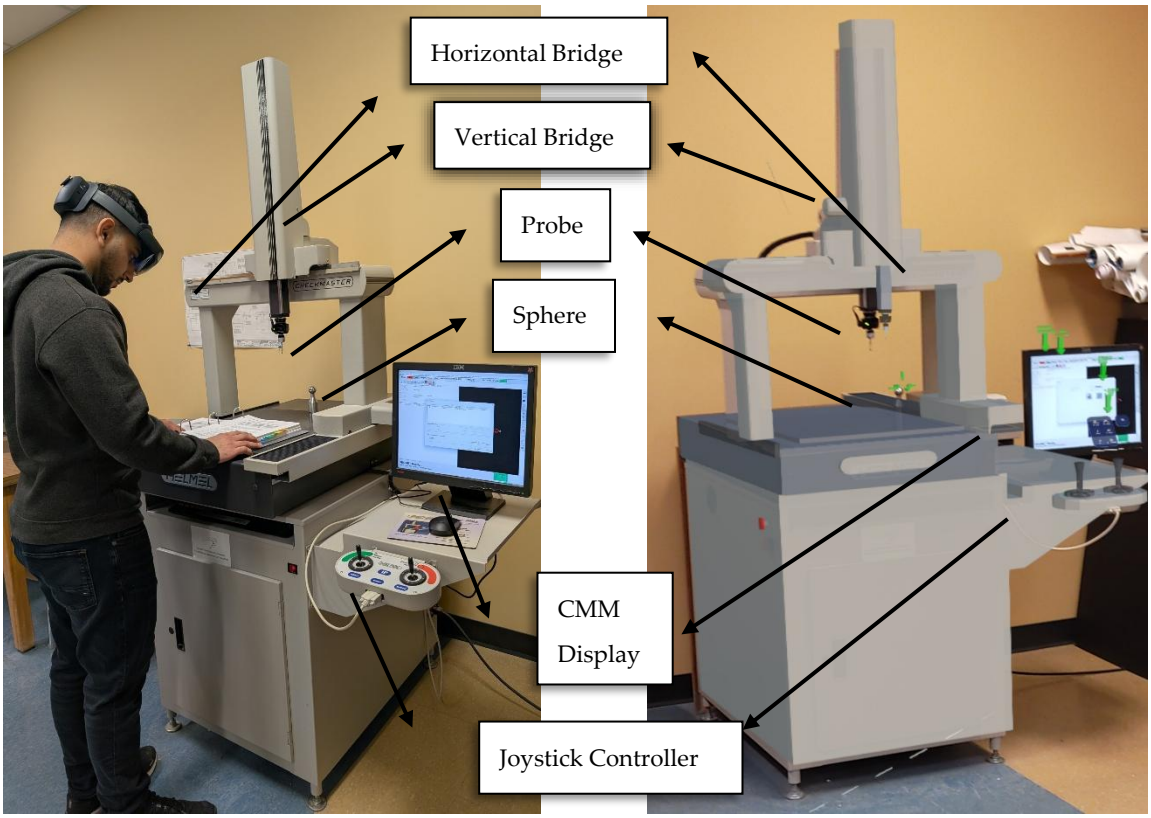
An AR-MLLM-based training system for Coordinate Measuring Machine (CMM) operations is developed for a case study of our proposed method [1]. Training operators to master CMM workflows remains a critical challenge in industries that require zero-defect manufacturing [43]. CMM operation requires precise configuration and calibration to guarantee accurate measurements, which is essential for ensuring that the final product meets the required specifications [45]. As shown in Figure 5, the workflow includes stylus qualification, establishment of the part coordinate system, feature measurement, and data analysis, which presents challenges for novice operators [1].



 doi:10.20944/preprints202601.1913.v1

Figure 5. CMM operation workflow.

To provide a realistic simulation of CMM training and an accurate representation of real-world scale and spatial relationships [46], a virtual CMM is precisely designed and deployed on a real machine, as shown in Figure 6. This includes probe, sphere, vertical bridge, horizontal bridge and joystick controller. For the stylus calibration task, virtual prompts are anchored on sphere and CMM machine display. These prompts are activated whenever users wear AR HMD and use MLLM model to provide guidance.



(**a**) (**b**)

Figure 6. Physical and virtual CMM machine: (**a**) User interacting with CMM machine using AR device; (**b**) AR prompts activated on CMM during operation.

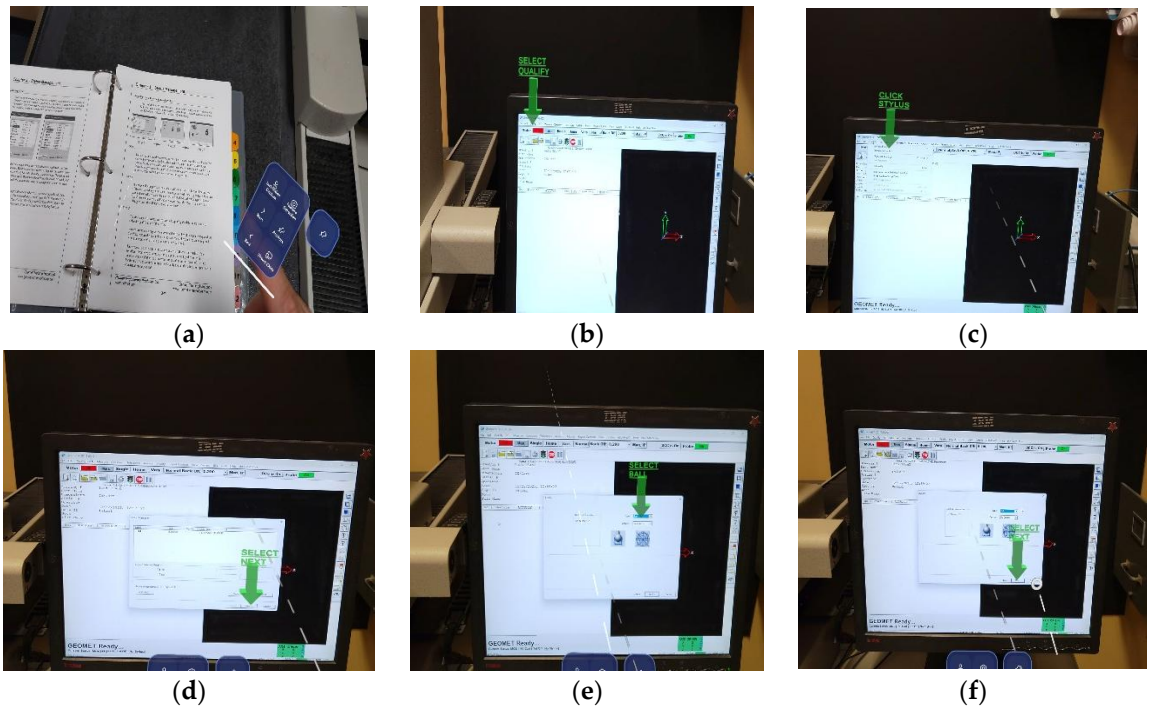
5.1. Manual Instruction and Machine Feedback

AR-MLLM operation training begins when users capture technical instructions from machine manuals using an MRTK hand interaction menu button. These virtual buttons provide the input for taking pictures and command navigation, allowing users to confirm a step or advance to the next one without using the machine's physical controls. The AR-MLLM model then performs visual–textual reasoning to interpret the captured instructions and returns the step-by-step prompts to the AR device. Figure *7* shows a stylus calibration task in which virtual arrows with labels are activated on the physical machine display to assist the user during the operation. The same workflow can be activated whenever users capture instructions for feature measurements, allowing the model to interpret the content and immediately render the corresponding guidance.

In complex machine operations, the HMI serves as the primary source of feedback on the machine's current state. Executing steps solely based on feedback leads to incorrect actions, as these displays do not provide visual guidance directly on the machine. This increases user workload and task completion time. To reduce these risks, the proposed AR-MLLM system also allows users to monitor machine feedback in real time and forward it to a GPT-5-based reasoning model to interpret the current state and render AR guidance to the machine for operator assistance. As depicted in

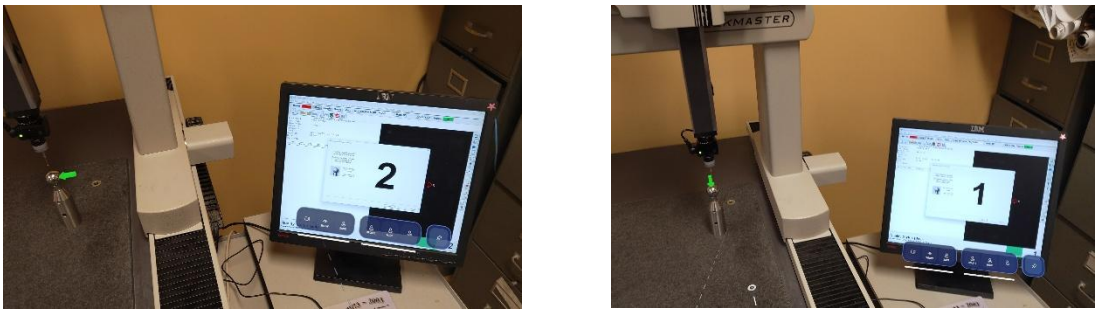
 doi:10.20944/preprints202601.1913.v1

Figure 8, the sphere prompts during stylus qualification are dynamically updated in response to CMM display feedback, providing step-by-step guidance and keeping the user aligned with the required actions.



(**a**) (**b**) (**c**) (**d**) (**e**) (**f**)

Figure 7. AR prompts activation on CMM display using AR-MLLM model: (**a**) Manual instructions for stylus qualification with MRTK menu buttons; (**b**) AR prompt for initiating qualification; (**c**) AR prompt for selecting stylus manager; (**d**) AR prompt for advancing to the next step; (**e**) AR prompt for selecting ball for stylus; (**f**) AR prompt for advancing to the final step.



(**a**) (**b**)

Figure 8. AR prefab activation based on machine feedback: (**a**) 2 points required: Right side anchored prompt activated; (**b**) 1 point required: Top-anchored prompt activated.

5.2. Activity Recognition

MLLMs can interpret physical space by processing visual inputs, but they still lack the ability to reliably recognize user activity as quickly as humans can understand task context. Thus, to help the AR-MLLM model accurately recognize user intent and reduce ambiguity in inferring the task, a user activity recognition prompt is used, as discussed in Section 4, to support human-like task interpretation during operation. For example, during feature measurement activity, the prompt specifies the operating context is "feature measurement on CMM machine", the entities of interest "stylus tip and target feature in contact", and the required inference "identify the feature and estimate its dimension".

Figure 9 shows the response generated by GPT-5 in an AR environment during the CMM feature measurement task. The AR-MLLM model identifies user activity by observing the stylus tip's

 doi:10.20944/preprints202601.1913.v1

position and computing its approximate diameter in real time, without prior measurement data. As shown in Figure 9, the model identifies the targets as a circle, a sphere, and a cylinder, and computes their approximate diameters in real time without prior measurement data.

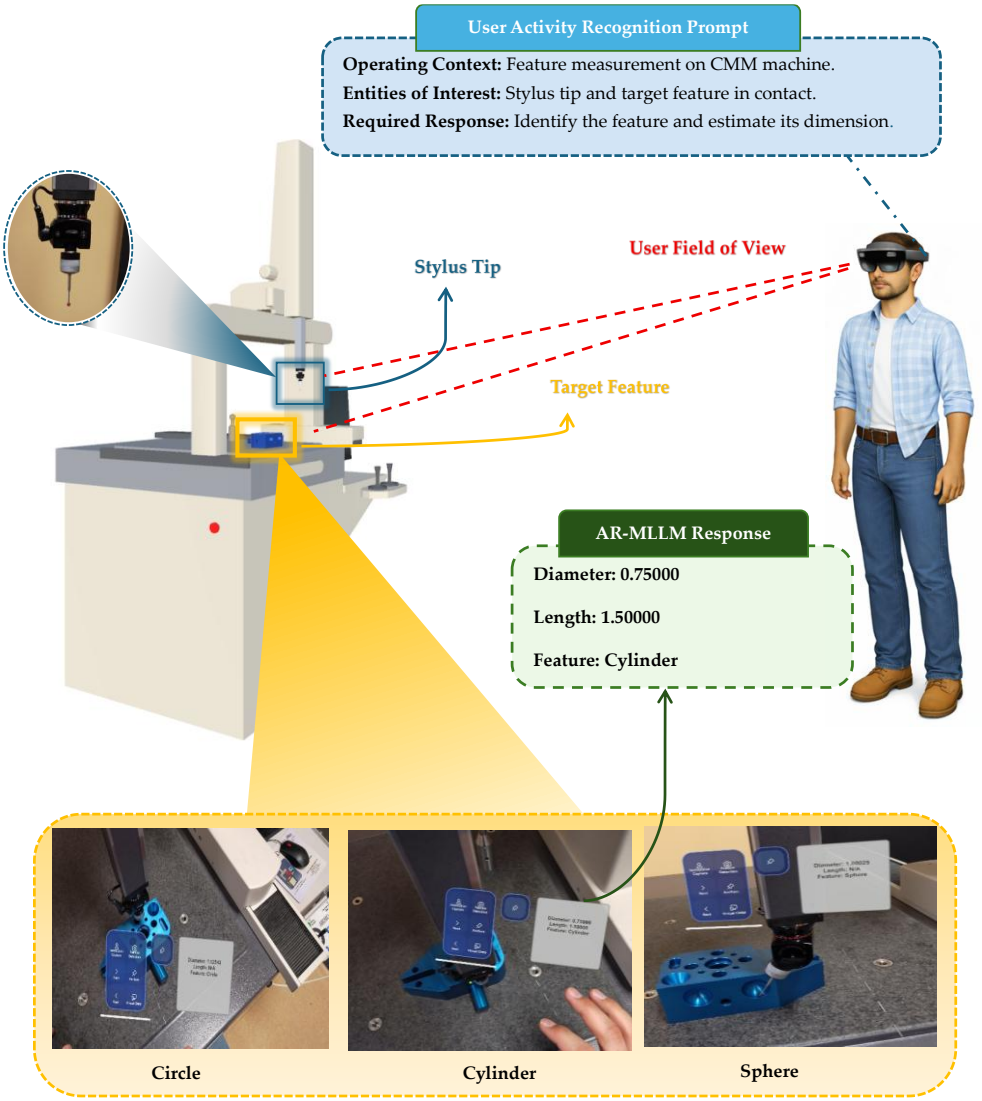


Figure 9. MLLM recognizes user activity and approximate diameter during feature measurement.

6. Evaluation

A user study was conducted to evaluate the usability and perceived workload of AR-MLLM training compared to manual-based training. In addition to this, the Bland-Altman analysis was used to compare the AR-MLLM feature measurements with CMM measurements [47]. Two well-established assessment measures are utilized: System Usability Scale (SUS) and NASA Task Load Index (NASA-TLX) [19,48]. The SUS consists of 10 questions that yield scores ranging from 0 to 100, providing an overall usability score. Each question measures a different aspect of usability, such as ease of use, confidence in using the system, complexity, and consistency. Participants rated each item on a 5-point Likert scale (1 = Strongly Disagree to 5 = Strongly Agree). The SUS score is computed using Equation (5): first transform the item responses, sum the transformed scores, and then multiply the total by 2.5 to convert it to a 0–100 scale. Higher scores indicate better usability, whereas lower scores indicate poorer usability.

$$\text{SUS Score} = (\sum \text{transformed scores}) \times 2, \quad (5)$$

 doi:10.20944/preprints202601.1913.v1

The NASA-TLX was used to assess the workload of users during CMM operation tasks, and the NASA Task Load Index (NASA-TLX) was used. It consists of six dimensions (mental demand, physical demand, temporal demand, performance, effort, and frustration). The NASA-TLX score was calculated using Equation (6) by averaging ratings across the six dimensions, where higher scores indicate a higher perceived workload, and lower scores indicate a lower perceived workload during the operation.

$$\text{NASA} - \text{TLX Score} = \left(\frac{\sum(\text{Rating of all dimensions}}{6}\right), \quad (6)$$

To evaluate the efficiency of the AR-MLLM model, the task completion time during the qualification process was measured. A shorter completion time indicates a more efficient system. Moreover, to verify the accuracy of the proposed model, task recognition and feature measurement of the workpiece during CMM operation were validated against CMM measurements using the Bland–Altman analysis. These results were then compared with those from the paper-based operation using the Wilcoxon rank-sum test to assess the significance of the system. A total of 15 participants were recruited from undergraduate and graduate students majoring in Mechanical Engineering, aged 20 to 32 years. The participants were divided into two groups to complete tasks using the proposed AR-MLLM and paper-based training methods. In the former training method, participants were first given pre-training to use an AR device (HoloLens 2). This included interactions with holograms in the AR environment and menu buttons using the built-in Microsoft app to ensure that all participants were aware of how to use the device before executing actual tasks and obtaining accurate results. AR-MLLM training starts when a user clicks the training start button and ends when probe calibration is successfully completed. Similarly, for manual-based operations, the training time was counted from reading the instructions to completing the probe calibration. The time-to-completion results and recognition accuracy of the AR-MLLM system were recorded using the HoloLens 2's built-in camera.

7. Results and Discussion

The task completion time, cognitive load, and usability of both methods for each participant were evaluated using a non-parametric Wilcoxon rank test. Two hypotheses were formulated: the null hypothesis (H_0), stating that there is no significant difference between the AR-MLLM and the method without AR-MLLM training, and the alternative hypothesis (H_1), proposing that there is a difference between the two methods, with a significance level of $p = 0.05$.

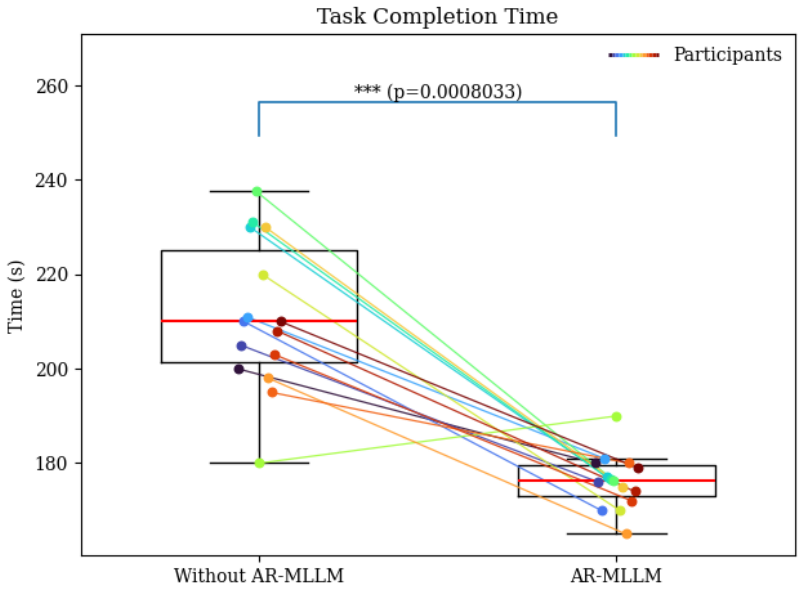
As shown in Figure 10(a), the results indicate that the task completion time for participants trained with the AR-MLLM method was lower than that for participants trained with the proposed method. The average time for AR-MLLM was 176.12 s, compared to 211.24 s for those who trained without the AR-MLLM method, with a difference of $p = 8.0x10^{-4}$, indicating a statistically significant reduction in task duration (Table 2).

Figure 10(b) shows the NASA-TLX analysis of workload for the AR-MLLM application. The survey further reveals that the AR-MLLM training conditions produced significantly lower workload scores across all subscales, with a significant difference between the two methods ($p = 6.5x10^{-4}$) as shown in Figure 10(b). Participants also rated the AR-MLLM system significantly higher in terms of usability, with an average score of 88.33, as shown in Table 2. However, some of the participants provided neutral responses but still results remained statistically significant ($p = 1.42x10^{-3}$).

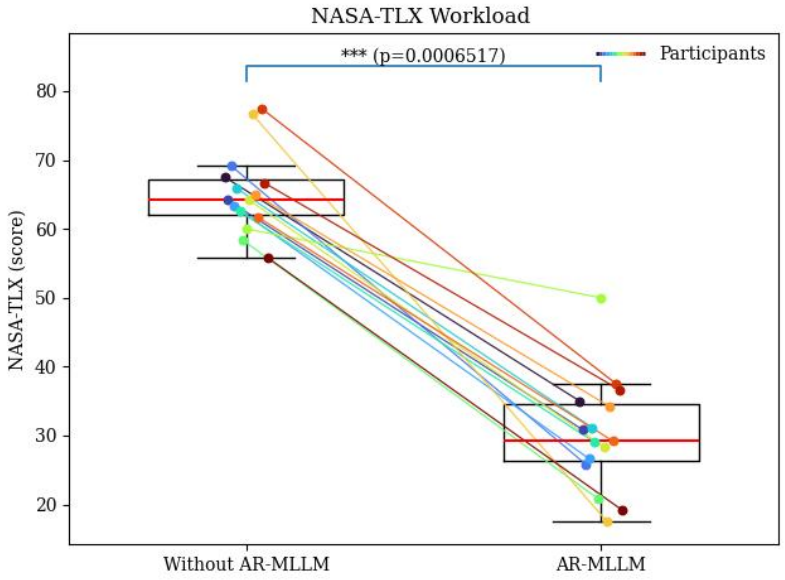
Table 2. Average values of AR-MLLM and without the AR-MLLM method.

Method	Task Execution time (s)	Usability (SUS) (score)	Task load (score)
AR-MLLM	176.12	80.33	30.12
Without AR-MLLM	211.24	24.33	65.22

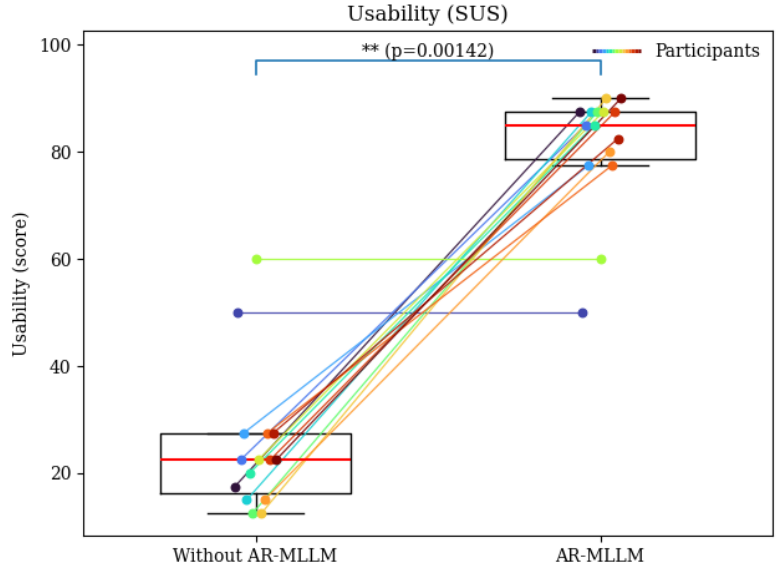
 doi:10.20944/preprints202601.1913.v1



(a)



(b)



(c)

 doi:10.20944/preprints202601.1913.v1

Figure 10. Survey results: (**a**) Task completion time; (**b**) NASA-TLX workload; (**c**) Usability (SUS).

To further validate the model's accuracy and performance, a Bland-Altman analysis was performed to compare AR-MLLM measurements with CMM measurements. It plots the difference between the two methods against their mean for each trial, including the mean bias and 95% limits of agreement. Figure 11 demonstrates the differences for AR-MLLM and CMM measurements mostly fell within the limits of the agreement for all features, showing a close agreement between the two methods across trials with LOA $-2.12\,x\,10^{-3}$ to $1.22\,x\,10^{-3}$ in for circle, LOA $9.5\,x\,10^{-4}$to $5.97\,x\,10^{-3}$ in for cylinder and LOA$-5.55\,x\,10^{-4}$ to $1.36\,x\,10^{-3}$in for sphere.

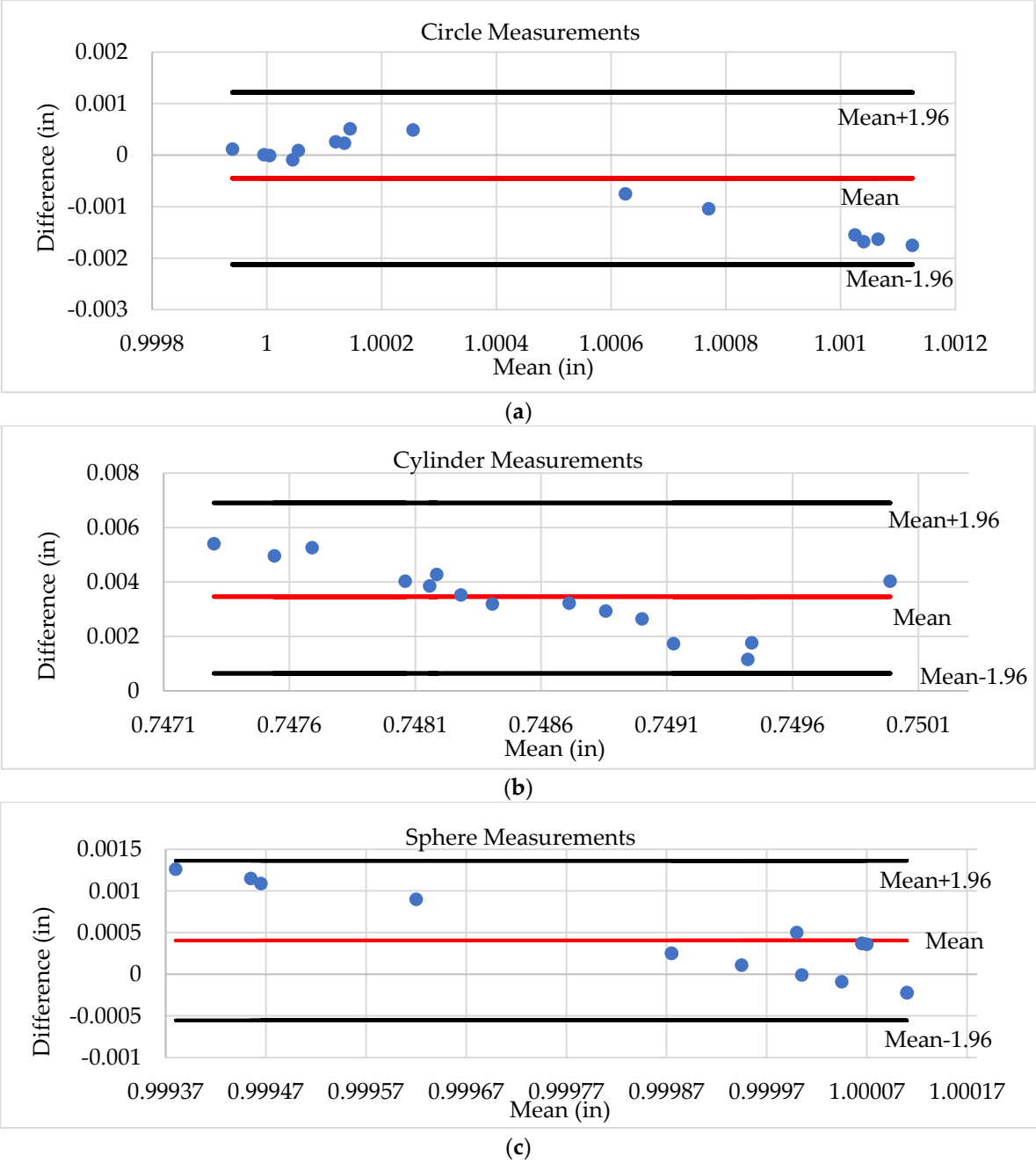


Figure 11. Bland–Altman analysis of comparing feature measurements of AR-MLLM method and CMM: (**a**) Circle diameter; (**b**) Cylinder diameter; (**c**) Sphere diameter.

As depicted in Figure 12 (a, b), the AR-MLLM model slightly underestimates the circle diameter, with a mean difference of $-4.5x10^{-4} \pm 8.5x10^{-4}$in. In the cylinder measurements, the AR-MLLM method showed a positive mean difference of $3.46.12\,x\,10^{-3}$in with $\pm\,1.28.12\,x\,10^{-3}$in standard

 doi:10.20944/preprints202601.1913.v1

deviation. This suggests that the AR-MLLM measured the cylinder diameter approximately 0.0035 in larger than the CMM on average. Similarly, sphere measurements followed the same trend, with a small average overestimation of the mean difference with $4.0\ x\ 10^{-4} \pm 4.9\ x\ 10^{-4}$ as shown in Table 3.

Overall, all these results demonstrate good agreement between the two methods and show that errors remain small and bound within the limits of agreement, indicating consistent performance over the observed measurement range. Figure 12 (c, d) further highlights the measurement accuracy of the AR-MLLM model and shows that the model achieves accuracies of 99.96%, 99.53%, and 99.95% for circles, cylinders, and spheres with corresponding MAPE values of 6.79%, 46.36% and 4.77%. These results confirm the practical capability of the proposed AR-MLLM solution in complex precision measurement operations.

Table 3. Feature measurement results.

Feature Type	Standard Deviation (SD)	Mean Bias Difference	Mean Absolute Percentage Error (MAPE)	Accuracy (%)
Circle	0.00085	-0.00045	6.79698	99.93203
Cylinder	0.00128	0.00346	46.36545	99.53635
Sphere	0.00049	0.00040	4.77005	99.95230

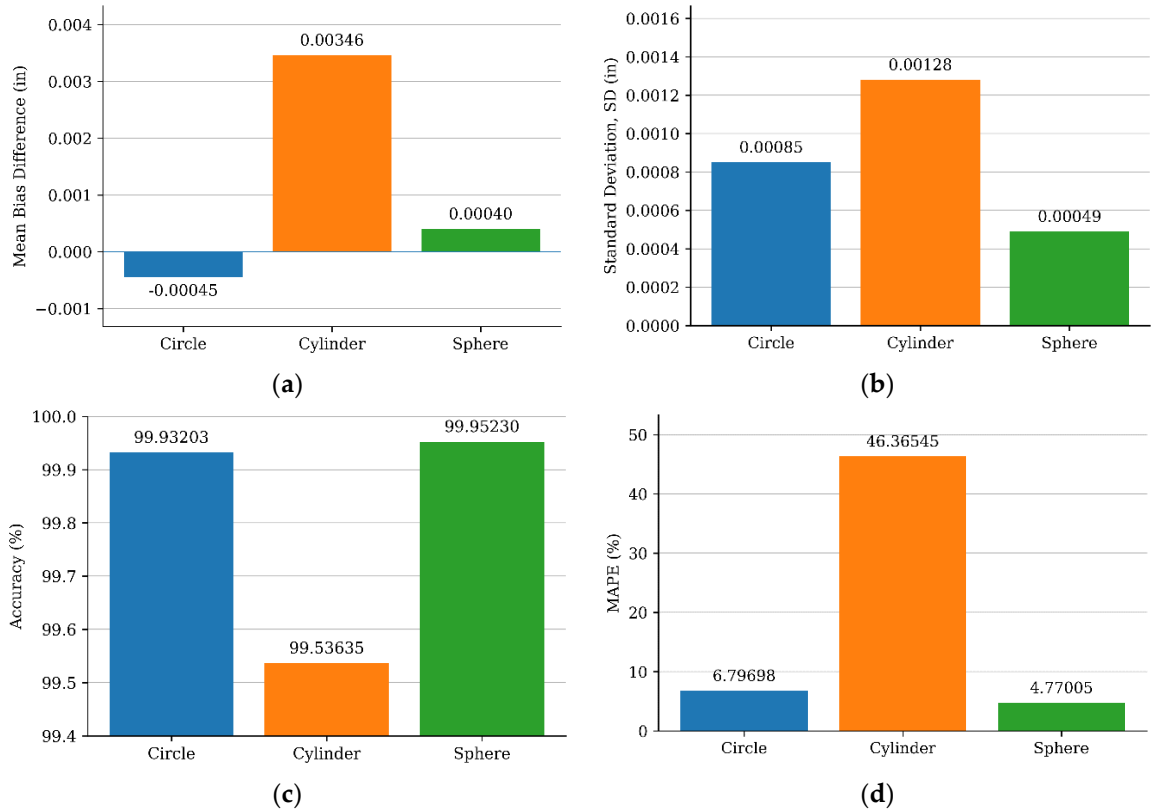


Figure 12. Feature measurement analysis: (**a**) Standard deviation; (**b**) Mean Bias Difference; (**c**) Accuracy; (**d**) Mean absolute percentage error (MAPE).

8. Conclusions

This research developed an advanced AR-MLLM-based training system for complex machine operations, where tasks are critical and require constant supervision. The use of ChatGPT-5 and prompt engineering in this research provides greater expertise and enhances operational efficiency by interpreting real-time machine feedback and recognizing user activity. Unlike traditional AR training systems that rely on fixed, pre-scripted instructions, our system provides robust training

 doi:10.20944/preprints202601.1913.v1

methods by automatically updating the steps in the AR environment and directly superimposing digital overlays on the machine.

This AR-MLLM training system allows the user to capture live visuals of the working environment and send them to ChatGPT-5 for processing and generating the output. These outputs are then returned to the AR application and provide guidance directly on the machine. We validated the proposed system using Coordinate Measuring Machine (CMM) operations. During the operation, the user captures images from the machine manual and the CMM display. These images are sent to ChatGPT-5 using carefully designed prompts, enabling the model to understand the current step and provide accurate, context-aware guidance through digital overlays on the CMM machine.

We also evaluated the system by comparing the CMM task completion time with and without AR-MLLM training. The results show that participants using the proposed method completed the operation faster, indicating improved efficiency. To further assess accuracy, we estimated the workpiece feature dimensions using the AR-MLLM system and compared them with CMM-measured dimensions. This comparison demonstrates that the system can support users by recognizing their actions and providing feature measurements. Finally, we compared the user experience with conventional paper-based training and found that the participants reported higher satisfaction and lower workload when using our approach. Together, these results show that AR-MLLM training can be effective for complex operations that normally require constant supervision, while also improving productivity, reducing time and resource use, and increasing worker satisfaction in the manufacturing industry.

In future work, the AR-MLLM model will be applied to other domains that require expert-level skills. It will also be tested with a broader user base and extended to include voice interaction in addition to text and visual inputs. Moreover, the machine feedback could be directly connected to the MLLM and trained within the system, eliminating the need to manually capture instructions and enabling the AR system to trigger digital overlays automatically.

Author Contributions: Conceptualization, W.A. and Q.P.; methodology, W.A.; software, W.A.; validation, Q.P. and W.A.; formal analysis, W.A.; investigation, W.A.; resources, Q.P.; data curation, Q.P.; writing—original draft preparation, W.A.; writing—review and editing, Q.P.; visualization, W.A.; supervision, Q.P.; project administration, Q.P.; funding acquisition, Q.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by Discovery Grants from the Natural Sciences and Engineering Research Council (NSERC) of Canada, the Graduate Enhancement of Tri-Council Stipends (GETS) program from the University of Manitoba, and the University of Manitoba Graduate Fellowships (UMGF).

Data Availability Statement: The experimental data are provided in Figures 10 and 11 of this paper.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

AR	Augmented Reality
CMM	Coordinate Measuring Machine
LLM	Large Language Model
MLLM	Multimodal Large Language Model
VLM	Vision Language Model
HMI	Human Machine Interface
JSON	JavaScript Object Notation
SD	Standard Deviation
LOA	Limits of Agreement
MAPE	Mean Absolute Percentage Error
SUS	System Usability Scale
TLX	Task Load Index

References

1. Ahmed, W.; Khan, U.; Peng, Q. Augmented Reality-Based Operation Training for Coordinate Measuring Machines Using User-Centered Interface Approach.; CAD'25: Shenzhen, China, June 23, 2025.
2. Trojanowska, J.; Kaščak, J.; Husár, J.; Knapčíková, L. Possibilities of Increasing Production Efficiency by Implementing Elements of Augmented Reality. *Bulletin of the Polish Academy of Sciences: Technical Sciences* **2022**, *70*, doi:10.24425/bpasts.2022.143831.
3. Kwon, H.J.; Lee, S. Il; Park, J.H.; Kim, C.S. Design of Augmented Reality Training Content for Railway Vehicle Maintenance Focusing on the Axle-mounted Disc Brake System. *Applied Sciences (Switzerland)* **2021**, *11*, doi:10.3390/app11199090.
4. Amouzgar, K.; Willebrand, J. A Novel XR-Based Real-Time Machine Interaction System for Industry 4.0: Usability Evaluation in a Learning Factory. *J. Manuf. Syst.* **2025**, *82*, 254–283, doi:10.1016/j.jmsy.2025.05.019.
5. Adel, A. Future of Industry 5.0 in Society: Human-Centric Solutions, Challenges and Prospective Research Areas. *Journal of Cloud Computing* **2022**, *11*, 1–15, doi:10.1186/s13677-022-00314-5.
6. You, H.; Ye, Y.; Zhou, T.; Zhu, Q.; Du, J. Robot-Enabled Construction Assembly with Automated Sequence Planning Based on ChatGPT: RoboGPT. *Buildings* **2023**, *13*, doi:10.3390/buildings13071772.
7. Xu, F.; Nguyen, T.; Du, J. Augmented Reality for Maintenance Tasks with ChatGPT for Automated Text-to-Action. *J. Constr. Eng. Manag.* **2024**, *150*, doi:10.1061/jcemd4.coeng-14142.
8. Srinidhi, S.; Lu, E.; Rowe, A. XaiR: An XR Platform That Integrates Large Language Models with the Physical World. In Proceedings of the 2024 IEEE International Symposium on Mixed and Augmented Reality (ISMAR); IEEE, October 2024; pp. 759–767.
9. Butt, J. A Strategic Roadmap for the Manufacturing Industry to Implement Industry 4.0. *Designs (Basel).* **2020**, *4*, 1–31.
10. Morales Méndez, G.; del Cerro Velázquez, F. Augmented Reality in Industry 4.0 Assistance and Training Areas: A Systematic Literature Review and Bibliometric Analysis. *Electronics (Switzerland)* **2024**, *13*, doi:10.3390/electronics13061147.
11. Malta, A.; Farinha, T.; Mendes, M. Augmented Reality in Maintenance—History and Perspectives. *J. Imaging* **2023**, *9*, doi:10.3390/jimaging9070142.
12. Van Campenhout, L.; Vancoppenolle, W.; Dewit, I. From Meaning to Expression: A Dual Approach to Coupling. *Designs (Basel).* **2023**, *7*, doi:10.3390/designs7030069.
13. Lakka, E.; Malamos, A.G.; Pavlakis, K.G.; Andrew Ware, J. Designing a Virtual Reality Platform to Facilitate Augmented Theatrical Experiences Based on Auralization. *Designs (Basel).* **2019**, *3*, 1–11, doi:10.3390/designs3030033.
14. Tobiskova, N.; Hattinger, M.; Sanderson Gull, E. Evaluating an Augmented Reality Prototype for Enhanced User Guidance in an Industrial Production Context. In Proceedings of the Advances in Transdisciplinary Engineering; IOS Press BV, April 9 2024; Vol. 52, pp. 419–430.
15. Frizziero, L.; Santi, G.M.; Donnici, G.; Leon-Cardenas, C.; Ferretti, P.; Liverani, A.; Neri, M. An Innovative Ford Sedan with Enhanced Stylistic Design Engineering (SDE) via Augmented Reality and Additive Manufacturing. *Designs (Basel).* **2021**, *5*, doi:10.3390/designs5030046.
16. Xue, Z.; Yang, J.; Chen, R.; He, Q.; Li, Q.; Mei, X. AR-Assisted Guidance for Assembly and Maintenance of Avionics Equipment. *Applied Sciences (Switzerland)* **2024**, *14*, doi:10.3390/app14031137.
17. Kang, G.H.; Kwon, H.J.; Chung, I.S.; Kim, C.S. A Study on the Development of Augmented Reality Contents for Air Compressor of Railway Vehicles. In Proceedings of the 2023 Prognostics and Health Management Conference - Paris, PHM-Paris 2023; Institute of Electrical and Electronics Engineers Inc., 2023; pp. 59–63.
18. Koteleva, N.; Valnev, V.; Frenkel, I. Investigation of the Effectiveness of an Augmented Reality and a Dynamic Simulation System Collaboration in Oil Pump Maintenance. **2021**, doi:10.3390/app.
19. Daling, L.M.; Tenbrock, M.; Isenhardt, I.; Schlittmeier, S.J. Assemble It like This! – Is AR- or VR-Based Training an Effective Alternative to Video-Based Training in Manual Assembly? *Appl. Ergon.* **2023**, *110*, doi:10.1016/j.apergo.2023.104021.
20. Yong, J.; Wei, J.; Wang, Y.; Dang, J.; Lei, X.; Lu, W. Heterogeneity in Extended Reality Influences Procedural Knowledge Gain and Operation Training. *IEEE Transactions on Learning Technologies* **2023**, *16*, 1014–1033, doi:10.1109/TLT.2023.3286612.

21. Shankhwar, K.; Smith, S. An Interactive Extended Reality-Based Tutorial System for Fundamental Manual Metal Arc Welding Training. *Virtual Real.* **2022**, *26*, 1173–1192, doi:10.1007/s10055-022-00626-6.
22. Peckham, O.; Raines, J.; Bulsink, E.; Goudswaard, M.; Gopsill, J.; Barton, D.; Nassehi, A.; Hicks, B. Artificial Intelligence in Generative Design: A Structured Review of Trends and Opportunities in Techniques and Applications. *Designs (Basel).* **2025**, *9*.
23. Zhang, C.; Chen, J.; Li, J.; Peng, Y.; Mao, Z. Large Language Models for Human–Robot Interaction: A Review. *Biomimetic Intelligence and Robotics* **2023**, *3*, doi:10.1016/j.birob.2023.100131.
24. Xu, F.; Nguyen, T.; Du, J. Augmented Reality for Maintenance Tasks with ChatGPT for Automated Text-to-Action. *J. Constr. Eng. Manag.* **2024**, *150*, doi:10.1061/jcemd4.coeng-14142.
25. Fuchter, S.K.; Schlichting, M.S.; Filho, G.G. Study on the Use of ChatGPT for Generating Code for Web-Based Augmented Reality Applications. In Proceedings of the Measurement: Sensors; Elsevier Ltd., May 1 2025; Vol. 38.
26. Zichar, M.; Papp, I. Contribution of Artificial Intelligence (AI) to Code-Based 3D Modeling Tasks. *Designs (Basel).* **2024**, *8*, doi:10.3390/designs8050104.
27. Stover, D.; Bowman, D. TAGGAR: General-Purpose Task Guidance from Natural Language in Augmented Reality Using Vision-Language Models. In Proceedings of the Proceedings - SUI 2024: ACM Symposium on Spatial User Interaction; Association for Computing Machinery, Inc, October 7 2024.
28. Ye, Y.; You, H.; Du, J. Improved Trust in Human-Robot Collaboration With ChatGPT. *IEEE Access* **2023**, *11*, 55748–55754, doi:10.1109/ACCESS.2023.3282111.
29. Fan, H.; Zhang, H.; Ma, C.; Wu, T.; Fuh, J.Y.H.; Li, B. Enhancing Metal Additive Manufacturing Training with the Advanced Vision Language Model: A Pathway to Immersive Augmented Reality Training for Non-Experts. *J. Manuf. Syst.* **2024**, *75*, 257–269, doi:10.1016/j.jmsy.2024.06.007.
30. Wang, H.; Li, Y.F. Large Language Model Empowered by Domain-Specific Knowledge Base for Industrial Equipment Operation and Maintenance. In Proceedings of the 2023 5th International Conference on System Reliability and Safety Engineering, SRSE 2023; Institute of Electrical and Electronics Engineers Inc., 2023; pp. 474–479.
31. Fan, H.; Fuh, J.; Lu, W.F.; Kumar, A.S.; Li, B. Unleashing the Potential of Large Language Models for Knowledge Augmentation: A Practical Experiment on Incremental Sheet Forming. In Proceedings of the Procedia Computer Science; Elsevier B.V., 2024; Vol. 232, pp. 1269–1278.
32. Liu, P.; Yuan, W.; Fu, J.; Jiang, Z.; Hayashi, H.; Neubig, G. Pre-Train, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing. *ACM Comput. Surv.* **2023**, *55*, 1–35, doi:10.1145/3560815.
33. Chen, B.; Yi, F.; Varro, D. Prompting or Fine-Tuning? A Comparative Study of Large Language Models for Taxonomy Construction. In Proceedings of the Proceedings - 2023 ACM/IEEE International Conference on Model Driven Engineering Languages and Systems Companion, MODELS-C 2023; Institute of Electrical and Electronics Engineers Inc., 2023; pp. 588–596.
34. Boys Smith, N.; Salingaros, N.A. AI Judging Architecture for Well-Being: Large Language Models Simulate Human Empathy and Predict Public Preference. *Designs (Basel).* **2025**, *9*, doi:10.3390/designs9050118.
35. Nan, L.; Zhao, Y.; Zou, W.; Ri, N.; Tae, J.; Zhang, E.; Cohan, A.; Radev, D. Enhancing Text-to-SQL Capabilities of Large Language Models: A Study on Prompt Design Strategies. In Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing; Association for Computational Linguistics (ACL).: Singapore, December 2023; pp. 14935–14956.
36. Zheng, H.T.; Xie, Z.; Liu, W.; Huang, D.; Wu, B.; Kim, H.G. Prompt Learning with Structured Semantic Knowledge Makes Pre-Trained Language Models Better. *Electronics (Switzerland)* **2023**, *12*, doi:10.3390/electronics12153281.
37. Jang, M.; Lukasiewicz, T. Consistency Analysis of ChatGPT. In Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing; Bouamor, H., Pino, J., Bali, K., Eds.; Association for Computational Linguistics: Singapore, December 2023; pp. 15970–15985.
38. Kojima, T.; Gu, S.S.; Reid, M.; Matsuo, Y.; Iwasawa, Y. Large Language Models Are Zero-Shot Reasoners. In Proceedings of the NIPS'22: Proceedings of the 36th International Conference on Neural Information Processing Systems; Curran Associates Inc.: New Orleans USA, January 29 2022; pp. 22199–22213.

 doi:10.20944/preprints202601.1913.v1

39. Angelopoulos, J.; Manettas, C.; Alexopoulos, K. Industrial Maintenance Optimization Based on the Integration of Large Language Models (LLM) and Augmented Reality (AR). In Proceedings of the Advances in Artificial Intelligence in Manufacturing II; Alexopoulos, K., Makris, S., Stavropoulos, P., Eds.; Springer Nature Switzerland: Cham, 2025; pp. 197–205.
40. Unity Technologies. Unity 6000.0.29f1 Available online: https://unity.com/releases/editor/whats-new/6000.0.29f1 (accessed on 27 November 2025).
41. Ungureanu, D.; Bogo, F.; Galliani, S.; Sama, P.; Duan, X.; Meekhof, C.; Stühmer, J.; Cashman, T.J.; Tekin, B.; Schönberger, J.L.; et al. HoloLens 2 Research Mode as a Tool for Computer Vision Research. **2020**, doi:10.48550/arXiv.2008.11239.
42. König, S.; Siebers, S.; Backhaus, C. Image Quality Assessment of Augmented Reality Glasses as Medical Display Devices (HoloLens 2). *Applied Sciences (Switzerland)* **2025**, *15*, doi:10.3390/app15147648.
43. International Telecommunication Union. Recommendation ITU-R BT.709-6: Parameter Values for the HDTV Standards for Production and International Programme Exchange. Available online: https://www.itu.int/rec/R-REC-BT.709-6-201506-I (accessed on 20 December 2025).
44. Yan, X.; Xiao, Y.; Jin, Y. Generative Large Language Models Explained. *IEEE Comput. Intell. Mag.* **2024**, *19*, 45–46, doi:10.1109/MCI.2024.3431454.
45. Torok, J.; Kocisko, M.; Teliskova, M.; Janak, M. Increasing of the Work Productivity of CMM Machine by Applying of Augmented Reality Technology. In Proceedings of the MATEC Web of Conferences; EDP Sciences, August 1 2016.
46. Martinez Gasca, O.; Van Dorpe, L.; Dewit, I.; Van Campenhout, L. SAR Miniatures: Physical Scale Models as Immersive Prototypes for Spatially Augmented Environments. *Designs (Basel).* **2025**, *9*, doi:10.3390/designs9010010.
47. Bland, J.M.; Altman, D.G. Agreement between Methods of Measurement with Multiple Observations per Individual. *J. Biopharm. Stat.* **2007**, *17*, 571–582, doi:10.1080/10543400701329422.
48. Ghobrial, M.; Seitier, P.; Lagarrigue, P.; Galaup, M.; Gilles, P. Effectiveness of Machining Equipment User Guides: A Comparative Study of Augmented Reality and Traditional Media. In Proceedings of the Materials Research Proceedings; Association of American Publishers, 2024; Vol. 41, pp. 2320–2328.