



MARIE

Multimodal Activity Recognition in an Industrial Environment

Presenter: Ron Fritzemeier, Director, Mission Solutions

Project Leads: Giuseppe “Beppe” Raffa (Principal Engineer), Richard Beckwith,
Intelligent Systems Research Lab (ISR)

Changing the Narrative



Human & AI: From Competition to Collaboration

Smart Task Spaces & Human/AI Collaboration

AI-based “task spaces” recognize, guide, and improve the execution of multi-step human tasks

Human/AI Collaboration enables model fine-tuning leveraging domain experts' interaction, rather than AI researchers work



MARIE: Multimodal Activity Recognition in Industrial Environments

“How can AI help technicians perform complex tasks?”

- Factories rely on **well-defined tasks**
 - **Improving execution** is time consuming and requires expert effort
 - **New personnel training** is complex and time consuming
-
- **AI Technologies** can teach technicians to consistently replicate expert performance and support continuous process improvement
 - **Human/AI Collaboration** will make this **scalable** and adaptable to manufacturing tasks and other verticals
 - **MARIE** can increase **compliance and safety** while reducing errors and faults



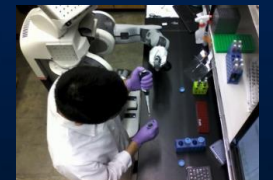
Tutor and guide
for new activities



Recognize and validate
actions



Document and generate
reports



Detect variations
and **assist** users



AI in Manufacturing



Goals



Support complex workplace activities

- Assembly, maintenance, troubleshooting, training

Develop robust task support system

- Provides "buddy-checks" and training
- Helps process improvement

Improved scalability via Human/AI collab

- Simplified data collection/labeling
- Reduced training data volume
- Continuous learning

Value for Techs

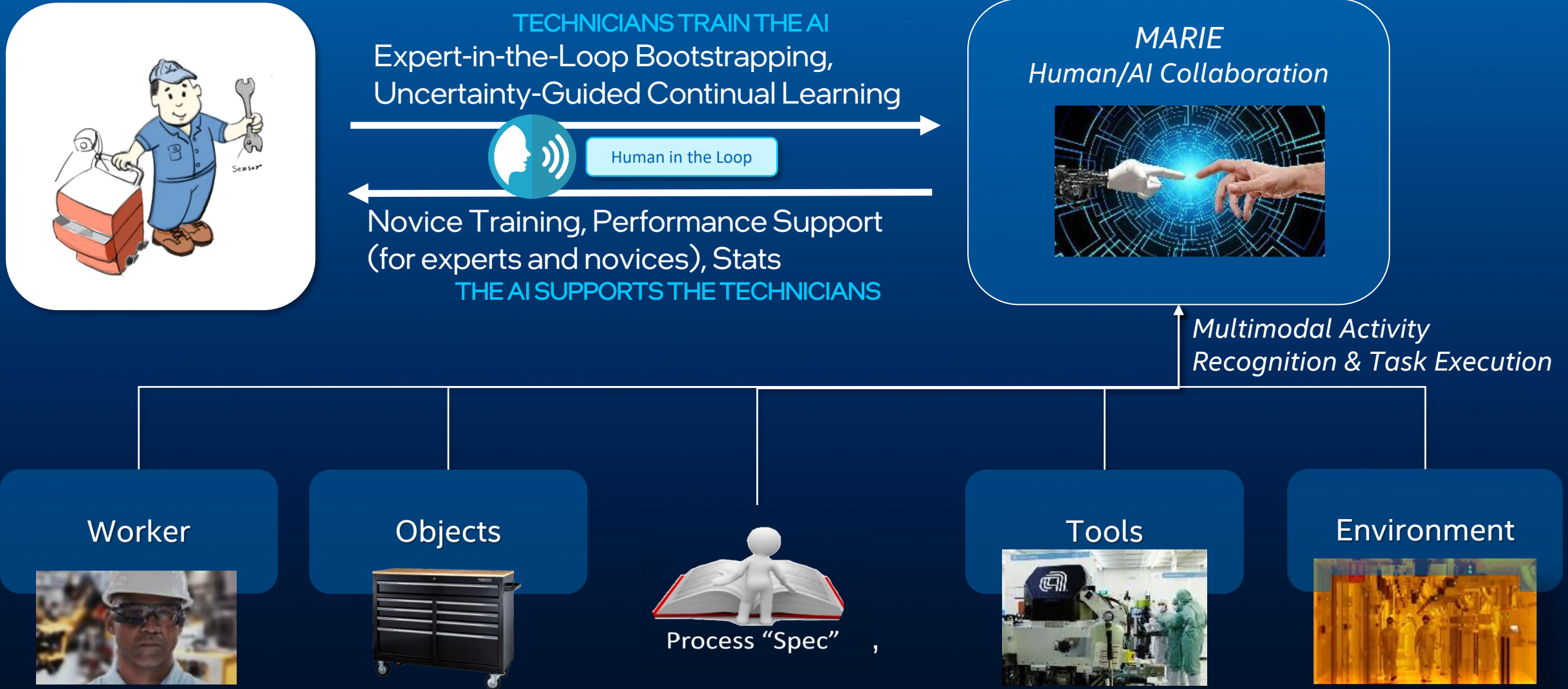
- Provide info from spec and answer questions
- Flag errors as they execute a critical step and support resolution
- On the job training

Value for Manufacturing supervisors

- Performance measures (e.g. tool up-time)
- Identify errors and more error prone steps
- Understand correlation with output (e.g. yield)
- Deviation from norm, different ways a process is executed
- Spec improvement and roll out to new sites

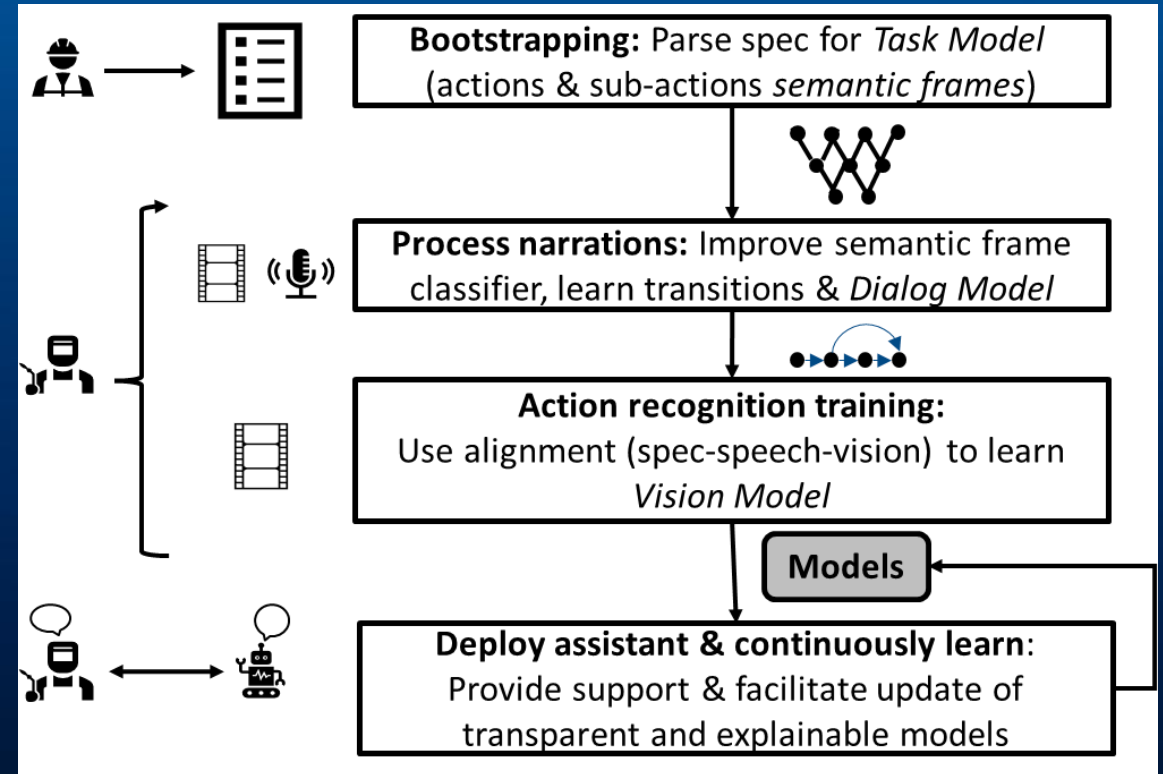


Our “Human in the Loop” Approach



Human-in-the-Loop Learning

- The system starts by parsing the instructions (“specs”) as written by the expert
 - Extraction of ontology and presented action sequence (i.e., a nascent task model)
- MARIE has experts show examples of actions and objects in the procedure
 - Building nascent vision models
- MARIE improves model performance by interacting with experts as they work
 - Fine – tuning of uncertainty-aware vision models and variants of process sequence
- MARIE is fully deployed and helps the technician through situated dialogues
 - Task Assistance
- MARIE, through continued uncertainty guidance, user dialogue for feedback
 - Continual learning, continuous improvement



Grounding for Human-in-the-Loop Learning

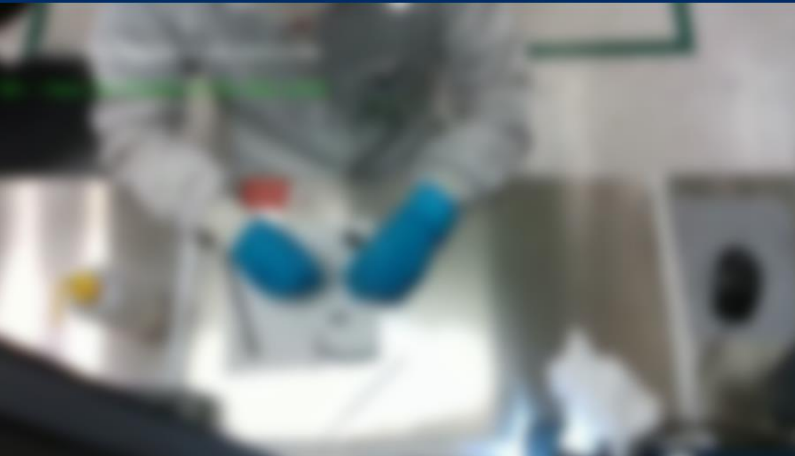


Process “Spec”

Reading an instruction manual or spec, MARIE extracts not only references to actions and objects involved in a procedure but also the presented process execution sequence.

MARIE grounds text references through verbal interaction. MARIE can then use visual data for situational eval and/or use dialog to improve situational awareness.

UNGROUNDED



MARIE asks to see the objects and actions, providing training data for the pre-finetuned recognition system

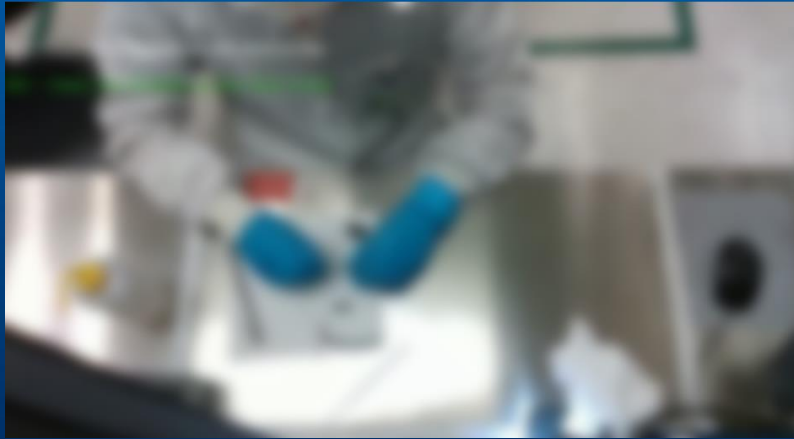
GROUNDING



MARIE and user exploit grounding and situational awareness to collaboratively fine-tune models

Visual Perception: Examples

Action Recognition



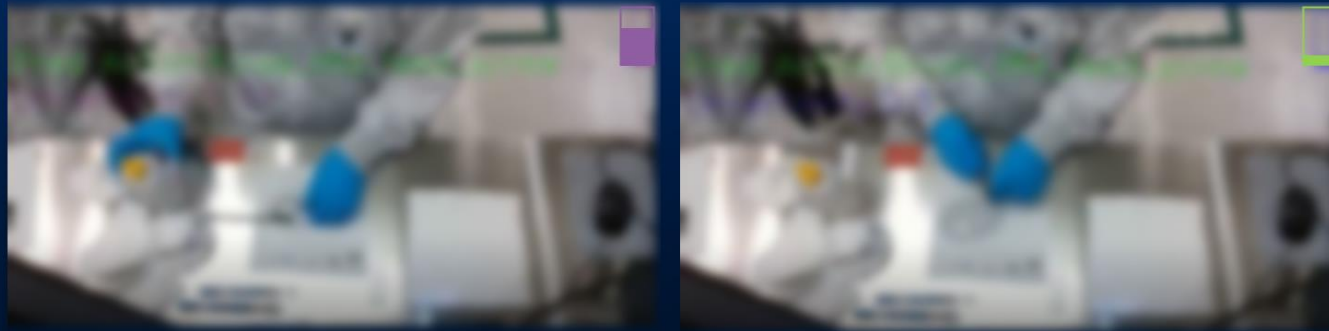
Visual Anomaly Detection



Object Recognition



Uncertainty based Continual Learning



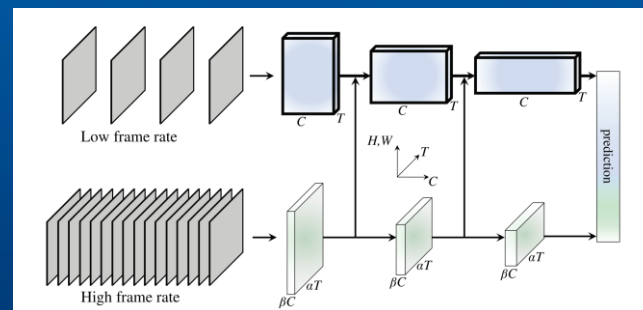
Object Segmentation



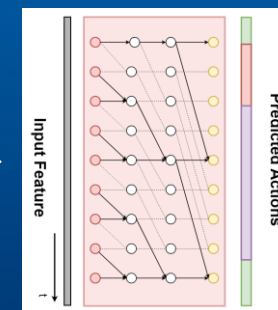
MARIE: *Are you brushing right now?*

Under the Hood: Visual Perception

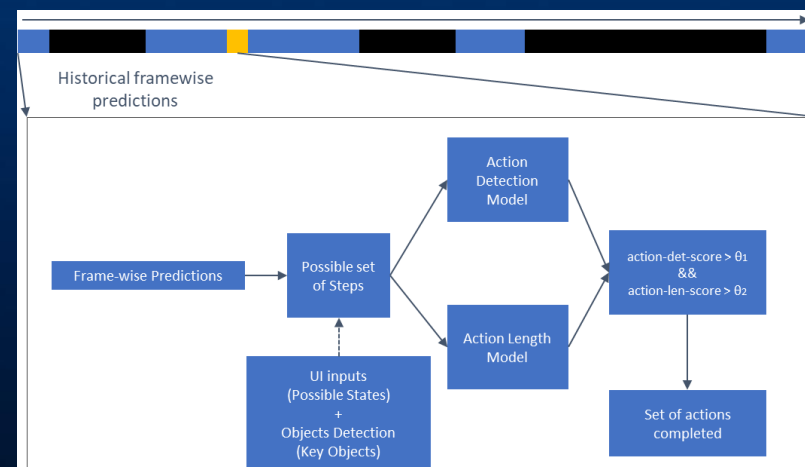
- Real-time fine-grained action recognition
 - CNN-based feature extraction + Causal Temporal Convolution
- Object Recognition and localization
 - Training via few-shot learning from technician demonstrations
- Task Modeling
 - Down-stream component to model allowed action sequences and timing for precise temporal localization
- Uncertainty-guided continual learning
 - Leverage predictive uncertainty estimation (e.g., Bayesian Neural Network)
 - Apply HITL Continual Learning based on model uncertainty estimates and importance sampling



3DCNN Feature Extraction



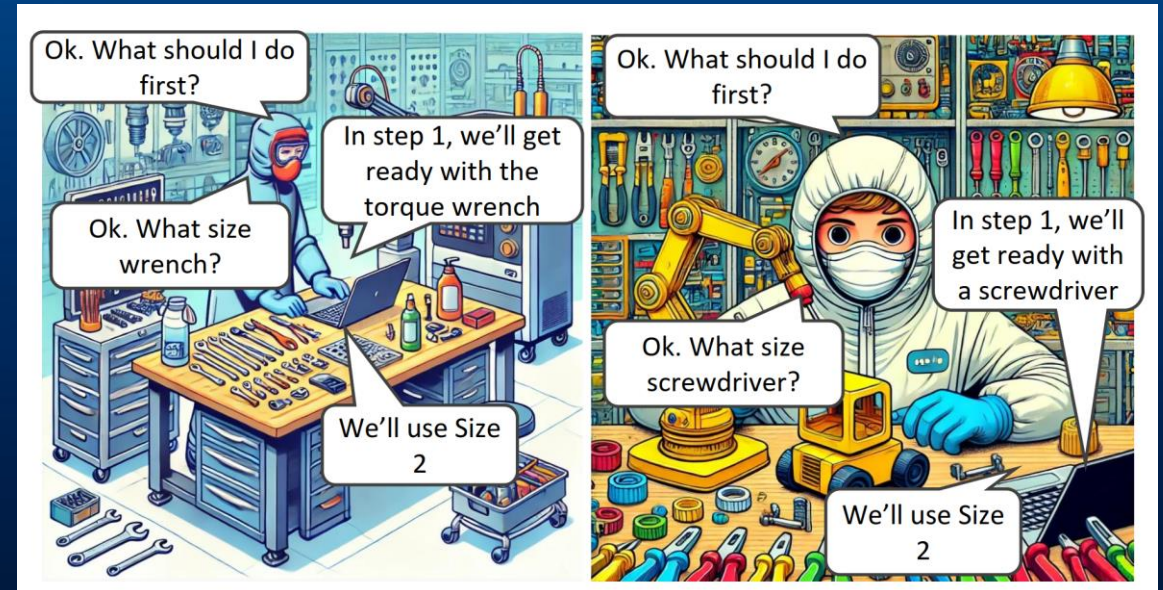
Causal Temporal Convolution



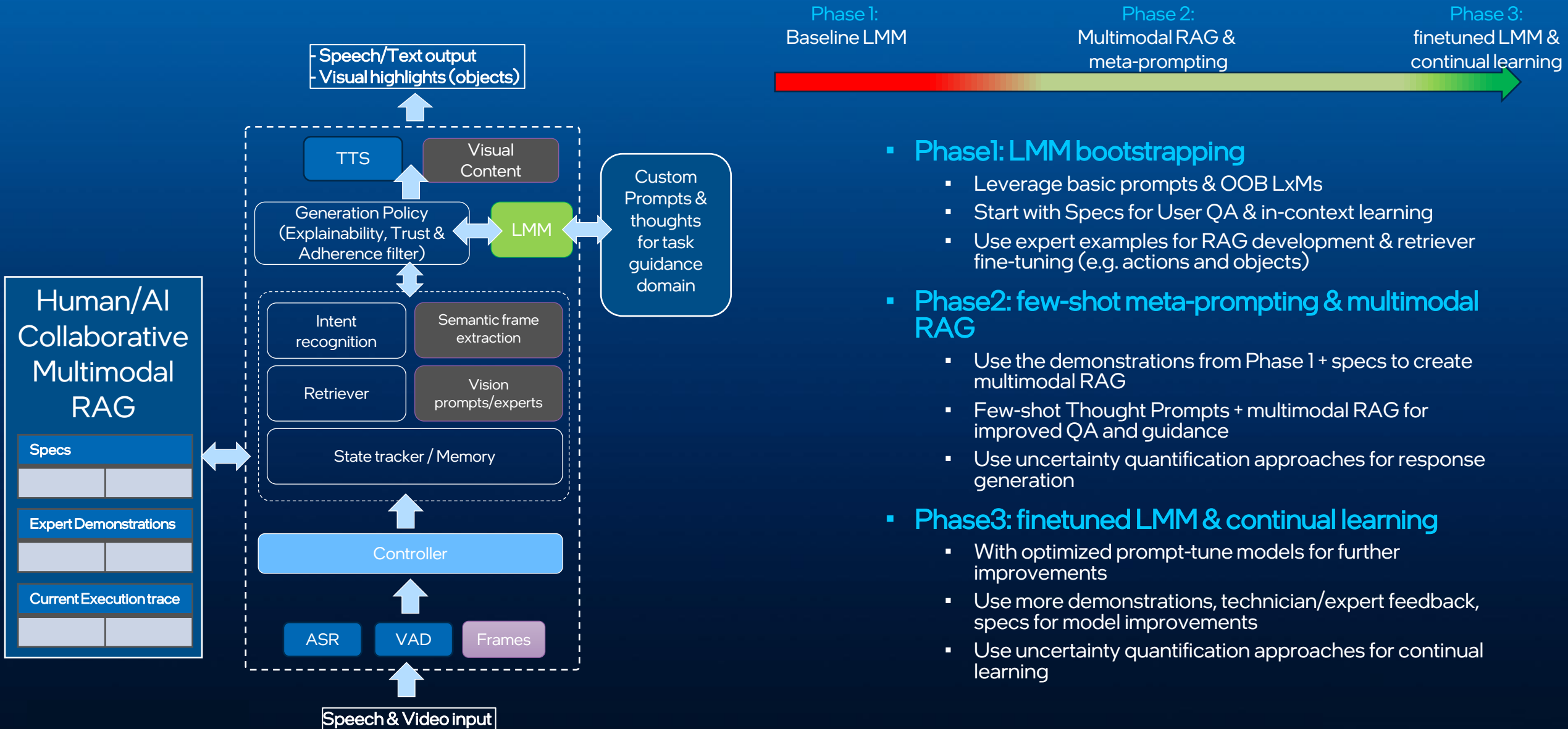
Task Model

Mixed-Initiative Vision-LLM based Dialogue

- MARIE LLM subsystem enables different interactions through an agent-based framework
 - Chat & social obligations
 - QA
 - Task guidance
 - Question generation / follow up
- It also enables effective and trustworthy interaction using multimodal RAG and content filters



Under the Hood: Human/AI Augmented LLM



Human/AI Collaboration in Manufacturing

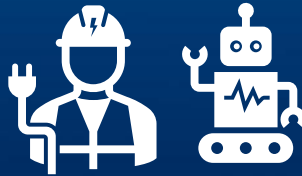
Research Findings & Trust



Dialogue Policies



Responsible & Transparent AI



Human & AI benefit from HAIC



Improved annotation time



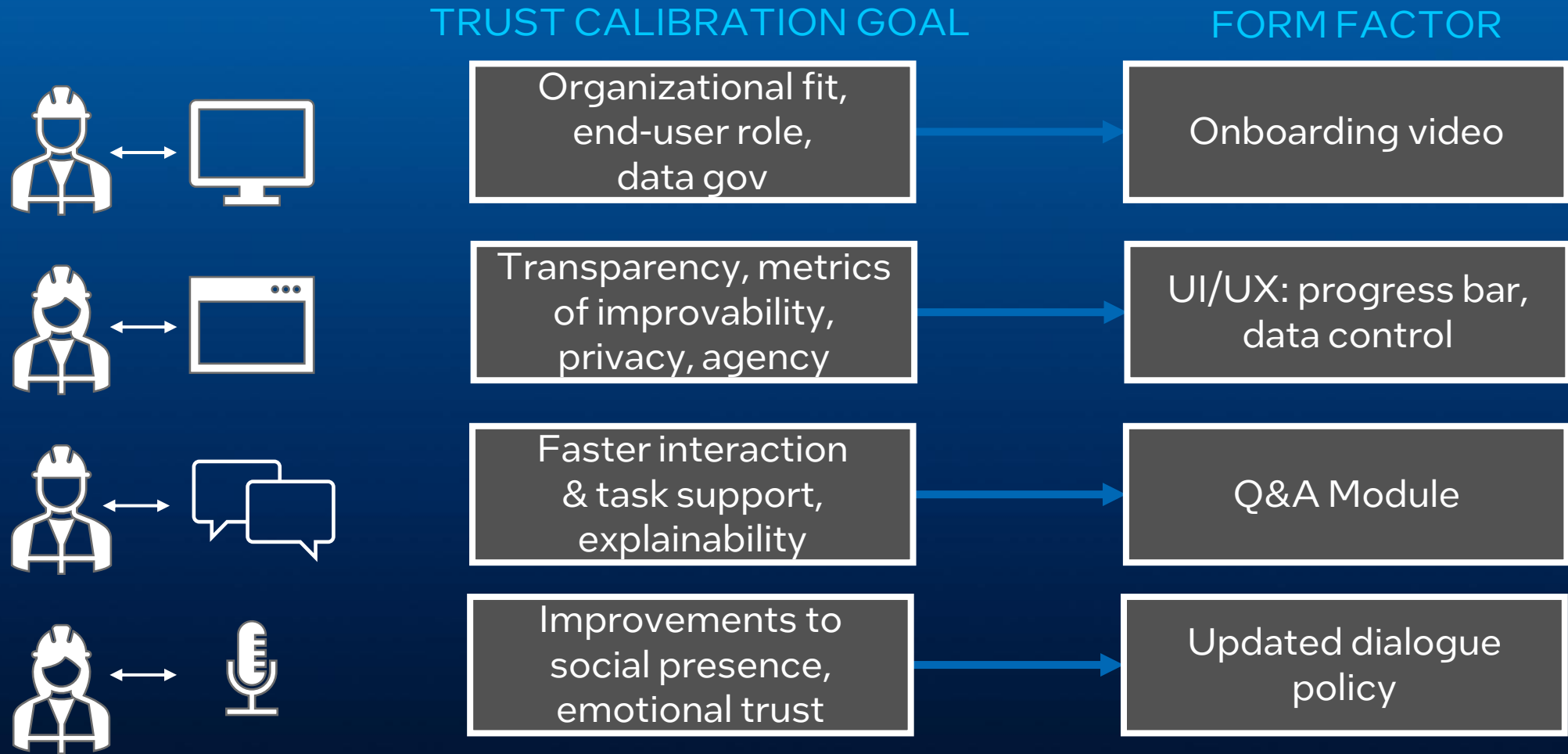
Improved spec compliance

Factors Influencing Trust in AI

Factors Influencing Trust in AI	
AI-related	Context-related
Ability Integrity Benevolence Popularity Familiarity Ease of use	Task difficulty Perceived risks and benefits Situational characteristics Domain's reputation Developer's reputation

Kim, S.S., Watkins, E.A., Russakovsky, O., Fong, R. and Monroy-Hernández, A., 2023, June. Humans, AI, and Context: Understanding End-Users' Trust in a Real-World Computer Vision Application. FAccT 2023.

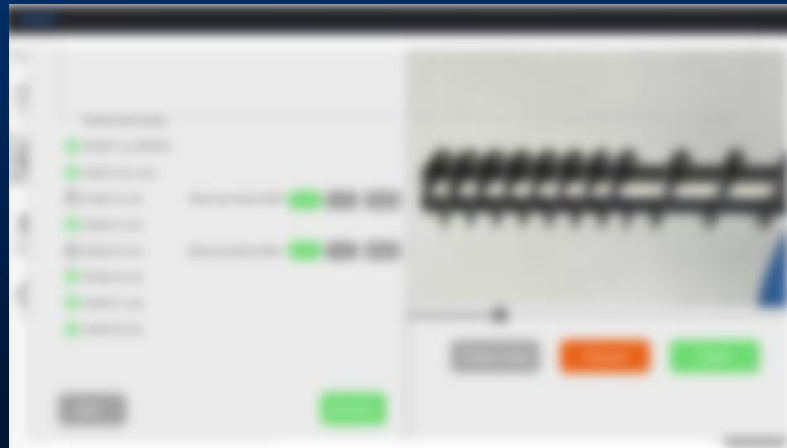
Supporting trust calibration across design + engineering



MARIE POC running in Intel Factory



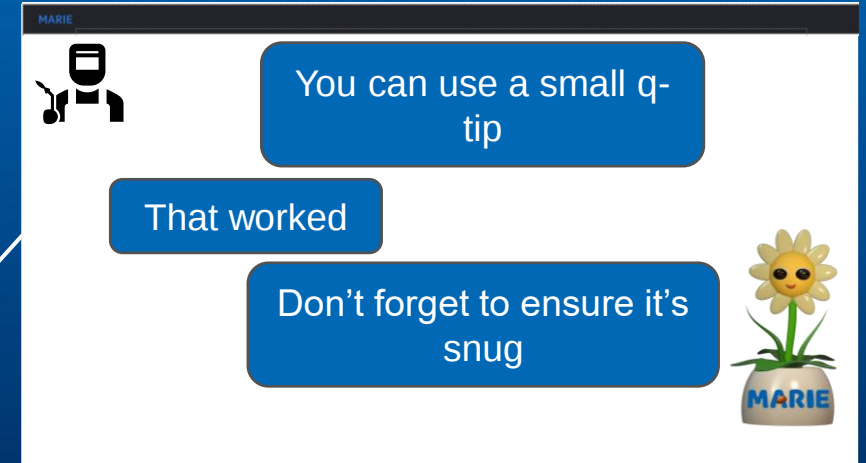
Spec → Process Flow & UI Design



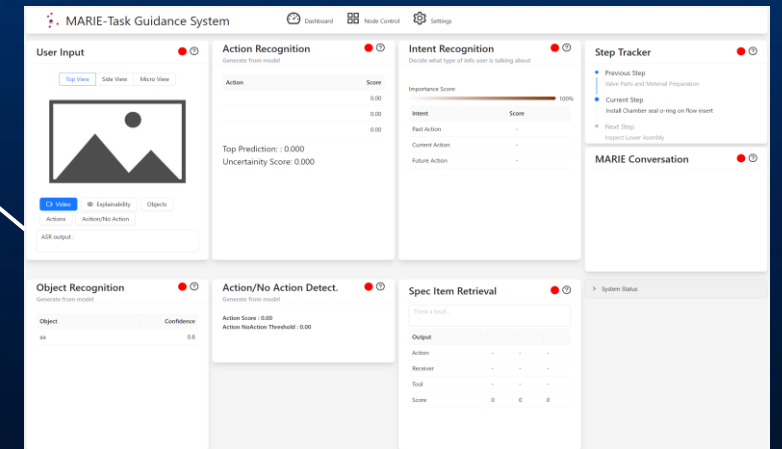
User feedback &
Continual Learning



MARIE



Mixed initiative dialogue

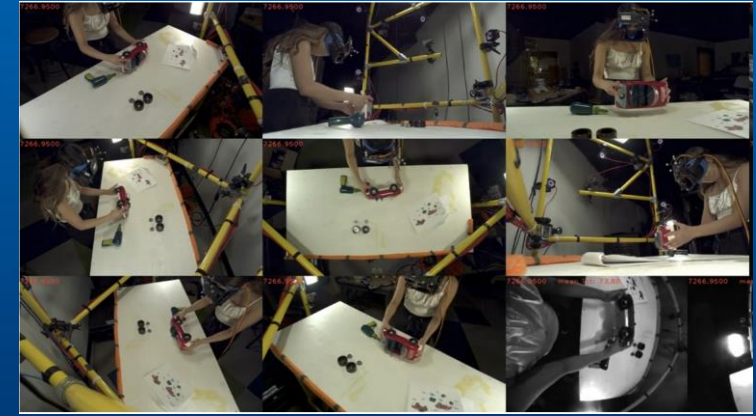


Backend system for
performance analysis and transparency

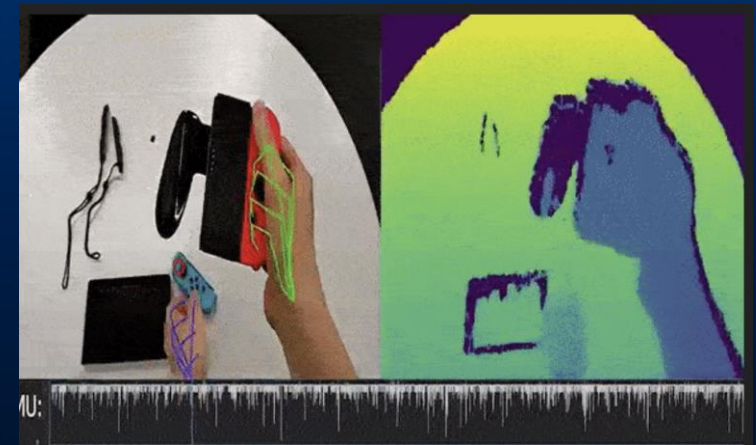
How does MARIE learn from experts?



Expanding to other domains and technical challenges



Assembly101



HOLOASSIST

thank you

MARIE

Multimodal Activity Recognition in an Industrial Environment

Giuseppe Raffa, giuseppe.raffa@intel.com



Q&A