ArtWork

SMART AND CONNECTED WORKER

ITEA 4 – 22019

Work package 4

Context-based Worker Assistance System

Deliverable 4.3

Sensor Fusion-based Task Planning Tool

Document type	: Deliverable
Document version	: 1.0
Document Preparation Date	: 16.12.2024
Classification	: Public
Contract Start Date	: 2024-03-01
Contract End Date	: 2027-02-28

ITEA4

Final approval	**Name**	**Partner**
Review Task Level	Pedro Colarejo	LOAD
Review WP Level	Sabino Roselli	Chalmers
Review Board Level	Thomas Bär	DAIMLER

Executive Summary

This deliverable presents the results and methodological advances achieved in the development of context-aware task planning for smart and connected worker assistance systems. Building on previously established perception and sensing capabilities, the work focuses on transforming real-time observations of workers, tools, and environments into structured task understanding and adaptive planning decisions. The central contribution of this phase is the establishment of a systematic link between perception outputs and higher-level reasoning mechanisms that can support intelligent, situation-dependent system behavior in industrial settings.

Across the addressed industrial domains, the work demonstrates how heterogeneous inputs such as worker localization, action recognition, and intention prediction can be combined into coherent task representations. These representations enable the system to reason task progress, anticipate upcoming actions, and evaluate contextual constraints such as safety conditions, workspace configuration, and resource availability. Rather than relying on rigid, predefined workflows, the proposed planning approaches support flexibility and adaptation, which are essential in environments characterized by manual operations, variability in worker behavior, and frequent deviations from nominal processes.

The deliverable also highlights how context-aware task planning contributes to measurable project objectives. By enabling timely and relevant assistance actions, the developed concepts support improvements in productivity, error prevention, and system responsiveness, while maintaining alignment with safety and privacy requirements. The approaches are designed with validation in mind and are mapped to concrete industrial scenarios, providing a clear basis for assessing performance against project KPIs such as task completion efficiency, robustness under uncertainty, and user acceptance.

Importantly, this work establishes the foundation for the next task of this work package, where planning will be integrated with execution and feedback mechanisms to form a closed-loop, context-based worker assistance system. In this closed-loop setup, perceptual feedback will continuously inform planning decisions, allowing the system to adapt dynamically to changing conditions and user behavior. This progression enables end-to-end validation of intelligent assistance concepts in real industrial pilots, demonstrating tangible benefits for both operational performance and worker support.

Glossary

GPSS - Generic Photo-Based Sensor System

RITA - Robot In The Air

HMMs - Hidden Markov Models

DBNs - Dynamic Bayesian Networks

RNNs - Recurrent Neural Networks

LSTM - Long Short-Term Memory

GNNs - Graph Neural Networks

RGB-D – camera that provides both colour and depth data

LLMs - Large Language Models

VLMs - Vision-Language Models

FFSM - fuzzy finite state machine

SEFFSM - structure-evolving fuzzy finite-state machine

IPM - Inverse Perspective Mapping

List of Figures

Figure 1: Structure-Evolving Fuzzy Finite State Machine 16

Figure 2: Multi Camera Perception...17

Figure 3: Human-Robot collaboration demo in a shared workspace 19

Figure 4: dual-pane user interface 22

Figure 5: Analyzing video frames or segments where the differences occurred 23

Figure 6: Robot demonstrator with contextual trajectory stitching method 24

Content

Executive Summary 3
Glossary 4
List of Figures 4
1 Introduction 7
2 Industrial Challenges 8
2.1 Challenges in Truck industry 8
2.2 Challenges in Bus industry 9
2.3 Challenges in Textile industry 9
3 State of the Art 11
3.1 Sensor Fusion 11
3.2 Deviation Detection 12
3.2.1 Deviation Detection in Human-Centric Textile and Automotive Production Processes 12
3.2.2 Reference Procedure and Deviation Typology 13
3.2.3 Automotive Industry as a Reference Baseline 13
3.2.4 Textile Industry Specificities and Challenges 13
3.2.5 AI-Driven Multimodal Deviation Detection 14
3.2.6 Open Challenges in the Artwork Context 14
3.3 Task Re-Planning 14
4 Approach 15
4.1 Sensor Fusion based on Fuzzy Finite State Machines (FFSM) 15
4.2 Computer vision techniques 17
4.3 Deviation Detection 17
4.3.1 Reference Method: Normative Backbone with Expert Variability 18
4.3.2 Sensor Fusion Focused on Evidence of Task Completion 18
4.3.3 Online Plan–Execution Alignment 18
4.3.4 Deviation Handling and Adaptive Task Planning 18
4.3.5 Conclusion 19
4.4 Task Re-Planning 19
5 Preliminary Results 21
5.1 Sensor Fusion based on Fuzzy Finite State Machines (FFSM) 21

5.2 Tablet-Based Deviation Detection at Sewing Workstations 21

5.3 Task re-planning 24

6 Conclusions and Future Work 26

7 Bibliography.. 27

1 Introduction

In this phase we build on WP4.2, where the focus was placed on sensing and perception as the foundation for understanding industrial work environments. Through the development of methods for worker and tool detection, localization, action recognition, and three-dimensional environment modeling, WP4.2 established the capability to observe what is happening on the factory floor in real time. These perception capabilities enable the system to capture the current state of the environment, including the position and actions of human operators, the availability and use of tools, and the spatial context in which tasks are performed. While this represents a critical step toward smarter and safer industrial systems, perception alone is not sufficient to enable adaptive and intelligent behavior.

This deliverable addresses the next logical step: transforming perceptual information into context-aware task planning. The central objective of this phase is to move from observing the environment to reasoning about it, enabling systems to decide what to do next based on the current task context. Context-aware task planning aims to interpret the worker's ongoing activity, infer intentions, and anticipate upcoming steps, allowing assistance systems, robots, or digital tools to adapt their behavior accordingly. This is particularly relevant in industrial settings characterized by manual or semi-automated work, high variability, and frequent human–machine interaction, where rigid, predefined workflows are insufficient.

Building on the outputs of WP4.2, this phase leverages detection, localization, and intention prediction as inputs to higher-level planning mechanisms. These mechanisms reason task progress, spatial constraints, safety conditions, and available resources to support adaptive assistance, dynamic task sequencing, and coordinated actions between humans and machines. The goal is not to replace human decision-making, but to augment it by providing timely, situation-aware support in hybrid environments that may include robots, intelligent tools, or digital worker assistance systems.

The challenges addressed in this phase span multiple industrial domains, including truck production, bus assembly, and textile manufacturing. Across these domains, context-aware task planning must cope with uncertainty in perception, variability in worker behavior and skill levels, and strict safety and operational constraints. This deliverable therefore focuses on approaches that integrate perceptual information into robust task representations and planning strategies, laying the groundwork for adaptive, safe, and efficient industrial assistance systems in the subsequent stages of the project.

2 Industrial Challenges

The industrial challenges addressed in this work stem from the growing complexity and variability of modern production environments, where humans increasingly interact with intelligent systems and tools. Ensuring safe, efficient, and adaptive operation requires a reliable understanding of task context under real-world constraints. The following sections summarize the key challenges identified across the targeted industrial domains.

2.1 Challenges in Truck industry

The GPSS (Generic Photo-Based Sensor System) project aims to streamline material delivery and logistics by implementing an AI-powered automated pallet truck transporter system. Unlike traditional solutions that rely on expensive on-board sensors and processing units, GPSS utilizes a network of ceiling-mounted cameras to act as a "global eye." This allows the system to manage fleet navigation and safety boundaries from a bird's-eye perspective, ensuring efficient activity across large-scale warehouse and factory environments.

Simultaneously, the RITA (Robot In The Air) project targets the "kitting" process, the labor-intensive task of gathering and organizing parts for assembly. RITA introduces a gantry-mounted collaborative robot that assists operators from above.

To enable true human-robot coexistence (in GPSS) and collaboration (RITA), both projects seek to reduce physical distance between robots and workers. However, replacing steel fences safety cages with "virtual" ones introduces several technical challenges:

- Off-Board Perception: Since the transporters and robot arms lack local LiDAR or proximity sensors, safety is entirely dependent on the off-board cameras with low latency and no blind spots.
- Multi-Camera Data Fusion: In Volvo warehouses, no single camera can maintain a global view of the entire line. The system must "hand off" the tracking of workers and transporters between cameras without losing their identity or coordinates. This requires complex data fusion algorithms
- Human Pose Detection: Safety in a kitting cell requires more than just detecting a "person." The system must perform high-fidelity pose estimation (tracking joints) to ensure the robot does not collide with a worker's arm during these tasks.
- Continuous Verification & Validation (V&V): Because these systems are dynamic and AI-driven, traditional safety certifications are insufficient. There is a constant need for testing and validation frameworks to ensure the AI behaves predictably and correctly in all situations such as changing lighting conditions or changing in the environment.

2.2 Challenges in Bus industry

The bus manufacturing industry presents unique challenges in assisted assembly processes, where workers must follow complex, multi-step procedures while managing tools, materials, and dynamic workflows. The primary challenge addressed in this work package is the development of a context-aware assistance system designed to support workers in real-time by providing relevant working instructions via a monitor.

The assistance system must dynamically adapt to the worker's current task context to ensure the correct information is displayed at the right moment. This requires multi-modal sensor fusion to integrate diverse data sources, including:

- Position Tracking
    - Real-time monitoring of the worker's location, as well as the positions of tools and materials in the workspace.
    - Ensures the system understands which resources are being accessed or utilized.
- Action Recognition
    - Detection of the worker's current actions (e.g., picking up a tool, handling a material, or performing a specific task).
    - Enables the system to correlate physical activities with predefined workflow steps.
- Event-Based History Tracking
    - Maintenance of a detailed event log capturing:
        - Material/tool usage (e.g., "Material X taken", "Tool Y picked up").
        - Action initiation (e.g., "Worker started action Z").
        - Resource return events (e.g., "Tool Y placed back on the toolboard").
    - This historical data helps the system infer the worker's progress and predict the next logical step.

The most critical challenge lies in fusing these heterogeneous data streams to accurately determine the current working step the worker is performing. This involves:

- Disambiguating overlapping or ambiguous events (e.g., distinguishing between a tool being used for a primary task vs. a secondary adjustment).
- Mapping sensor data to predefined workflow models to identify whether the worker is on track, deviating, or requires guidance.
- Dynamic adaptation of displayed instructions based on real-time context, ensuring minimal cognitive load while maintaining productivity.

By resolving these challenges, the assistance system aims to reduce errors, improve efficiency, and enhance worker situational awareness in high-complexity bus assembly processes.

2.3 Challenges in Textile industry

The textile manufacturing environment presents significant challenges in ensuring that operators consistently execute tasks according to the most efficient and defect-free methods. Production lines are characterized by high variability, frequent changes in product configurations, short life cycles, and strong dependency on operator skills. In this context, the primary challenge addressed in this work package is the development of a context-aware Assistance System capable of monitoring task execution in real time and supporting operators with corrective and adaptive gui-dance.

The proposed Assistance System aims to observe and interpret how tasks are performed on the shop floor and to compare the detected execution patterns with predefined reference methods. To achieve this, the system relies on multi-modal sensor fusion, combining visual perception and contextual data sources, including:

- Operator identification and positioning, enabling the association of actions and performance metrics with a specific operator.
- Action and method recognition, through camera-based analysis of gestures and movements, allowing comparison between the detected execution and the reference method.
- Task phase identification, determining the current stage of the operation, and overall progress.
- Time and performance monitoring, measuring execution times, and detecting idle or inefficient periods.
- Contextual event logging, recording relevant execution events to build a historical trace for analysis and improvement.

Based on this contextual understanding, the Assistance System can provide real-time feedback to the operator, such as alerts, visual cues, or corrective suggestions, and can also update reference methods when better-performing execution patterns are consistently observed. By addressing these challenges, the Assistance System supports continuous improvement of production methods, reduces dependence on scarce “time and methods” specialists, and contributes to higher efficiency, consistency, and quality in textile manufacturing operations.

3 State of the Art

This section provides an overview of the existing methods used to deal with the industrial challenges listed in the previous section.

3.1 Sensor Fusion

Sensor fusion refers to the integration of heterogeneous sensor signals (e.g., position tracking, action recognition, and event logs) to produce a coherent estimate of the current world or workflow state with reduced uncertainty. Sensor fusion has a long history in robotics and automation and is essential for real-time context estimation in complex tasks where single sensors are insufficient [1].

Probabilistic and Bayes-based Methods

Probabilistic methods explicitly model sensor uncertainty and temporal dependencies. Hidden Markov Models (HMMs) and Dynamic Bayesian Networks (DBNs) serve as foundational frameworks for inferring hidden states (e.g., current assembly step) from sequences of noisy observations. HMMs maintain a probability distribution over possible states and update these distributions as new sensor events arrive, enabling robustness to sensor noise and intermittent data. DBNs elevate this by incorporating richer temporal dynamics and interdependencies across variables. Particle filters extend these approaches by representing state distributions with weighted samples and updating them over time. These methods work well in scenarios with structured workflows but can become computationally expensive as workflow complexity increases because of the explosion in state space and dependencies.

A concrete example of probabilistic activity recognition in assembly tasks uses Bayesian filtering over structured state representations such as multi-hypergraphs to infer current assembly actions from sensor streams. This approach demonstrates how probabilistic inference can handle combinatorial state complexity in real manual work processes [2].

Machine Learning-Based Approaches

Deep learning techniques have transformed sensor data interpretation by learning rich representations directly from raw sensor event sequences. Recurrent Neural Networks (RNNs) and related architectures like Long Short-Term Memory (LSTM) models can learn temporal dependencies in sequential data. Transformer models, with their self-attention mechanisms, have been employed to identify salient events and relationships in long event streams without strict reliance on sequential recurrency. Graph Neural Networks (GNNs) model relationships between entities such as workers, tools, and materials, representing them as graph nodes where edges denote interactions. This representation captures spatial and relational dependencies that are difficult to express with purely sequential models [3]. Hybrid architectures combining deep learning with symbolic or probabilistic reasoning are increasingly explored to balance flexibility and interpretability.

Rule-Based and Constraint-Satisfaction Systems

In contrast to data-driven methods, rule-based systems rely on explicitly defined logical rules (e.g., “Tool Y must be picked before Action Z”). These systems enforce logical consistency and interpretability, making them suitable for safety-critical scenarios where unpredictable decisions cannot be tolerated. Constraint-satisfaction frameworks formalize workflow constraints using temporal logic or mathematical constraints, enabling the direct detection of rule violations. These systems, however, lack the adaptive learning capacity of probabilistic and machine learning approaches and can struggle to generalize when workflows evolve. Hybrid systems leverage rule-based frameworks for their structure while using learning components to deal with exceptions and variability.

Strengths and Limitations

- Probabilistic models: strong in uncertainty handling, but scale poorly with complex workflows.
- Machine learning models: adaptable and powerful at pattern recognition, but often opaque and data hungry.
- Rule-based systems: deterministic and interpretable but limited in flexibility for unanticipated variations.

3.2 Deviation Detection

State-of-the-art deviation detection is evolving toward human-centric, AI-assisted systems capable of continuously aligning observed operator actions with expected work order procedures. While automotive manufacturing provides methodological foundations, textile manufacturing—exemplified by Petratex—drives innovation in flexibility, semantic interpretation, and expert knowledge capture. Within Artwork, deviation detection is positioned as a core mechanism to support scalable training, quality consistency, and resilience in highly variable textile production environments.

3.2.1 Deviation Detection in Human-Centric Textile and Automotive Production Processes

In the context of the Artwork project and the Petratex industrial environment, deviation detection refers to the identification of mismatches between the expected execution of a work order, as defined in digital work instructions, and the actual actions performed by an operator at a workstation. This capability is increasingly critical in service-based industrial models, where production capacity is sold as manufacturing hours and processes, and where high variability, short production line life cycles, and frequent onboarding of new operators are the norm.

Deviation detection is a key enabler for quality assurance, operator training, knowledge transfer, and production consistency, particularly in environments where expert operators must support multiple workstations simultaneously. Unlike traditional production monitoring systems, which focus on machine signals and throughput, deviation detection in Artwork targets the human execution of procedures, interpreting how operators interact with machines, tools, materials, and the workstation context.

3.2.2 Reference Procedure and Deviation Typology

In state-of-the-art systems, the reference procedure is derived from work instructions associated with a work order, often expressed as sequences of actions, machine configurations, and material handling steps. Deviations are commonly classified as:

- **Temporal deviations**, such as missing or delayed execution of instruction steps;
- **Structural deviations**, involving incorrect sequencing of operations;
- **Spatial deviations**, including improper posture, hand positioning, or workspace usage;
- **Semantic deviations**, where the operator manipulates an incorrect tool, material, or machine parameter relative to the work instruction [4],[5].

Recent approaches emphasize semantic alignment between observed actions and instruction intent, rather than strict step-by-step enforcement, allowing greater robustness in variable production contexts [6].

3.2.3 Automotive Industry as a Reference Baseline

The automotive industry provides a mature baseline for deviation detection research, supported by highly standardized workstations and well-formalized assembly instructions. Vision-based systems using RGB-D cameras and pose estimation are commonly employed to recognize operator actions and compare them against expected task sequences using temporal models such as LSTMs and Transformers [7],[8].

In this sector, deviation detection is primarily used for assembly validation, operator certification, and quality defect prevention. Hybrid approaches combining operator actions with machine states and tool data have demonstrated improved reliability [9]. However, the high level of process standardization in automotive manufacturing limits the applicability of these solutions to more flexible and less formalized environments such as textile production.

3.2.4 Textile Industry Specificities and Challenges

Textile manufacturing, particularly in the JVC/Petratex context, presents significantly higher procedural variability. Operators perform manual or semi-manual tasks involving flexible materials, frequent machine reconfiguration, and continuous adaptation to customer-specific requirements. Many procedures rely heavily on tacit expert knowledge, which is difficult to formalize in traditional work instructions [10].

State-of-the-art textile-oriented approaches focus on:

- Fine-grained analysis of operator gestures and hand-material interaction;
- Context-aware interpretation of actions relative to fabric type, machine setup, and work order constraints;
- Distinguishing between non-compliant deviations and acceptable operational variations that reflect expert adaptation [11].

This distinction is fundamental in Artwork, where the objective is not to constrain operators, but to capture, normalize, and disseminate expert know-how across the factory floor.

3.2.5 AI-Driven Multimodal Deviation Detection

Since 2019, deviation detection has increasingly relied on multimodal AI architectures, combining video streams, depth data, machine telemetry, and contextual work order information [12]. In the Artwork approach, these modalities are used to digitally recognize machines, tools, materials, and operator actions during the execution of a work order.

Emerging research explores the use of semantic reasoning layers and Large Language Models (LLMs) to translate low-level action descriptions into clear, operator-oriented instruction language, adapted to different skill levels [13]. This enables deviation detection systems to function not only as compliance mechanisms, but also as training and assistance tools.

3.2.6 Open Challenges in the Artwork Context

Despite advances, several challenges remain highly relevant for Petratex and Artwork:

- Work instructions may be incomplete, implicit, or evolve during production;
- Models must generalize across machines, fabrics, and rapidly changing work orders;
- Detected deviations must be explainable and actionable for operators and supervisors;
- Human acceptance and compliance with EU data protection and worker monitoring regulations must be ensured [14].

3.3 Task Re-Planning

In this part we talk about current approaches to tackle contextual robot task planning to assist a human worker. In general, robotic task planning is currently undergoing a fundamental shift, moving away from static, pre-programmed instruction sequences toward dynamic systems capable of reasoning about changing environments. This is primarily driven by either foundation models which provide high-level semantic reasoning or reactive motion generation, which focuses on physical safety and execution.

A current trend is to shift towards the integration of LLMs and Vision-Language Models (VLMs) as the cognitive core of robotic systems. The advantage here is the vast knowledge they contain to ideally provide zero-shot planning without task-specific training. Further, robotic foundation models are being explored that, given some sensor input such as a camera, can move a robot end-effector simply by telling it about the task by text. While the potential of such solutions is high, they for now lack robustness and can be challenging to deploy given the high computational resources required.

Other works instead use LLMs or VLMs as high level planner and rely on more classical controllers to execute the actual motions. While those methods lack generality, safety and convergence to task locations can be guaranteed, which is currently not possible for robotic foundation models.

4 Approach

Out of the many different methods discussed in the previous section, this section dives into the ones implemented by the different Artwork's partners to deal with the industrial challenges.

4.1 Sensor Fusion based on Fuzzy Finite State Machines (FFSM)

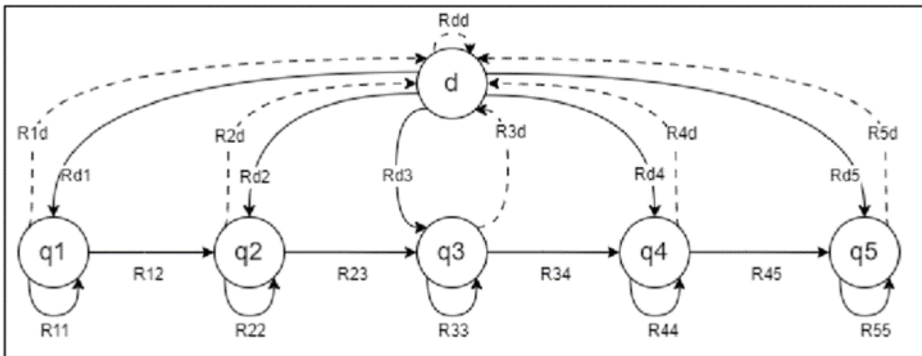
Our proposed solution for real-time operation recognition in industrial environments addresses the critical need for a framework that balances adaptability, interpretability, and computational efficiency. Existing machine learning-based methods, while powerful in pattern recognition, often suffer from opacity and high computational demands, making them less suitable for real-time applications where latency and trust in decision-making are paramount. Conversely, rule-based systems, though transparent and interpretable, lack the flexibility to accommodate dynamic or unpredictable worker behaviors, which are common in complex industrial tasks. To bridge this gap, the approach leverages fuzzy logic as a foundational element, enabling nuanced handling of uncertain or ambiguous sensor inputs while maintaining clarity in decision-making processes. By integrating fuzzy set theory, the framework transforms raw sensor data (such as tool usage, motion patterns, or material interactions) into interpretable linguistic variables, allowing for gradual state transitions rather than rigid binary classifications. This not only enhances robustness against noise but also introduces a layer of uncertainty representation that aligns with real-world operational variability.

At the core of this methodology is a hybridized sensor fusion strategy that combines signal preprocessing with adaptive state modeling. Input signals from diverse sources, such as cameras or wearable sensors, undergo smoothing techniques to mitigate abrupt transitions and noise, ensuring smoother and more reliable data interpretation. This preprocessing step is complemented by a fuzzy finite state machine (FFSM), which dynamically models operations as overlapping states with varying activation levels, rather than discrete transitions. The FFSM's ability to handle partial activations and ambiguous conditions makes it particularly well-suited for industrial scenarios where tasks may involve overlapping phases or unanticipated deviations. However, recognizing the limitations of static rule sets in conventional FFSMs, the proposed approach introduces an evolving structure mechanism that enables the system to learn and adapt to new patterns without compromising interpretability. This adaptive component ensures that the framework remains responsive to worker behavior while retaining the transparency and efficiency required for real-time deployment. By synthesizing these elements, the solution aims to deliver a scalable, interpretable, and computationally lightweight approach to operation recognition in dynamic industrial settings.

The core of the proposed framework is a structure-evolving fuzzy finite-state machine (SEFFSM) that marries real-time sensor fusion with adaptive reasoning. Incoming sensory streams (captured from vision systems, wearable accelerometers, or industrial instrumentation) first undergo a pre-processing stage that smooths abrupt transitions using decaying-tail functions (Gaussian or exponential). This step mitigates impulsive noise and ensures that subsequent state updates reflect gradual, physically plausible changes in the worker's activity. The pre-filtered signals are then mapped onto fuzzy linguistic variables (e.g., Active, Inactive) via membership functions, yielding degrees of activation that encode uncertainty rather than hard binary labels.

With fuzzy input values, the SEFFSM operates on a set of fuzzy states that can coexist with partial activation. Transition rules are encoded as fuzzy implications whose firing strengths depend on both the current state's degree of truth and the input of memberships. Each rule carries a user-defined weight, allowing domain experts to encode prior knowledge or priorities. The system's output is computed by aggregating contributions from all concurrently active rules, thereby producing a smooth, interpretable response that captures the continuous nature of industrial operations.

To remain robust in the face of novel or evolving worker behaviors, the SEFFSM incorporates an online structure-evolution mechanism. When the aggregate firing strength of all active rules falls below a configurable threshold, the system recognizes that its current rule set inadequately describes the observed transition. In such cases, a default fallback state is invoked, and the observed pre- and post-state pair (along with the corresponding sensor readings) is logged for later review. A lightweight rule-generation routine then proposes candidate transitions, which experts can validate based on frequency, consistency, and contextual relevance. This incremental learning approach preserves interpretability while granting the system the flexibility to adapt to unforeseen task variations, all while maintaining computational efficiency suitable for real-time industrial deployment.



State	Operation
d	Default
q1	Grabbing Glue Gun
q2	Gluing with Glue Gun
q3	Grabbing Screwdriver
q4	Grabbing Screw
q5	Screwing the Screw

Figure 1: Structure-Evolving Fuzzy Finite State Machine

Figure 1 depicts a Structure-Evolving Fuzzy Finite State Machine (SEFFSM) representation, illustrating how industrial operations (such as tool usage or task execution) are modeled as fuzzy states with gradual transitions rather than rigid discrete steps. The diagram features a central default state (d) and five operational states (q1 to q5), each corresponding to specific tasks like "Grabbing Glue Gun" or "Screwing the Screw," as defined in the accompanying table. Arrows between states (e.g., R12, R23) represent fuzzy transition rules, where the system dynamically evaluates the degree of activation for each state based on sensor inputs, such as motion patterns or tool interactions, pre-processed via smoothing techniques (e.g., Gaussian decay) to mitigate noise. The dashed loops (e.g., R11, R22) indicate self-transitions with partial activation, reflecting the overlapping or ambiguous nature of real-world operations, while the default state (d) serves as a fallback for unrecognized transitions, enabling the system's adaptive learning mechanism. This aligns with the SEFFSM's core principles—fuzzy logic for uncertainty handling, pre-filtering for robustness, and structure evolution to incorporate new patterns—ensuring real-time interpretability and scalability in dynamic industrial environments such as bus production.

4.2 Computer vision techniques

Inverse Perspective Mapping (IPM) and Multi-View Image Stitching work together to create a seamless top-down representation of an environment from camera images from different angles. IPM is the mathematical process that eliminates the perspective inherent in standard cameras by re-projecting pixels from a 2D image plane onto a common ground plane as seen in Figure 2(a). Once each individual camera has been transformed via IPM, Multi-View Image Stitching is used to align and merge these overlapping projections into a single, cohesive map Figure2(b).

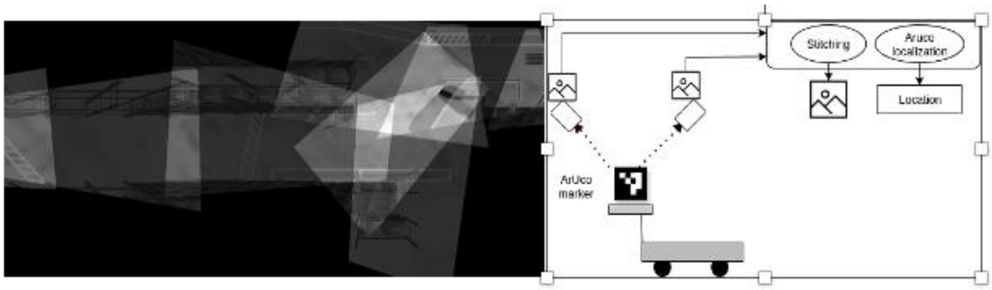


Figure 2: Multi Camera Perception

An ArUco marker (often called an ArUco code) is a square landmark used in computer vision to help robots and cameras understand their position in space. These markers that are attached to different objects (e.g. robots, robotic arms, human pose) as seen in Figure 2(b) are designed specifically for fast and reliable 3D tracking. They are used as fixed reference points to calculate three distinct pieces of information: position (relative to the camera), orientation (the angle of the marker), and localization (the absolute position within a space). One major issue with ArUco code localization is occlusion, which can cause a total loss of tracking if the line-of-sight is interrupted. Because the algorithm requires a clear view of all four corners and the internal grid to calculate a pose, therefore small obstruction, like a shadow or a passing object, will prevent the marker from being detected entirely.

4.3 Deviation Detection

In the Artwork/Petratex context, deviation detection must operate in highly variable, human-centric production environments, where work orders change frequently; manual operations dominate, and expert know-how plays a central role. The proposed approach treats deviation detection not as rigid error checking, but as a continuous alignment process between a planned reference procedure and the observed execution of a work order at a workstation, using multimodal sensor fusion.

4.3.1 Reference Method: Normative Backbone with Expert Variability

Each work order is represented through a two-layer reference method. The normative layer encodes the work-instruction backbone: mandatory steps, optional steps, admissible sequences, tool–machine–material constraints, and safety-critical conditions. This layer defines what must happen to ensure quality and compliance.
The empirical layer is derived from expert operator demonstrations, captured during teaching or simulation sessions. It models typical execution patterns, step durations, micro-actions, and common adaptations observed in real production.

This separation allows the system to distinguish non-compliant deviations from legitimate operator variability, a critical requirement in textile manufacturing where adaptation is often necessary.

4.3.2 Sensor Fusion Focused on Evidence of Task Completion

Rather than relying on fine-grained gesture classification, the system uses sensor fusion to infer evidence that a task step is being executed or completed. Inputs may include video and depth cameras (operator posture, hand–object interaction, workstation zones), tool or material identification, and machine telemetry when available. Minimal operator inputs (e.g. start, pause, blocked) are used to increase robustness and explainability.

These heterogeneous signals are fused into a step-level belief score, representing the likelihood that a given work-instruction step is currently active or completed.

4.3.3 Online Plan–Execution Alignment

Deviation detection is performed through online alignment between the planned task graph and the incoming evidence stream. The task planner continuously maintains hypotheses about the current step, allowing temporal flexibility and alternative sequences defined in the reference method.

Deviations are detected as violations of constraints, rather than simple mismatches, and are classified as:

- Hard deviations, such as skipped mandatory steps, wrong tools or materials, or unsafe machine states;
- Soft deviations, including unusual timing or uncommon action patterns that remain within acceptable bounds.

This mechanism supports both structured automotive-like procedures and flexible textile workflows.

4.3.4 Deviation Handling and Adaptive Task Planning

Detected deviations are immediately contextualized and linked to actionable responses. Hard deviations may trigger guidance, operator feedback, or task interruption, while soft deviations are logged for monitoring or coaching. Where necessary, the task planning tool can perform local replanning, proposing recovery paths, alternative sequences, or controlled rework aligned with the work order objectives.

Repeated, expert-validated variations are candidates to be incorporated into the empirical layer, enabling continuous refinement of work instructions and scalable knowledge transfer.

4.3.5 Conclusion

By combining a dual-layer reference method with sensor fusion and online plan–execution alignment, the proposed approach enables robust, explainable, and human-centred deviation detection. It supports quality assurance and training at Petratex without over-constraining operators, while progressively formalizing expert know-how into reusable digital work instructions—fully aligned with the goals of the Artwork Task Planning and Teaching Instructions framework.

4.4 Task Re-Planning

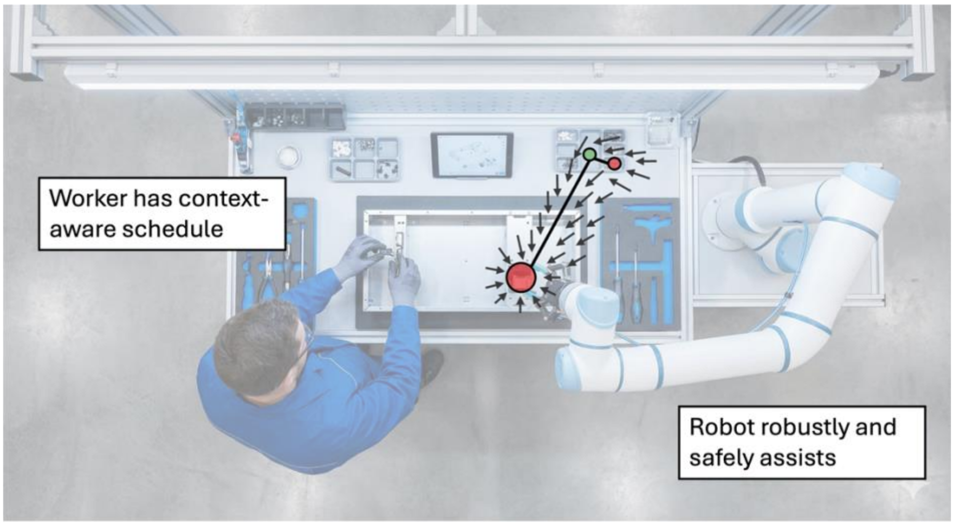


Figure 3: Human-Robot collaboration demo in a shared workspace

Given that we are gathering information about the environment, we want aim to use it for a robot assistant that is both context-aware, inherently safe, and reactive to dynamic changes in the workspace (See Figure 3: Human-Robot collaboration demo in a shared workspace). To achieve this, we propose coupling a robust graph-based representation of motion tasks with a high-frequency reactive control layer.

At the core of our solution is a method we call the Gaussian Graph. Instead of memorizing a single, fixed path for a specific task, our system breaks down human demonstrations or predefined trajectories into a network of small, stable motion regions. These regions can be seen as "steppingstones" of valid movement spread across the workspace. By connecting them based on their direction and proximity, we create a navigable roadmap of the robot's capabilities. This allows the robot to "stitch" together different parts of previously learned tasks to form entirely new trajectories. For example, if a worker requests a tool at a new location, the robot can intelligently combine the beginning of one motion with the end of another to reach that specific target, effectively adapting to the new context instantly without needing to be retaught. Notably, the context does not have to be goal location only and can contain any other information. The context simply determines the valid motion regions, so tool positions, body posture or assembly instructions are inherently compatible. Crucially, this navigation is strictly confined to areas where the robot has learned safe behavior, ensuring it never attempts to move through unverified or dangerous parts of the workspace.

Now given the high-level plan of motion regions, we employ a low-level reactive controller based on dynamical systems [15]. While the Gaussian Graph determines the path, this layer ensures that the execution is safe. The first part to achieve this is that the low-level controller computes vector fields that lead towards the goal location. Therefore, the robot movement becomes stable and can be combined with compliant control, such that the work can move the robot by hand if needed without disturbing the robot's task. Further, we utilize obtained distance fields using RGB-D cameras to create a continuous representation of the surroundings [16]. This allows the robot to smoothly adjust its motion to maneuver around obstacles and humans that are in the same space without stopping its motion. By combining the context-aware planning using the Gaussian Graph and the fast-acting, safe controller we aim to achieve a robotic assistant that is both robust enough to handle changing tasks but at the same time aware enough to work safely alongside humans.

The method is integrated into a ROS2 node that controls a robot manipulator by [1]￼.

[1] The code is open source and can be found at https://github.com/KilianFt/DSStitchingNode

5 Preliminary Results

5.1 Sensor Fusion based on Fuzzy Finite State Machines (FFSM)

The proposed SEFFSM solution for real-time operation recognition in industrial environments has been successfully implemented as a laboratory demonstrator at IFAK, showcasing its efficacy in dynamic, sensor-driven task monitoring. This hybrid framework, combining fuzzy logic with adaptive state modeling, enables robust and interpretable operation identification by transforming raw sensor inputs (e.g., from vision systems or wearable devices) into gradual, linguistically meaningful states. The system's ability to handle partial activations, ambiguous transitions, and incremental learning (via a fallback default state and expert validated rule evolution) ensures scalability and real-time responsiveness, critical for applications like context-aware assistance systems for bus production. The approach has been documented in the peer-reviewed article (Real-Time Operation Identification Using a Structure Evolving Fuzzy Finite State Machine"), published at the 2025 IEEE 8th International Conference on Industrial Cyber-Physical Systems (ICPS).

5.2 Tablet-Based Deviation Detection at Sewing Workstations

As a preliminary result of the implementation of the proposed approach to JVC/Petratex a software application was developed and deployed on tablets installed at individual sewing workstations at JVC. This implementation serves as an initial proof of concept for runtime comparison between reference work instructions and actual operator execution, focusing on usability, learning support, and deviation awareness in a real production environment.

The application presents a **dual-pane user interface**, explicitly designed to support operator understanding and self-correction:

- **Reference Execution Area (left panel):**
  Displays the digital work instruction associated with the current work order and textile piece. This includes structured step descriptions and a reference video illustrating the correct execution of the sewing operation, as defined during expert teaching or instruction authoring.
- **Runtime Execution Area (right panel):**
  Displays a live or recorded video stream capturing the actual operation performed by the operator at the workstation.

This side-by-side layout enables immediate visual comparison between the expected and observed execution, reinforcing learning through direct observation.

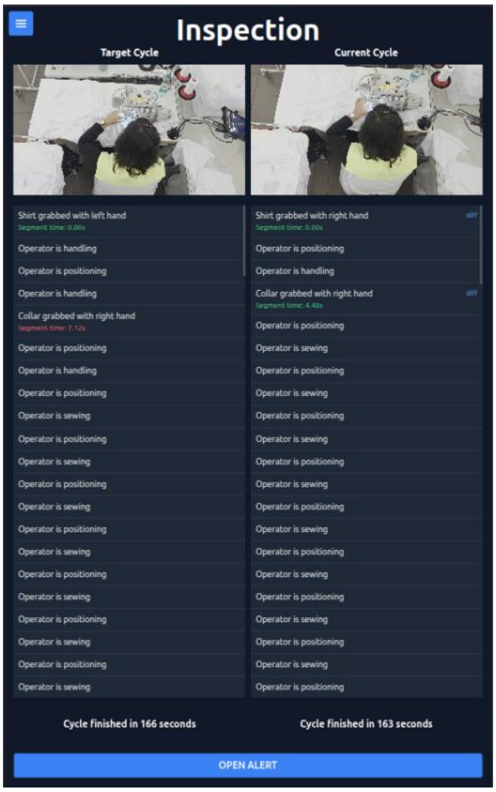


Figure 4: dual-pane user interface

During operation, the system continuously analyses the runtime video stream and generates movement descriptions and action indicators derived from the operator’s interaction with the workstation. These descriptions are compared against the reference work instruction, taking into account:

- the **sequence of operational steps**,
- the **relative timing and duration** of each step,
- and the **alignment between observed actions and expected instruction phases**.

Detected differences between the reference execution and the observed execution are automatically signaled to the operator, without interrupting the production flow. The system does not enforce hard stops at this stage but instead focuses on awareness and learning.

When deviations are detected, the operator can access the specific video frames or segments where the differences occur. This allows the worker to review their own execution in direct comparison with the reference instruction, fostering self-guided learning and procedural understanding.

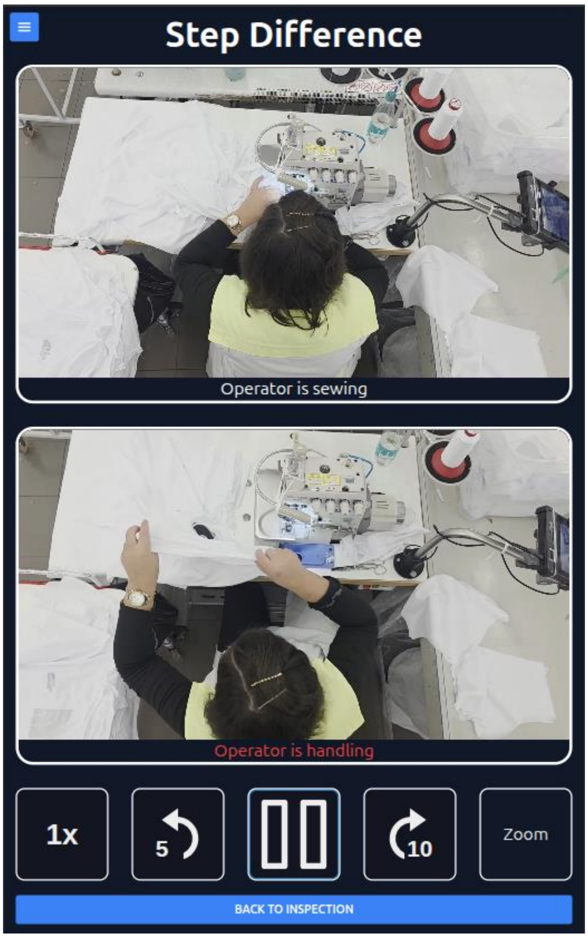


Figure 5: Analyzing video frames or segments where the differences occurred

This interaction model supports:

- on-the-job training for less experienced operators,
- reinforcement of correct techniques,
- and reflection on alternative execution strategies.

Importantly, the system treats deviations primarily as learning opportunities, rather than immediate errors, which is consistent with the JVC/Petratex production context where controlled variability and operator adaptation are common.

5.3 Task re-planning

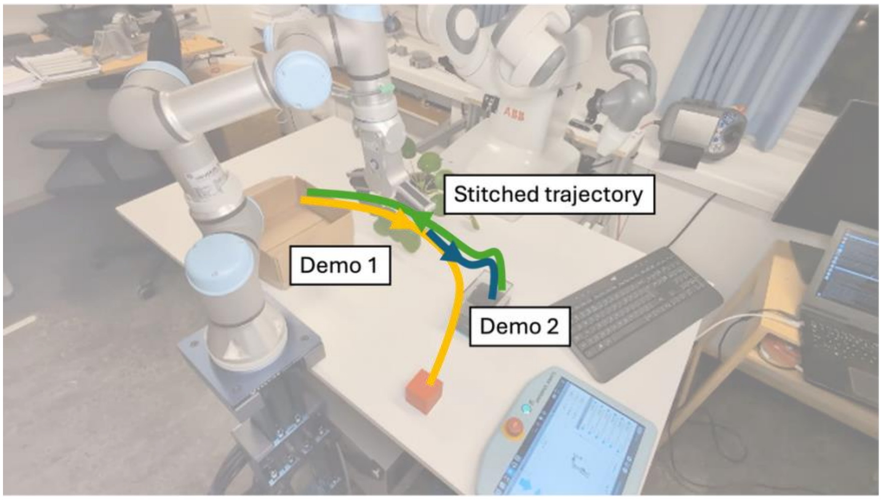


Figure 6: Robot demonstrator with contextual trajectory stitching method

For contextual task replanning, we have developed the core method that represents a workspace as a task and simplifies learning from demonstrations with a worker. In this work, we prove that our graph representation converges towards goal locations and show how to effectively combine different demonstrations. Further, we tested the method on a pick and place task with a UR3e[2] where an operator gives two demonstration sets: 1) the robot moving from package to picking box A and 2) the robot picking box B (See Figure 6: Robot demonstrator with contextual trajectory stitching method). It is able to stitch together the demonstrations and form a sequence of first moving towards box B, picking it and then moving back towards the package to drop it. Importantly, as mentioned before, the path the robot takes is represented as a vector field, so it is robust to disturbances. Further, note that the task can be changed online at any point which enables smooth task re-planning to integrate with context-aware worker systems. Future work contains the extension of this demonstrator to handle picking more robustly and integrate compliant controllers with obstacle avoidance to smoothly handle collaborative work.

Positioning within the Artwork Approach

This preliminary implementation validates key assumptions of the Artwork deviation detection strategy:

- that **visual comparison** between reference and runtime execution is intuitive for operators,
- that **timing- and sequence-based deviation detection** can be meaningfully applied in textile sewing operations,

[2] https://www.universal-robots.com/products/ur3e/

- and that deviation feedback can be delivered in a **non-intrusive, operator-centric manner**.

While the current solution relies mainly on video-based analysis, it establishes the foundation for future integration with additional sensor modalities (e.g. machine signals, tool identification) and tighter coupling with the Task Planning Tool for adaptive guidance and replanning.

6 Conclusions and Future Work

This deliverable has advanced the project from environment sensing toward context-aware task planning, demonstrating how perception outputs can be transformed into actionable task-level reasoning. By integrating detection, localization, and intention prediction into planning processes, the work establishes a scalable foundation for adaptive decision-making in complex industrial environments. The presented approaches address key industrial requirements related to flexibility, safety, and operational robustness across the targeted industries.

The results contribute directly to the project KPIs by enabling measurable improvements in task efficiency, error reduction, and system responsiveness, while maintaining compliance with safety and privacy constraints. The proposed planning concepts are designed to operate under realistic industrial conditions and will be validated in representative scenarios within truck production, bus assembly, and textile manufacturing. These validation scenarios provide a structured framework for assessing performance, robustness, and user acceptance in real operational contexts.

This work prepares the transition to the next project phase, where a closed-loop, context-based worker assistance system will be implemented. In this phase, task planning will be tightly coupled with execution and feedback, allowing the system to continuously adapt based on observed worker actions and environmental changes. This closed-loop integration will enable KPI-driven validation of end-to-end performance, demonstrating tangible benefits in safety, productivity, and worker support within industrial pilot deployments.

7 Bibliography

[1] “Sensor Fusion.” Wikipedia, The Free Encyclopedia, 2026. [Online]. Available: https://en.wikipedia.org/wiki/Sensor_fusion

[2] T. Felske, S. Lüdtke, S. Bader, and T. Kirste, “Activity Recognition in Assembly Tasks by Bayesian Filtering in Multi-Hypergraphs,” Feb. 01, 2022, *arXiv*: arXiv:2202.00332. doi: 10.48550/arXiv.2202.00332.

[3] K. Muthineni, A. Artemenko, D. Abode, J. Vidal, and M. Nájar, “PosGNN: A Graph Neural Network Based Multimodal Data Fusion for Indoor Positioning in Industrial Non-Line-of-Sight Scenarios,” *IEEE Open J. Veh. Technol.*, 2025, Accessed: Jan. 30, 2026. Available: https://ieeexplore.ieee.org/abstract/document/11235985/

[4] Q. Xiong, J. Zhang, P. Wang, D. Liu, and R. X. Gao, “Transferable two-stream convolutional neural network for human action recognition,” *J. Manuf. Syst.*, vol. 56, pp. 605–614, Jul. 2020, doi: 10.1016/j.jmsy.2020.04.007.

[5] M.-P. Pacaux-Lemoine, Q. Berdal, C. Guérin, P. Rauffet, C. Chauvin, and D. Trentesaux, “Designing human–system cooperation in industry 4.0 with cognitive work analysis: a first evaluation,” *Cogn. Technol. Work*, vol. 24, no. 1, pp. 93–111, Feb. 2022, doi: 10.1007/s10111-021-00667-y.

[6] C. Yang *et al.*, “Ontology-based knowledge representation of industrial production workflow,” *Adv. Eng. Inform.*, vol. 58, p. 102185, Oct. 2023, doi: 10.1016/j.aei.2023.102185.

[7] “A Comprehensive Survey of Vision-Based Human Action Recognition Methods.” Accessed: Feb. 06, 2026. [Online]. Available: https://www.mdpi.com/1424-8220/19/5/1005

[8] J. Izquierdo-Domenech, J. Linares-Pellicer, and J. Orta-Lopez, “Towards achieving a high degree of situational awareness and multimodal interaction with AR and semantic AI in industrial applications,” *Multimed. Tools Appl.*, vol. 82, no. 10, pp. 15875–15901, Apr. 2023, doi: 10.1007/s11042-022-13803-1.

[9] “Human–Robot Collaboration in Manufacturing Applications: A Review.” Accessed: Feb. 06, 2026. [Online]. Available: https://www.mdpi.com/2218-6581/8/4/100

[10] W. H. Akhtar, C. Watanabe, Y. Tou, and P. Neittaanmäki, “A New Perspective on the Textile and Apparel Industry in the Digital Transformation Era,” *Textiles*, vol. 2, no. 4, pp. 633–656, Dec. 2022, doi: 10.3390/textiles2040037.

[11] E. Fenoglio, E. Kazim, H. Latapie, and A. Koshiyama, “Tacit knowledge elicitation process for industry 4.0,” *Discov. Artif. Intell.*, vol. 2, no. 1, p. 6, Mar. 2022, doi: 10.1007/s44163-022-00020-w.

[12] Z. Wang and J. Yan, “Multi-sensor fusion based industrial action recognition method under the environment of intelligent manufacturing,” *J. Manuf. Syst.*, vol. 74, pp. 575–586, Jun. 2024, doi: 10.1016/j.jmsy.2024.04.019.

[13] C. Trivedi *et al.*, “Explainable AI for Industry 5.0: Vision, Architecture, and Potential Directions,” *IEEE Open J. Ind. Appl.*, vol. 5, pp. 177–208, 2024, doi: 10.1109/OJIA.2024.3399057.

[14] “European Agency for Safety & Health at Work - Information, statistics, legislation and risk assessment tools.” Accessed: Feb. 06, 2026. [Online]. Available: https://osha.europa.eu/en

[15] S. Sun and N. Figueroa, “SE(3) Linear Parameter Varying Dynamical Systems for Globally Asymptotically Stable End-Effector Control,” in *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Oct. 2024, pp. 5152–5159. doi: 10.1109/IROS58592.2024.10801844.

[16] U. Ali *et al.*, “Enabling Safe, Active and Interactive Human-Robot Collaboration via Smooth Distance Fields,” in *2025 20th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, Mar. 2025, pp. 439–446. doi: 10.1109/HRI61500.2025.10974001.