Ref. Ares(2026)8510239 - 31/08/2026

Artificial Intelligence in Manufacturing leading to Sustainability and Industry5.0

Project Acronym:

AIMS5.0

Grant agreement no: 101112089

Deliverable no. and title	**D2.4 - Report on AI tools and algorithms for sustainable industrial processes Y3**	
Work package	WP2	AI tools, methods and algorithms for sustainable industrial processes
Task	T2.1-4	All WP2 Tasks
Subtasks involved	All WP2 subtasks	
Lead contractor	Infineon Technologies AG Thomas Gutt, mailto:thomas.gutt@infineon.com	
Deliverable responsible	BME – Budapest University of Technology and Economics Pal Varga – pvarga@tmit.bme.hu	
Version number	V1.0	
Date	31/08/2026	
Status	Final	
Dissemination level	PU	

Copyright: AIMS5.0 Project Consortium, 2026

Authors

Participant no.	Part. short name	Author name	Chapter(s)
1	EDI	Kārlis Freivalds	All
2	BME	Pal Varga	1,2,3,6
3	EDI	Una Untule	3
4	EDI	Laura Leja	2,3
5	BME	Ruba Almahasneh	All
6	BME	Gergely Hollósi	2
7	BME	Daniel Ficzere	2
8	IFD	Germar Schneider	3
9	IDEKO	Itziar Ricondo	3

Document History

Version	Date	Author name	Reason
v0.1	15.07.2023	Alfred Hoess	Initial Template
V0.2	27.03.2026	Kārlis Freivalds	Structure and initial contents
V0.3	12,08.2026	Ruba Almahasneh	chapter2
V0.4	12.08.2026	Pal Varga	Intro, AIMS Toolbox, Executable Summary, Structure, Conclusion
V0.9	21.08.2026	Kārlis Freivalds	Final streamlining and review
V1.0	31.08.2026	Pal Varga	Finalization after review

Publishable Executive Summary

This deliverable, D2.4, presents the final technological results of WP2 in the AIMS5.0 project, with a particular focus on the development, integration, validation, and reuse of Artificial Intelligence methods, algorithms, tools, and industrial applications. A central outcome is the final version of the AIMS5.0 AI Toolbox, conceived as a common framework for accelerating the adoption of AI in industrial environments. Rather than replacing existing AI frameworks, the Toolbox extends and connects available solutions by providing a structured environment for knowledge exchange, reusable tools, deployment support, composability, and methodological guidance. Its aim is to reduce development and integration efforts, facilitate rapid prototyping, support reliable implementation, and make successful AI solutions easier to transfer across use cases and industrial domains.

The AI Toolbox covers several levels of technological granularity, ranging from individual algorithms and reusable tools to complete industrial applications, supported by a methodology layer that defines specifications, best practices, deployment approaches, and implementation guidance. The final implementation includes a catalogue of available tools, deployment mechanisms, supporting libraries, an Application Wiki, and a repository for accessing Toolbox components. The architecture distinguishes between deeply integrated AI services and more modular, shallow integrations, allowing technologies with different levels of maturity and integration requirements to coexist within the same framework.

The technological portfolio assembled within WP2 is broad and reflects the diversity of industrial AI requirements. It includes computer vision and image understanding, language and generative AI, industrial data analytics, anomaly and fault detection, predictive maintenance, robotics and automation, manufacturing simulation, optimization and scheduling, semantic technologies, explainability, factuality evaluation, and trustworthy AI. The Toolbox therefore spans both general-purpose AI capabilities and highly specialized industrial solutions.

A key achievement of the project is that these developments are not treated as isolated research prototypes. Instead, the technologies are systematically linked to the AIMS5.0 industrial use cases and are described according to a common structure covering their technical principles, evaluation, achieved KPIs, contribution to the project objectives, and application scenarios. The Toolbox has already been made available to the use cases, and almost all use cases are evaluating selected methods or tools originating from WP2.

The reported results contribute directly to the project’s industrial performance objectives. These include improvements in prediction and classification performance, reductions in quality-control effort, shorter AI-system reconfiguration times, lower data requirements when adapting models to new machines, increased productivity, improved manufacturing uptime, faster AI development and integration, reduced time-to-market, and lower energy consumption. Taken together, the results of WP2 make an important contribution to European industrial competitiveness. European manufacturing increasingly depends on the ability to introduce advanced AI technologies rapidly, reliably, and at reasonable cost, while maintaining high standards of quality, sustainability, security, explainability, and human oversight.

Overall, AIMS5.0 demonstrates that industrial AI creates the greatest value when algorithms are embedded in reusable engineering processes, interoperable infrastructures, validated industrial applications, and collaborative knowledge networks. The WP2 results establish such a foundation. By transforming heterogeneous AI developments into accessible and transferable industrial capabilities, the project strengthens Europe’s capacity to innovate, supports more sustainable and resilient manufacturing, and helps European Industry5.0 remain competitive in a rapidly evolving global technological landscape.

Table of contents

1. Introduction 6
1.1 Objective and Scope of the Document 6
1.2 Structure of the Deliverable Report 7
2. AI Toolbox – Final Version 8
2.1 The AI Toolbox Concept 8
2.2 Final Implementation Status 10
2.3 AI Toolbox Repository 11
3. Technological Results: AI Algorithms and Tools 18
3.1 Triple-based factual correctness evaluation of AI generated summaries (TU/e) 19
3.2 SPARQL-CAR: A Semantic Error Detection and Repair Framework for Reliable NL2SPARQL (TU/e) 20
3.3 Invertible differentiable batched GPU-accelerated simulation for the laser beam shaping system of EOS M290 PBF-LB-M system (TU/e) 22
3.4 Phase Retrieval algorithm package: FourierGSNet, GS direct unrolling, Gradient Descent fitting method, and Newton-Raphson fitting method (TU/e) 24
3.5 Automatic Ontology Learning Framework (TU/e) 26
3.6 Improved human motions for robot testing IPL devices (PCL, TUE) 27
3.7 Data analysis method (PCL, RUG) 28
3.8 Automatic event and data sharing to orchestrate service delivery (Husqvarna) 28
3.9 AI models development (Husqvarna) 29
3.10 Automatic tool breakage detection in milling operations (GOIMEK, SAVVY, IDEKO, SORALUCE, MULTIVERSE) 30
3.11 Object detection and segmentation (EDI, BME) 32
3.12 Single object 6DoF pose estimation (BME,EDI)) 34
3.13 Grasp affordance estimation (EDI) 36
3.14 Grasp planning and execution (EDI) 38
3.15 Socket detection (EDI) 40
3.16 Robot arm feedback control (EDI, HUA) 42
3.17 AI-based robot control script synthesis (EDI, HUA) 43
3.18 Computer Vision (Signify, DAC) 45
3.19 Intelligent Scheduler (Signify, TU/e) 47
3.20 Improved overall productivity of the manufacturing factory by ~20% Energy-aware scheduling (FUH, IFD, IFAT, SYSTEMA, VIF) 48
3.21 Intelligent Energy management (FUH, IFD, IFAT, SYSTEMA, VIF) 49
3.22 MES Airbus (SYSTEMA, IFD, IFAT, IFAG) 50
3.23 PH and EC analysis (TIAG, RSA, AIT, GUT, PYT) 51
3.24 Edge computing architecture and relevant SW (TIAG, RSA, AIT, GUT, PYT) 51
3.25 Power consumption control (TIAG, RSA, AIT, GUT, PYT) 52
3.26 Visual plant mass estimation (TIAG, RSA, AIT, GUT, PYT) 53
3.27 Wireless Communication Enhancement algorithms for smart antenna (GUT) 53
3.28 ThreatGet (TIAG, RSA, AIT, GUT, PYT) 55
3.29 Rapid modelling (SYSTEMA, IFD, IFAT, IFAG, TUD, HSZG, WUW) 56
3.30 WIP level prediction using ML (SYSTEMA, IFD, IFAT, IFAG, TUD, HSZG, WUW) 57

3.31 AI-based decision support (SYSTEMA, IFD, IFAT, IFAG, TUD, HSZG, WUW) 58
3.32 Integrated transportation and production (SYSTEMA, IFD, IFAT, IFAG, TUD, HSZG, WUW) . 59
3.33 Digital Kanban System (SYSTEMA, IFD, IFAT, IFAG, TUD, HSZG, WUW) 60
3.34 AMHS (HOST, IFD, IFAG, IFAT, FMX, TUD, UzL) 60
3.35 Tool and Person Detection with Classification for Mobile Systems (UzL, IFD) 63
3.36 Virtual Commissioning using a Digital Twin (FMX, IFD, HOST, TUD, IFAT) 64
3.37 Machine Learning, deep Learning (UniPD, IFD; IFAG; IFAT, KAI, STATWOLF, LFOUNDRY) .. 65
3.38 Components RF design acceleration (THALES, FORTH) 68
3.39 RF components fabrication optimization (THALES, FORTH) 69
3.40 New manufacturing paradigm for RF components (THALES, FORTH) 70
3.41 Greener/more sustainable manufacturing for RF components (THALES, FORTH) 71
3.42 Development of RF components capable to support diverse applications (THALES, FORTH) 72
3.43 Sensors and actuators for cleanroom monitoring (FhG-IISB, IFD, IFAG, IFAT, PFEIFFER, SYSTEMA) 72
3.44 Semantic web-based DR (IFAG, FUH, NXP, IFD) 75
3.45 Digital Reference Ecosystem (IFAG, FUH, NXP, IFD) 77
3.46 Open Access Platform (IFAG, FUH, NXP, IFD, DAC) 79
3.47 Production Trace Algorithm (NXP, Semaku, TU/e) 80
3.48 Supply Chain Simulation Model (NXP, Semaku, TU/e) 81
3.49 Linear Programming (LP) Model (NXP, Semaku, TU/e) 81
3.50 Stochastic Optimization Models (NXP, Semaku, TU/e) 82
3.51 Supply Chain Master Planning Heuristic (NXP, Semaku, TU/e) 83
3.52 LLM for Root Cause Analysis (NXP, Semaku, TU/e) 83
3.53 Single & Multi-Stage Quality Control Models (NXP, Semaku, TU/e) 84
3.54 Test Wafer Harmonization (IFAT, IFD, IFAG, TUD) 85
4. Expert Groups – Final Contributions and the Book 87
4.1 Machine Learning Models and Techniques 87
4.2 Image Processing 88
4.3 Anomaly Detection, Fault Detection, Predictive Maintenance 89
4.4 Reinforcement Learning, Optimization, Scheduling 90
4.5 Robotics, Automation, Process Management 91
4.6 Natural Language Processing, Large Language Models 92
5. Conclusions 94

1. Introduction

1.1 Objective and Scope of the Document

This deliverable presents the final results of the AIMS5.0 activities related to AI methods, algorithms, tools, and their integration into the AI Toolbox. Its objective is to consolidate the technological developments achieved within WP2 and to demonstrate how these developments support the project's industrial use cases. The document provides a common reference for the AI solutions developed by the project partners, facilitating knowledge sharing, the reuse of technological results, and collaboration across use cases and industrial domains.

A central part of the deliverable is the final version of the AIMS5.0 AI Toolbox. The AI Toolbox is designed not to replace existing AI frameworks and toolboxes, but rather to extend and connect them through a common platform for industrial AI development and knowledge exchange. It supports a harmonized and structured development process and promotes reusable, composable, and deployable AI solutions. The Toolbox addresses different levels of AI service granularity, ranging from algorithms and reusable tools to complete applications, complemented by a methodology layer providing specifications, best practices, deployment guidance, and implementation support.

This deliverable therefore reports comprehensively on the final implementation status and contents of the AI Toolbox, including its catalogue, deployment mechanisms, supporting libraries, Application Wiki, and repository. It distinguishes between deeply integrated, reusable AI services and more specialized shallow-integrated tools and covers a broad spectrum of capabilities, including computer vision, language and generative AI, data analytics, robotics and automation, manufacturing and simulation, semantic technologies, and trustworthy AI.

Furthermore, the deliverable presents the technological results developed by the project partners. The individual technologies are described according to a common structure covering the technology itself, its evaluation and achieved KPIs, its contribution to the AIMS5.0 success metrics, and its application within the project's use cases and beyond. The results demonstrate how the AI Toolbox elements are being applied and validated in WP5 and WP6 and how they contribute to improvements in prediction and classification performance, quality control, reconfiguration effort, productivity, time-to-market, sustainability, and other industrial objectives.

Finally, D2.4 consolidates the work of the cross-WP2 Expert Groups. Their joint technical activities have resulted in thematic contributions addressing Machine Learning Models and Techniques; Image Processing; Anomaly Detection, Fault Detection and Predictive Maintenance; Reinforcement Learning, Optimization and Scheduling; Robotics, Automation and Process Management; and Natural Language Processing and Large Language Models. These contributions combine state-of-the-art reviews with technological developments and industrial use cases from AIMS5.0 and identify remaining research challenges towards sustainable, resilient, and human-centric Industry 5.0.

1.2 Structure of the Deliverable Report

The deliverable is organized to provide a consolidated view of the final WP2 results, ranging from the implementation of the AIMS5.0 AI Toolbox and the individual technological developments to the joint technical contributions of the WP2 Expert Groups.

Section 2 presents the final version of the AIMS5.0 AI Toolbox. It revisits the Toolbox concept and its different levels of AI service granularity and describes the final implementation status, including the catalogue, deployment tools, Toolbox libraries, Application Wiki, and repository. The section also presents the distinction between deep and shallow integration and summarizes the broad range of AI capabilities available through the Toolbox.

Section 3 presents the technological results developed within AIMS5.0 in the form of AI algorithms, tools, methods, and applications. Each technological result is described using a common structure covering the technology description, evaluation and achieved KPIs, contribution to the project's Objective 2 success metrics, and application scenarios within the AIMS5.0 use cases and beyond. The section therefore provides the main link between the technological developments of WP2, their integration into the AI Toolbox, and their validation in the WP5 and WP6 use cases.

Section 4 summarizes the final contributions of the cross-WP2 Expert Groups and their joint work towards the AIMS5.0 book on Artificial Intelligence in Manufacturing leading to Sustainability and Industry 5.0. The section covers Machine Learning Models and Techniques; Image Processing; Anomaly Detection, Fault Detection, and Predictive Maintenance; Reinforcement Learning, Optimization, and Scheduling; Robotics, Automation, and Process Management; and Natural Language Processing and Large Language Models. The contributions combine reviews of the state of the art, results and industrial use cases developed within AIMS5.0, and the identification of research gaps relevant to Industry 4.0 and Industry 5.0.

Section 5 concludes the deliverable by summarizing the main results of WP2 and the role of the AI Toolbox, technological developments, and Expert Group activities in supporting the wider objectives and industrial use cases of AIMS5.0.

2. AI Toolbox – Final Version

2.1 The AI Toolbox Concept

While a number of AI tools and toolboxes are already available, industrial applications pose new challenges for the design and development of AI solutions. The goal of the AIMS5.0 AI Toolbox is not to replace existing toolboxes, but rather to extend existing solutions by providing a common platform for information and knowledge exchange for industrial AI applications. More precisely, the AIMS5.0 AI Toolbox guides and supports the design and development process through a more harmonized and structured approach, thereby maximizing the usability of AI models at a larger industrial scale.

To support AI service design and implementation, the AIMS5.0 AI Toolbox has the following well-defined objectives:

- Decrease the time-to-market: By providing design templates and examples, as well as reusable and compatible tools, the Toolbox helps reduce time-to-market through rapid prototyping;
- Reliability and security: By describing best practices, the Toolbox helps designers improve the reliability and security of AI solutions;
- Rapid evaluation: To support the testing of industrial proofs of concept, the Toolbox provides tools that can be easily evaluated;
- Composition: The AI Toolbox provides tools that can interoperate through standard interfaces and be composed to implement complex applications. Composition helps improve the reusability of AI tools;
- Knowledge sharing: The toolbox provides means to exchange and share valuable knowledge and best practices;
- Extensive support: The AIMS5.0 AI Toolbox supports natively any existing AI tools and toolboxes.

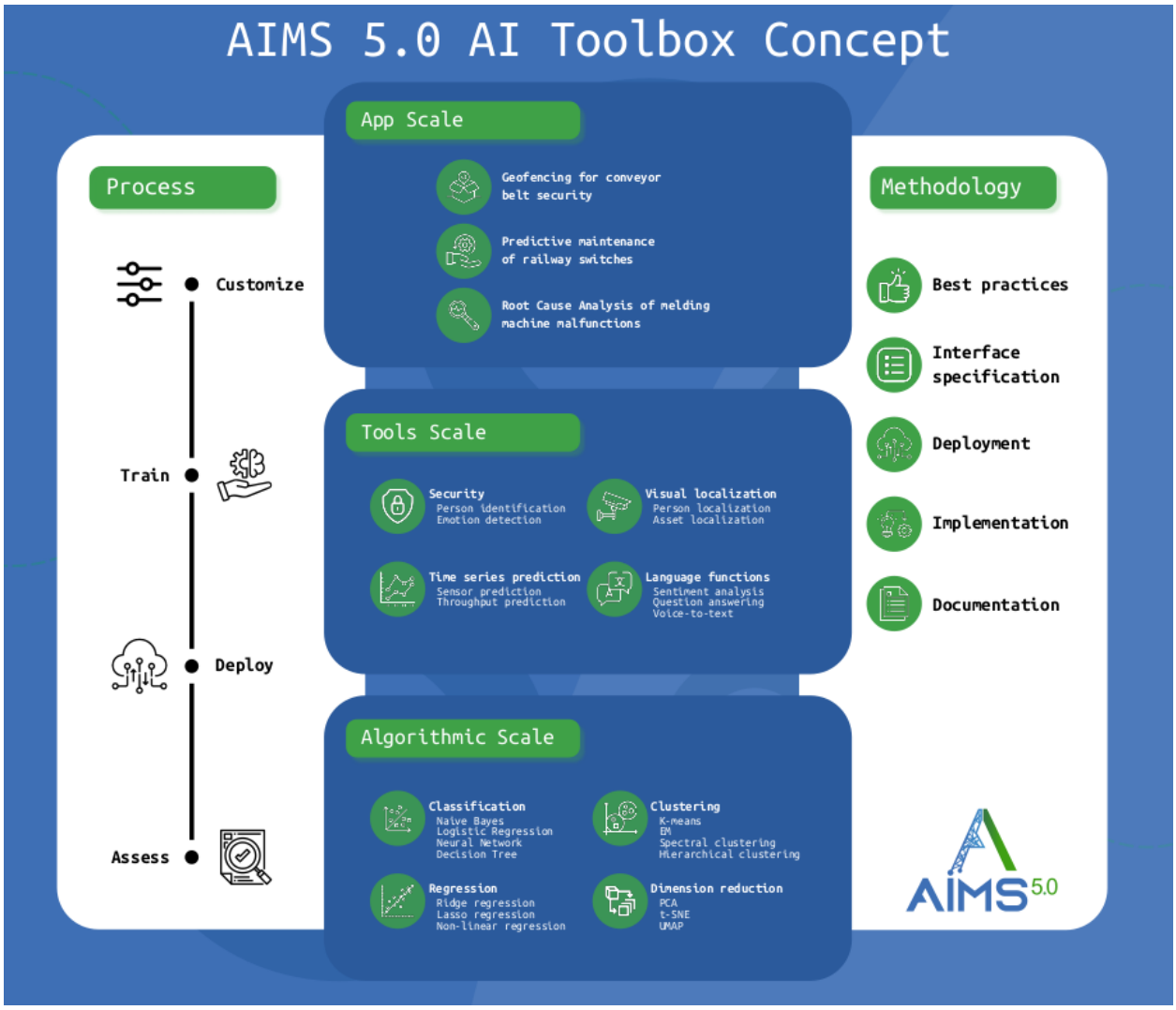


Figure 1. AIMS5.0 AI Toolbox concept with different levels of AI service granularity, supported by a common process and methodology. The examples shown at the different levels are for illustrative purposes only; many other categories and examples are possible.

Figure 1 shows the main concept of the AIMS5.0 AI Toolbox. There are four levels of AI service granularity, which differ primarily in terms of granularity. The three implementation-based levels are supported by a fourth, predominantly theoretical methodology layer. The following sections present the four levels in greater detail and compare them with the objectives and requirements.

2.1.1 Application scale

Application scale represents the highest level of AI service granularity, where AI services provide solutions for complete use cases, such as a predictive maintenance service specifically designed for railway switches. The aim of the AIMS5.0 AI Toolbox is to reach the application scale through the composition of reusable and composable tools.

2.1.2 Tools scale

Tools scale represents a compromise between the algorithmic scale and the application scale. Tools essentially provide applied AI methods, which commonly consist of one or more AI models combined with internal logic to solve recurring problems. For example, asset localization in a

factory may consist of a few small steps, but the same problem can arise across numerous use cases. The tool scale provides a high degree of reusability and supports rapid development. Furthermore, careful design of service interfaces can enable good composability.

2.1.3 **Algorithmic scale**

The Algorithmic scale is highly granular, as well-known AI and ML algorithms are implemented as AI services. At the algorithmic scale, the focus is primarily on designing appropriate service interfaces to enable the composition of services. However, using individual algorithms, while offering a high degree of reusability, requires considerable effort to implement complex use cases. Furthermore, using individual AI algorithms requires substantial knowledge of AI methods and techniques, which can hinder rapid development.

2.1.4 **Methodology scale**

As a theoretical layer, the methodology provides specifications, requirements, and best practices for implementing AI services. Although the Methodology scale can be considered an independent level within the AI service hierarchy, it is best viewed as the foundation for all other service levels, as it provides the means to fulfil the objectives and requirements of the AI Toolbox services. Consequently, an appropriate methodology should be provided at all service levels to ensure integrated and unified interfaces, facilitate adaptation to heterogeneous environments, and satisfy composability requirements when implementing AI services.

Although this hierarchical approach remains valid from a practical perspective, a flat Toolbox concept is recommended. As shown in the previous deliverable on the AI Toolbox, the most promising implementation level is the tools scale. While implementing AI applications using tools requires only limited knowledge of basic AI methods, appropriate service interfaces can be designed to satisfy composability requirements.

2.2 Final Implementation Status

- Catalogue: Applications consist of tools that are chained together. The catalogue contains a list of the available tools included in the AI Toolbox, together with information on their usage, categorized by field of application;
- Deployment tools: Prompt evaluation of AI tools requires mechanisms for seamlessly deploying tools as services. The primary method for deploying these notebooks as services involves containerized services with well-defined interfaces. However, this approach can be extended to support various deployment scenarios for cutting-edge industrial frameworks, such as Arrowhead. To deploy implemented tools as services, the AI Toolbox requires deployment tools that utilize the deployment configuration specified by the tools;
- Toolbox libraries: An essential requirement for aligning with the philosophy of the AI Toolbox is the ability to deploy these notebooks across various architectures without the need to rewrite the code. These libraries handle interface compatibility and management for the tools, as responsibility for managing interface integration lies with the AI Toolbox;
- Application Wiki: AI Toolbox users require a comprehensive guide to its use. This Wiki provides all essential information on implementing specific applications using integrated tools, as well as guidance on best practices and contributing to the AI Toolbox.

2.3 AI Toolbox Repository

https://aitoolbox.tmit.bme.hu/Tools

Integration model

- **Deep integration** - tightly integrated into the toolbox infrastructure
- **Shallow integration** - lighter or more modular integration

Deep AI Toolbox Tools

The **Deep AI Toolbox** layer contains tools intended for **deeper integration, orchestration, and reusable AI services.** These tools focus less on a single specialized AI task and more on providing general-purpose AI capabilities that can be integrated into broader industrial AI workflows.

Deep AI Toolbox Tools

The **Deep AI Toolbox** layer contains tools intended for **deeper integration, orchestration, and reusable AI services**. These tools focus less on a single specialized AI task and more on providing general-purpose AI capabilities that can be integrated into broader industrial AI workflows.

Tool	Summary	Main Capability
LLM Tool	Provides Large Language Models as deployable services using models available through the Hugging Face ecosystem.	Generative AI, reasoning, NLP
Object Detection Tool	Provides object detection capabilities based on Ultralytics models and can serve as a reusable detection service.	Computer vision
Demo Tool	Demonstrates how a tool can be deeply integrated into the AI Toolbox architecture.	Integration and demonstration
Prompt Context Injection Tool	Adds external or task-specific context to NLP prompts to improve the relevance and quality of LLM responses.	LLM orchestration, NLP

Deep AI Layer – Overall Summary

The **Deep AI layer** can therefore be viewed as the **AI service and integration layer** of the Toolbox. It provides reusable AI capabilities that can be incorporated into larger workflows, applications, and industrial AI systems. Its key characteristics are:

- Service-oriented AI capabilities
- LLM and generative AI integration
- Reusable computer-vision services
- Context-aware AI interaction
- Demonstration of deep Toolbox integration
- Support for composing AI capabilities into larger solutions

In an industrial architecture, these tools can act as **higher-level AI services** that consume or combine capabilities provided by specialized shallow tools.

2. Shallow AI Toolbox Tools

The **Shallow AI Toolbox** contains a much broader collection of **specialized, task-oriented AI tools**. These tools generally address specific problems related to perception, prediction, optimization, robotics, document processing, language processing, anomaly detection, or explainability.

A. Computer Vision and Image Understanding

Tool	Summary
6DoF Pose Estimation Tool	Estimates the three-dimensional position and orientation of objects, supporting robotic manipulation and industrial automation.
Object Detection with YOLOv8	Performs fast real-time object detection using the YOLOv8 architecture.
BiRefNet Image Segmentation	Provides high-accuracy pixel-level image segmentation for separating objects or regions from images.
Zero-Shot Object Detection with OWLv2	Detects objects without requiring conventional task-specific training, using open-vocabulary/zero-shot detection.
MobileSAM	Provides lightweight image segmentation using a mobile-oriented implementation of the Segment Anything approach.
SAM 3	Performs segmentation and tracking of object instances in images and videos using text or image prompts.
RT-DETR	Provides real-time transformer-based object detection with an end-to-end architecture.
FoundationPose	Provides unified 6D pose estimation and tracking for previously unseen objects.
Image Classification Tool	Assigns predefined category labels to complete images without explicitly locating individual objects.
Multi-Object Tracking Tool	Tracks multiple objects over consecutive video frames for applications such as surveillance, robotics, and industrial monitoring.
Human Pose Estimation Tool	Extracts human key points and skeletal structure for movement and interaction analysis.
Socket Detection Tool	Detects socket positions on a conveyor using a YOLOv8m model, supporting automated industrial inspection or assembly.
Multimodal VLM Tool	Combines visual and textual information for image understanding, image captioning, visual question answering, and multimodal reasoning.

Overall: These tools form the **visual perception layer**, supporting tasks such as detection, segmentation, classification, tracking, pose estimation, and multimodal visual reasoning.

B. Language, Document and Generative AI

Tool	Summary
DeepSeek-R1	Provides a deployment-ready reasoning-oriented LLM for advanced language and reasoning tasks.
Qwen3.5	Provides a multimodal foundation model combining language, visual understanding, and advanced reasoning capabilities.
GPT-OSS 120B	Provides an open-weight large language model oriented toward reasoning, agentic applications, and developer use cases.
DeepSeek-OCR	Provides OCR capabilities with an emphasis on efficient/contextual processing of visual information.
LayoutLMv3	Performs document understanding by combining textual, visual, and layout information.
Embedding Tool	Generates vector representations of text for semantic search, retrieval, clustering, and similarity analysis.
Reranker Tool	Reorders retrieved documents according to their relevance, improving search and RAG performance.
Speech-to-Text Tool	Converts multilingual speech into text and supports speech translation.
Text-to-Speech Tool	Generates natural multilingual speech from text and supports voice-cloning applications.
Document Parsing Tool	Extracts structured information, tables, layouts, and content from documents and PDFs.
Factual Correctness Evaluation Tool	Evaluates whether AI-generated summaries are factually consistent with structured source information.
Automatic Ontology Learning Framework	Automatically learns domain ontologies from corpora or structured databases.
NL2SPARQL Semantic Error Detection Framework	Evaluates and detects semantic errors in SPARQL queries generated from natural-language requests.
ZebraLogicReasoner	Provides a reproducible environment for evaluating reasoning capabilities on structured logical puzzles.

Overall: This group forms the **language and knowledge-processing layer**, supporting LLMs, OCR, document intelligence, retrieval-augmented generation, speech processing, ontology construction, semantic reasoning, and factuality evaluation.

C. Tabular Data, Time Series and Anomaly Detection

Tool	Summary
Tabular Foundation Models (TabPFN)	Applies foundation-model approaches to structured tabular datasets for prediction.
Lag-Llama	Provides foundation-model-based forecasting for time-series data.
Tabular Prediction Tool	Performs classification and regression on structured tabular datasets.
Time-Series Anomaly Detection Tool	Identifies unusual patterns and deviations in temporal data such as sensor and operational signals.
DIFFI	Provides interpretable anomaly detection and feature-importance analysis using Depth-based Isolation Forest methods.
ExIFFI	Provides interpretable anomaly detection and feature importance using Extended Isolation Forest techniques.
Device-Level Quality Detection Tool	Detects device-level defects and connects machine-signal features with potential root causes.
FFT Feature Extraction for Tool Breakage Detection	Uses frequency-domain analysis of vibration signals to identify tool breakage during milling operations.

Overall: These tools constitute the **industrial data analytics layer**, covering prediction, forecasting, anomaly detection, feature-importance, quality monitoring, and machine-condition analysis.

D. Robotics, Motion and Industrial Automation

Tool	Summary
AMR Travel Time Prediction	Predicts autonomous mobile robot travel times from fleet and operational log data.
NLP Programming of Industrial Robots	Converts natural-language descriptions of robot tasks into tested and deployable robot-control programs.
Differentiable-Physics Motion Planning	Uses differentiable physics simulation to learn smooth and accurate robot trajectories.
Grasp Pose Estimation Tool	Generates 6DoF gripper poses for robotic pick-and-place operations from sensor information.
Motion Playground	Provides real-time modeling and visualization of motion data.
FoundationPose	Supports object pose estimation and tracking needed for robotic perception and manipulation.

6DoF Pose Estimation Tool	Estimates object position and orientation for robotic applications.
Socket Detection Tool	Provides spatial detection of sockets on conveyor systems for automated industrial operations.

Overall: These tools form the **robotics and physical-AI layer**, connecting perception and machine learning with robot manipulation, motion planning, programming, and autonomous systems.

E. Manufacturing, Simulation and Scientific AI

Tool	Summary
Phase Retrieval Algorithm Package	Provides several Fourier and optimization-based approaches for phase-retrieval problems, including unrolling, gradient descent, and Newton-Raphson methods.
Invertible Differentiable Simulation	Provides GPU-accelerated differentiable simulation for laser-beam shaping in additive manufacturing.
AE-PSL	Uses autoencoder-compressed parallel split learning for fine-tuning pretrained models while supporting distributed learning scenarios.
GAP-FL	Provides group-aware fairness analysis for differentially private federated learning.
Differentiable-Physics Motion Planning	Combines machine learning and physics simulation for industrial robot trajectory learning.

Overall: These tools extend the Toolbox into areas of **advanced industrial AI, simulation, distributed learning, privacy, fairness, and physics-informed optimization**.

F. Explainability, Evaluation and Trustworthy AI

Tool	Summary
Explainability Tool	Provides model interpretation for tabular, computer-vision, and deep-learning models.
Factual Correctness Evaluation	Evaluates factual consistency of AI-generated summaries.
NL2SPARQL Semantic Error Detection	Evaluates semantic correctness of AI-generated SPARQL queries.
GAP-FL	Addresses fairness analysis in privacy-preserving federated learning.

ZebraLogicReasoner	Supports reproducible evaluation of structured reasoning capabilities.
Automatic Ontology Learning Framework	Supports structured knowledge acquisition and ontology construction.

Overall: This group supports the **trust, transparency, evaluation, and governance layer** of the AI Toolbox.

3. Overall Catalogue Summary

The AI Toolbox currently demonstrates a broad spectrum of AI capabilities, ranging from **fundamental AI models and specialized algorithms to integrated AI services**.

Category	Main Role	Representative Capabilities
Deep AI	Higher-level reusable AI services and integration	LLM services, object detection services, prompt/context management, integration
Computer Vision	Machine perception	Detection, segmentation, classification, tracking, pose estimation
Language & Generative AI	Language and knowledge processing	LLMs, embeddings, OCR, document AI, speech, RAG
Data Analytics	Industrial data intelligence	Prediction, forecasting, anomaly detection, feature importance
Robotics	Physical AI and automation	Grasping, motion planning, robot programming, AMR prediction
Manufacturing & Simulation	Physics-informed industrial AI	Differentiable simulation, phase retrieval, manufacturing optimization
Knowledge & Semantic AI	Structured knowledge processing	Ontology learning, RDF/SPARQL processing, reasoning
Trustworthy AI	Reliability and governance	Explainability, factuality, fairness, evaluation

The AI tool Box‘s Categories and Algorithms:

The AI tool Box‘s repository of reusable AI services designed to facilitate the integration of heterogeneous AI capabilities into software systems. Its tool selection illustrates an emphasis on both general-purpose AI services, such as LLMs and object detection, and domain-specific capabilities, such as robotic pose estimation. The catalogue also suggests an architectural distinction between different levels of integration, which may be relevant when evaluating the toolbox's extensibility, interoperability, and ease of deployment:

1. Algorithmic scale

- Classification
  - Naive Bayes
  - Logistic Regression
  - Neural Network
  - Decision Tree

- Clustering
  - K-means
  - EM
  - Spectral clustering
  - Hierarchical clustering
- Regression
  - Ridge regression
  - Lasso regression
  - Non-linear regression
- Dimension reduction
  - PCA
  - t-SNE
  - UMAP

2. Tools scale

- Emotion detection
- Time-series prediction
- Sensor prediction
- Throughput prediction
- Language functions
  - Sentiment analysis
  - Question answering
  - Voice-to-text
- Visual localization
  - Person localization
  - Asset localization

3. Application scale

- Geofencing for conveyor-belt security
- Predictive maintenance of railway switches
- Root-cause analysis of welding-machine malfunctions

4. Methodology scale

- Best practices
- Interface specification
- Deployment
- Implementation
- Documentation
- Train
- Customize
- Deploy
- Assess

3. Technological Results: AI Algorithms and Tools

This section presents the various AI tools, methods, and applications developed within the project to support sustainable industrial processes. Each tool is described using a structured approach that includes its introduction, application scenario, alignment with the project's Objective 2 success metrics, target requirements and constraints, baseline definition, target-level measurement, and intermediate validation results at the end of Period 2.

One of the main objectives of WP2 is to create the AI Toolbox, a standardized platform that brings together AI-related tools and applications from multiple industrial domains to foster knowledge sharing and the exchange of best practices. The AI Toolbox integrates and extends diverse AI solutions to support human-centred digitalization and sustainable industrial transformation.

The AI Toolbox has already been made available to the use cases. Depending on the use case and its purpose, selected AI methods from the Toolbox are utilized, implemented, and evaluated within the application scenarios of the respective use cases. The AI Toolbox methods used in the AIMS5.0 use cases are summarized in the table below. It is important to note that almost all use cases are evaluating tools from the Toolbox (see **Table 1**).

Table 1: Addressed success metrics by each technology

Technology Result	UC	20% prediction	25% classification	30% QC time ↓	30–50% reconfig ↓	50% cost ↓	20% productivity ↑	>10% AI MES	20% uptime ↑	20% data ↓	>15% self-adj	Sustainability docs	70% ML time ↓	30% faster TTM	>10% energy ↓
Triple-based factual correctness evaluation				✓			✓								
Semantic Error Detection and Repair		✓			✓									✓	
Invertible GPU-accelerated simulation for laser beam shaping					✓		✓				✓			✓	
Phase Retrieval algorithm package			✓	✓											
Automatic Ontology Learning Framework			✓												
Improved human motions for robot testing IPL devices	UC2					✓	✓								
Data analysis method	UC2					✓	✓								
Automatic event & data sharing	UC3						✓	✓		✓	✓	✓		✓	✓
AI models development	UC3						✓	✓		✓	✓	✓		✓	✓
Automatic tool breakage detection	UC5		✓	✓	✓					✓	✓				
Object detection & segmentation	UC6	✓	✓												
Single-object 6DoF pose estimation	UC6	✓	✓		✓		✓				✓				
Grasp affordance estimation	UC6						✓				✓				
Grasp planning & execution	UC6						✓								
Socket detection	UC6	✓	✓		✓										
Robot arm feedback control	UC6	✓					✓				✓				
AI-based robot control script synthesis	UC6				✓						✓		✓		
Computer Vision	UC7	✓	✓	✓	✓	✓	✓	✓		✓		✓	✓	✓	
Intelligent Scheduler	UC7			✓											
Energy-aware scheduling	UC9											✓			✓
Intelligent Energy management	UC9										✓	✓			✓
MES Airbus	UC10						✓	✓			✓				
PH & EC analysis	UC11				✓							✓			✓
Edge computing architecture & software	UC11				✓							✓			✓
Power consumption control	UC11				✓							✓			✓
Visual plant mass estimation	UC11				✓							✓			✓
Wireless communication enhancement	UC11				✓										
ThreatGet	UC11				✓							✓			
Rapid modelling	UC12						✓								
WIP level prediction using ML	UC12						✓								
AI-based decision support	UC12						✓								
Integrated transportation & production	UC12						✓								
Digital Kanban System	UC12						✓								
Automated Material Handling Systems	UC13	✓							✓						
Tool & Person Detection for Mobile Systems	UC13	✓	✓		✓							✓			✓
Virtual Commissioning using Digital Twin	UC14					✓	✓		✓					✓	
Machine Learning & Deep Learning	UC15				✓					✓			✓	✓	
Components RF design acceleration	UC16						✓					✓			
RF components fabrication optimization	UC16						✓					✓			
New manufacturing paradigm for RF components	UC16						✓					✓			
Greener/more sustainable RF manufacturing	UC16						✓					✓			
RF components supporting diverse apps	UC16						✓					✓			
Sensors & actuators for cleanroom monitoring	UC17		✓	✓											
Semantic web-based Digital Reference	UC18														
Digital Reference Ecosystem	UC18														
Open Access Platform	UC18				✓		✓					✓	✓		
Production Trace Algorithm	UC19			✓		✓	✓								
Supply Chain Simulation Model	UC19			✓		✓	✓								
Linear Programming Model	UC19			✓		✓	✓								
Stochastic Optimization Models	UC19			✓		✓	✓								
Supply Chain Master Planning Heuristic	UC19			✓		✓	✓								
LLM for Root Cause Analysis	UC19			✓		✓	✓								
Single & Multi-Stage Quality Control Models	UC19			✓		✓	✓								
Test Wafer Harmonization	UC20					✓	✓								

The technological results of the AI Toolbox contribute to the following success metrics (see Annex 1 to the AIMS5.0 Grant Agreement, Part B, Table 1, Objective 2). These metrics are

addressed through the tools and applications developed within the project and integrated into the AI Toolbox. This section presents these tools and applications, highlighting how each contributes to achieving the defined success metrics (see the overview in **Table 1** above):

- Building a structured Toolbox repository containing more than 50 AI methods, tools, and algorithms that are used and integrated by WP5 and WP6 represents a novelty, as no such Toolbox is currently available for industrial processes and supply chains;
- This structured Toolbox is indexed according to the various stages and levels of engineering processes;
- Providing documented methods for the dynamic application of tools to improve sustainability;
- The following Objective 2 success metrics will be achieved within the frame of the AI toolbox evaluation in the specific use cases. Their achievement will be enabled through the AI Toolbox; however, validation is not possible at the WP2 level and requires validation within the WP5 and WP6 use cases;
- Industrial process phases made more effective and precise through the integration of tailored AI algorithms, methods, and tools:
  - 20% improvement in prediction precision;
  - 25% improvement in classification precision;
  - 30% improvement in time spent on quality control.
- Reduced reconfiguration time of the AI-based system by ~30-50%;
- Reduced manufacturing costs in the long term (3-5 years) by ~50%;
- Improved overall productivity of the manufacturing facility by ~20%;
- AI based MES capability > 10 %;
- Improvement uptimes of wafer transportation and storage systems by 20%;
- 20% reduction of required data for scaling new ML models to new machines;
- Self-adjusted -/ adapted manufacturing processes > 15 %;
- 70% reduction in the time required to develop and integrate an ML model in the digitalized industrial domain;
- Improved performance and to show an overall 30% faster time-to-market;
- Decreased energy consumption for PoC > 10%;
- Targeted guideline documents for integration of use-case specific methods.

3.1 Triple-based factual correctness evaluation of AI generated summaries (TU/e)

3.1.1 Technology description

The triple-based framework is designed to evaluate the factual correctness of AI-generated summaries. In this approach, a Large Language Model (LLM) is provided with both the source text and the AI-generated summary and tasked with extracting multiple triple representations of factual information from each. Each triple is a structured representation of a fact in the form of *(subject, predicate, object)*. The extracted triples are then compared, and factuality is assessed based on their intersection.

This framework can also be applied in manufacturing contexts, where large volumes of production logs are summarized using LLMs. By evaluating the factuality of these summaries, the framework helps reduce the need for manual verification, thereby improving the adaptability of AI-driven solutions and accelerating decision-making processes.

3.1.2 Evaluation

Committed KPI	Achieved KPI	Validation methodology
Verification of AI-generated summaries with ≥80% accuracy and ≥25% reduction in verification time.	Achieved at the research-prototype level: the framework reached a correlation of up to 0.98 with human factuality judgments on an open-source benchmark.	Compare the framework's factuality predictions with human judgments. Test multiple natural-language questions to assess the robustness of the LLM. Measure the time required to retrieve the data.

3.1.3 Reference to project objective 2 success metrics

- 30% improvement in time spent on quality control;
- Improved overall productivity of the manufacturing factory by ~20%;
- AI Toolbox integration: shallow integration.

3.1.4 Application scenario in this project and beyond

A possible application scenario is a root cause analysis problem in which an LLM generates possible causes, and the framework verifies each claim against manufacturing logs. This results in a verified summary containing a concise list of supported findings, thereby reducing the need for manual verification and accelerating decision-making.

Applicable to large volumes of text (e.g., logs, reports, and documents) requiring extensive manual review. An LLM generates a brief summary, while the framework verifies each claim against the original text, yielding concise and trusted summaries and reducing manual effort.

Currently, no use-cases are using this framework. Use-case partners are welcome to cooperate.

3.1.5 Reference to publication:

Latipov, I. A., Holenderski, M., & Meratnia, N. (2025). Triple-based Factual Correctness Evaluation of AI-Generated Summaries. *IFAC-PapersOnLine*, *59*(27), 226-231.

3.2 SPARQL-CAR: A Semantic Error Detection and Repair Framework for Reliable NL2SPARQL (TU/e)

3.2.1 Technology description

The NL2SPARQL semantic error detection and repair framework supports reliable natural-language access to knowledge graphs. It targets systems that translate natural-language questions into SPARQL queries and verifies whether the generated query preserves the meaning of the original question. This is important because generated SPARQL queries may be syntactically valid and executable while still retrieving misleading or semantically incorrect results from the knowledge graph.

The contribution consists of two components. The first component is the SPARQL semantic error taxonomy and benchmark. It semantic errors in generated SPARQL queries using an Abstract Syntax Tree (AST)-based comparison between a reference query and a candidate query. Rather than treating every failed query simply as "wrong," the taxonomy assigns fine-

grained semantic error labels, such as wrong IRI, missing filter, wrong variable, triple flip, or an extra graph pattern. This makes query failures easier to diagnose, compare, and explain.

The second component is SPARQL-CAR, a modular semantic checking and repair framework for NL2SPARQL generation. SPARQL-CAR combines an LLM-based query generator, a SPARQL engine-based syntax checker, and an LLM-based semantic judge. The syntax check ensures that generated queries are executable, while the semantic check evaluates whether the executable query accurately corresponds to the user's intent. If an error is detected, the framework can provide feedback and trigger a repair step. This transforms NL2SPARQL from a one-shot generation task into a controlled validation and repair pipeline.

3.2.2 Evaluation

Committed KPI	Achieved KPI	Validation methodology
Support the root cause analysis component by improving the reliability of natural-language access to knowledge graphs, in line with the use-case targets of ≥80% accuracy in retrieving relevant knowledge graph insights and ≥50% reduction in engineering analysis time.	Achieved at the research-prototype level: SPARQL-CAR improved Text-to-SPARQL translation quality through syntax and semantic checking and repair, achieving an F1 score of approximately 0.67 on QALD-9 in an open-source benchmark evaluation. Industrial validation on NXP traceability data is pending access to the data and definition of the target questions.	The framework is validated by comparing the predicted semantic error labels with the corresponding gold-standard labels for benchmark queries containing controlled semantic errors.

3.2.3 Reference to project objective 2 success metrics

- Committed KPI: Support Objective 2 by improving the reliability of AI-based root cause analysis and knowledge access through more robust natural-language querying over knowledge graphs;
- Contribution to Objective 2 metrics: The framework contributes to the use-case targets of ≥80% accuracy in retrieving relevant KG insights and ≥50% reduction in engineering analysis time by reducing the risk of executing misleading SPARQL queries. In UC19, this supports faster and more reliable root cause analysis over traceability data;
- AI Toolbox integration: Shallow integration. The semantic error taxonomy, benchmark generation scripts, semantic detection pipeline, and SPARQL-CAR check-and-repair components can be provided as reusable artefacts for validating LLM-based knowledge graph querying.

3.2.4 Application scenario in this project and beyond

The framework is developed for the UC19 traceability use case, where engineers may ask natural-language questions over manufacturing and traceability knowledge graphs. SPARQL-CAR validates and repairs generated SPARQL queries before execution, reducing the risk that a technically valid but semantically incorrect query returns misleading evidence.

Together with Semaku and NXP, we discussed the potential application of the taxonomy and SPARQL-CAR framework to traceability-related questions. The next step is to clarify data access and obtain input from NXP on relevant traceability questions and available gold-standard queries, enabling a concrete UC19 pilot using NXP data.

Beyond UC19, the framework can support any industrial setting where natural-language interfaces are used to query knowledge graphs, including root cause analysis, production traceability queries, quality investigations, and engineering decision support.

3.2.5 **Reference to publication:**

“NL2SPARQL-ErrorBench: A Benchmark and Taxonomy for Semantic Error Detection in SPARQL Queries“ paper is submitted to ISWC2026.

“SPARQL-CAR: A Semantic Check-and-Repair Framework for Reliable Text-to-SPARQL Generation“ paper is submitted to ESWC2026.

3.3 Invertible differentiable batched GPU-accelerated simulation for the laser beam shaping system of EOS M290 PBF-LB-M system (TU/e)

3.3.1 **Technology description**

This simulation computes the aberrated beam shape, i.e., the distorted intensity profile, of a focused Gaussian laser beam in a caustic plane within a certain range around the focal point of the focusing lens, given an aberrated phase map of a Spatial Light Modulator (SLM) located near the light source. The simulation models the phase–intensity response of the high-power laser in the M290 Laser-based Powder Bed Fusion of Metals (PBF-LB/M) system developed by EOS (Electro Optical Systems) GmbH, using analytical wave-optics models. The models used for all optical components in the system are invertible. The simulation is implemented in PyTorch as a neural network class, making it differentiable, and supports the parallel simulation of batches of samples on GPUs.

This simulation can be useful for the following purposes in the development of the EOS M290 PBF-LB/M system or similar additive manufacturing machines:

- Fast, large-scale simulations for the creation of large datasets;
- System parameter fitting;
- Optical aberration identification and quantification;
- Computations requiring inverse simulation from the caustic plane back to the SLM plane, e.g., using Gerchberg–Saxton algorithms.

3.3.2 **Evaluation**

Committed KPI	Achieved KPI	Validation methodology
End-to-end invertibility.	End-to-end invertible.	Run the forward simulation with a given SLM phase map to obtain the complex-valued light field in the caustic plane. Use the resulting light field as the input to the inverse simulation to determine whether the inverse simulation can recover the

		original SLM phase map and the intensity profile of the light source (the single-mode Gaussian CO_2 laser beam from the nLight AFX-1000 fiber laser).
Differentiability with respect to certain system parameters.	Differentiable with respect to Zernike coefficients that parameterize the phase aberration of SLM.	Implement a fitting algorithm based on this simulation that parameterizes SLM phase aberrations using Zernike coefficients and fits these coefficients to produce a distorted beam shape that matches a given target. Successful fitting with high reconstruction accuracy would demonstrate the differentiability of the simulation with respect to the Zernike coefficients.
GPU acceleration: 100 times faster than single thread simulation on CPU.	Simulating 10,000 samples on an NVIDIA GeForce RTX 4090 GPU with a batch size of 250 and a CUDA memory usage of 18,254 MB is 200 times faster than simulating 10,000 samples in a single thread on a 28-core CPU. Simulating 10,000 samples on an NVIDIA GeForce RTX 4090 GPU with a batch size of 1 and a CUDA memory usage of 586 MB is 100 times faster than simulating 10,000 samples in a single thread on a 28-core CPU.	Measure the simulation runtime while varying only the computational device (CPU/GPU) and batch size, keeping all other factors unchanged.

3.3.3 **Reference to project objective 2 success metrics**

AI Toolbox integration: shallow integration.

3.3.4 **Application scenario in this project and beyond**

A possible application scenario is the EOS M290 PBF-LB/M system, where the intensity profile of the laser, also referred to as the beam shape, is manipulated using a computer-generated holography (CGH) phase mask loaded onto a Spatial Light Modulator (SLM) with aberrations. The SLM modulates the phase of the laser reflected from its surface by controlling the voltage applied to each pixel of a liquid-crystal array, thereby tilting the crystals at different angles and introducing the desired phase delays at each pixel. However, the laser input to the SLM can have a power exceeding 500 W. The resulting heat causes phase deviations, i.e., aberrations, in the SLM, thereby distorting the produced beam shape. Distorted beam shapes caused by heat-induced aberrations in the SLM can reduce the efficiency and quality of the PBF-LB/M process. This simulation lays the foundation for acquiring large datasets for training deep learning models and developing algorithms for beam-shape correction, with the aim of improving the efficiency and quality of the PBF-LB/M process.

Currently, the following algorithms from the AIMS5.0 AI Toolbox—*Phase Retrieval Algorithm Package: FourierGSNet, GS Direct Unrolling, Gradient Descent Fitting Method, and Newton-Raphson Fitting Method*—use this simulation. Other use-case partners are welcome to collaborate.

3.4 Phase Retrieval algorithm package: FourierGSNet, GS direct unrolling, Gradient Descent fitting method, and Newton-Raphson fitting method (TU/e)

3.4.1 Technology description

This algorithm package consists of four algorithms for phase retrieval:

- FourierGSNet: A physics- and data-driven hybrid phase retrieval method for Coherent Diffractive Imaging (CDI), X-ray imaging, and laser beam shaping for the M290 PBF-LB/M system from EOS (Electro Optical Systems) GmbH;
- GS direct unrolling: A physics- and data-driven hybrid phase retrieval method for CDI, X-ray imaging, and laser beam shaping;
- Gradient descent method: A physics-driven phase aberration fitting method for laser beam shaping;
- Newton-Raphson method: A physics-driven phase aberration fitting method for laser beam shaping.

Phase retrieval is a well-known and important inverse problem in optics that aims to recover the phase of light from its intensity profile, i.e., an image measured by a camera, after the light propagates through an optical system. Phase retrieval is a notoriously ill-posed problem for which no generic analytical solution exists. Therefore, we developed two physics- and data-driven hybrid methods for high-accuracy phase retrieval. GS Direct Unrolling is a state-of-the-art method that provides high accuracy but has a very slow inference speed when applied to optical systems with complex forward paths. Therefore, FourierGSNet was developed based on GS Direct Unrolling to improve inference speed for complex optical systems, such as the laser beam-shaping system of the M290 PBF-LB/M machine, while maintaining high accuracy. We also developed two physics-driven fitting methods for the beam-shaping system. Although these methods are even slower than GS Direct Unrolling, they do not require training to achieve good accuracy.

3.4.2 Evaluation

Committed KPI	Achieved KPI	Validation methodology
FourierGSNet: Inference speed is 2× faster than GS Direct Unrolling on the X-ray imaging system and 5× faster on the M290 PBF-LB/M system.	Inference is 1.85× faster than GS Direct Unrolling on the X-ray imaging system and 4.5× faster on the M290 PBF-LB/M system.	Measure the single-sample inference time averaged over 3,000 test samples for different phase retrieval methods, while keeping all other factors unchanged.
FourierGSNet: Phase retrieval accuracy shows only a minor decrease compared with GS Direct Unrolling across all three applications.	Accuracy achieved as a percentage of the accuracy achieved by GS Direct Unrolling: ▪ App1: Coherent diffractive microscopy imaging: o MAE:101.4%;	Compute the spatial-domain metrics MAE and SSIM and the frequency-domain metric FRCM, averaged over the

	○ SSIM: 101.1%; ○ FRCM: 100%. ▪ App2: Near-field X-ray imaging: ○ MAE: 100%; ○ SSIM: 99.6%; ○ FRCM: 114.3%. ▪ App3: Laser beam shaping for EOS M290 PBF-LB/M system. ▪ MAE: ○ Chair: 108.2%; ○ Tear: 102.8%; ○ RecTophat: 108.3%; ○ Tophat: 150.5%; ○ Ring: 50%; ○ Gaussian: 149.7%; ▪ SSIM: ○ Chair: 101.1%; ○ Tear: 100.8%; ○ RecTophat: 100.7%; ○ Tophat: 115%; ○ Ring: 96.5%; ○ Gaussian: 134.8%. ▪ FRCM: ○ Chair: 72.4%; ○ Tear: 66.7%; ○ RecTophat: 47.9%; ○ Tophat: 162.6%; ○ Ring: 12.5%; ○ Gaussian: 179.9%. ▪ ReconsErr: ○ Chair: 102.6%; ○ Tear: 109.6%; ○ RecTophat: 124.5%; ○ Tophat: 164.5%; ○ Ring: 59.2%; ○ Gaussian: 148.4%.	respective test sets for all three applications. Compute the beam-shape reconstruction accuracy metric ReconsErr, averaged over the test set for the beam-shaping system.

3.4.3 **Reference to project objective 2 success metrics**

AI Toolbox integration: shallow integration.

3.4.4 **Application scenario in this project and beyond**

These algorithms can be used for:

- Object phase recovery in coherent diffractive microscopy imaging systems;
- Object phase recovery in near-field X-ray imaging systems;
- Prediction of aberrated SLM phase masks for beam-shape correction in the M290 Laser-based Powder Bed Fusion of Metals (PBF-LB/M) system.

Within the AIMS5.0 context, these algorithms are particularly relevant to scenarios requiring fast and precise beam shaping, aberration detection and correction, as well as fast and accurate X-ray imaging and coherent diffractive microscopy. FourierGSNet can enable rapid imaging and precise light control in industries such as biomedical research, laboratory microscopy, the semiconductor industry, the laser industry, and additive manufacturing.

Currently, no AIMS5.0 use case is using this simulation. Use case partners are welcomed to collaborate.

3.5 Automatic Ontology Learning Framework (TU/e)

3.5.1 Technology description

The automatic ontology learning framework supports ontology construction and extension from domain corpora or structured databases using a lightweight pipeline built around a Large Language Model (LLM), such as DeBERTa-v3-large. It combines data enrichment to provide additional context with data filtering to reduce noise. The framework supports three tasks: term typing (i.e., mapping terms to relevant concepts), taxonomy discovery (i.e., building hierarchical relationships between concepts within a taxonomy), and non-taxonomic relation extraction (i.e., identifying relevant triples for a given set of concepts and the relationships between them). This makes the framework suitable for specialised and privacy-sensitive industrial settings.

In practice, the framework takes a domain knowledge source, extracts candidate terms or relations, enriches them with lexical variants and short definitions, and predicts ontology types and semantic relationships. For relation-oriented tasks, candidate filtering reduces the search space and noise, thereby improving efficiency and feasibility in domains where manual ontology engineering is costly.

3.5.2 Evaluation

Committed KPI	Achieved KPI	Validation methodology
Improve ontology-learning quality in line with the Objective 2 target of a 25% improvement in classification precision for AI-supported industrial tools.	The framework increased term-typing precision from 36.3% to 82.8% on a benchmark ontology, thereby exceeding the 25% classification-precision improvement target.	The result was validated on held-out open-source benchmark test data using the official evaluation setup, comparing baseline and enriched variants based on micro-averaged precision, recall, and F1-score.

3.5.3 Reference to project objective 2 success metrics

- Committed KPI: Improve ontology-learning quality in line with the Objective 2 target of a 25% improvement in classification precision for AI-supported industrial tools;
- AI Toolbox integration: shallow integration.

3.5.4 Application scenario in this project and beyond

A possible application scenario is the extension of the digital reference ontology used for smart sustainable business planning, quality traceability, and incident handling. Manufacturing reports, incident logs, manuals, and structured master data can be mined to identify new domain terms, assign them to ontology types, and propose hierarchical and non-hierarchical relationships. This supports faster ontology maintenance, reduces manual ontology-engineering effort, and improves the semantic coverage of downstream knowledge graph applications. The same approach is applicable wherever specialised vocabularies evolve over time and privacy constraints favour self-hosted models over external API-based systems. At its current stage, the framework should be considered a research prototype rather than a fully developed end-user ontology-engineering platform.

3.5.5 Reference to publication:

Latipov, I. A., Holenderski, M. J., & Meratnia, N. (2025, October). IRIS at LLMs4OL 2025 Tasks B, C and D: Enhancing Ontology Learning Through Data Enrichment and Type Filtering. In International Semantic Web Conference 2025 (pp. 1-15). TIB Open Publishing (Technische Informationsbibliothek).

3.6 Improved human motions for robot testing IPL devices (PCL, TUE)

3.6.1 Technology description

- Human motions are generated and validated for use in the models
- Methods for using and translating data from one skeleton model to another
- Generate new data for training AI models and validate the correctness of the generated data.

3.6.2 Evaluation

Committed KPI	Achieved KPI	Validation methodology
Efficiency and Cost reduction: Avoid additional human testing for each newly developed IPL device. Shorter time to market: Achieve a 10–20% reduction in time to market.	Converting models to enable the use of public datasets, and developing and testing motion trails.	Evaluate the reliability of robot testing and define criteria for assessing human likeness.

3.6.3 Reference to project objective 2 success metrics

Technological result contributes to the overall objective 2 KPIs:

- Reduced manufacturing costs in the long term (3-5 years) by ~50%.
- Improved overall productivity of the manufacturing factory by ~20%.

3.6.4 Application scenario in this project and beyond

The technology is used in UC2.

3.7 Data analysis method (PCL, RUG)

3.7.1 Technology description

For the production process of OneBlade devices, a methodology was developed to analyse process data and identify potential improvements. The methodology and findings of the investigation are summarised in the thesis by M. Smits (RUG), *Detecting Defect-Prone Paths in Manufacturing Processes Using Process and Pattern Mining: A Case Study of the Philips OneBlade Assembly Line*. This report is confidential.

3.7.2 Evaluation

Committed KPI	Achieved KPI	Validation methodology
Process: Improved product quality for 1–3 Critical-to-Quality (CTQ) characteristics. Cost reduction: Up to 10% reduction in scrap due to poor quality.	FOR stable < 3%. Efficiency stable above 90%. Field call rate < 0.8%.	Measure the improvement in Cpk and the reduction in scrap levels.

3.7.3 Reference to project objective 2 success metrics

Technological result contributes to the overall objective 2 KPIs:

- Reduced manufacturing costs in the long term (3-5 years) by ~50%.
- Improved overall productivity of the manufacturing factory by ~20%.

3.7.4 Application scenario in this project and beyond

The technology is used in UC2.

3.8 Automatic event and data sharing to orchestrate service delivery (Husqvarna)

3.8.1 Technology description

To considerably reduce the administrative overhead associated with servitization through automated event and data sharing supported by AI.

3.8.2 Evaluation

Committed KPI	Achieved KPI	Validation methodology
Reduce cost of administration in service delivery with at least 80%.	Service administration in June 2025: 1,760 hours of work were required over one year to administer 200 contracts, representing a 75% reduction in administrative effort. A simulation with reduced monitoring and automatic creation of work orders based on Decision Rules demonstrated a further	To measure the business KPIs, we compare the number of contracts a service coordinator can manage within the same period, allowing us to quantify time savings per contract and demonstrate scalability.

	30% reduction in time spent, resulting in an overall administrative effort reduction of 82.5%.	

3.8.3 Reference to project objective 2 success metrics

Technological result contributes to following objective 2 KPIs:

- Improved overall productivity of the manufacturing factory by ~20%
- AI based MES capability > 10 %
- 20% reduction of required data for scaling new ML models to new machines
- Self-adjusted -/ adapted manufacturing processes > 15 %
- Providing documented methods for dynamic application of tools to improve sustainability
- Improved performance and to show an overall 30% faster time-to-market
- Decreased energy consumption for PoC > 10.

3.8.4 Application scenario in this project and beyond

The technology is used in UC3.

3.9 AI models development (Husqvarna)

3.9.1 Technology description

Reduce time to develop AI models, needed for service delivery

3.9.2 Evaluation

Committed KPI	Achieved KPI	Validation methodology
Reduce time to develop AI models, needed for service delivery, by 80-90%.	The process enabled by the SA platform reduces the total end-to-end cycle time from 40 days to an estimated 15 days.	We measure the time spent on AI model development and deployment at each stage, from the start of the process to deployment: Data Ingestion & Preparation; Model Development & Training; Model Optimization & Validation; and Deployment & Integration.

3.9.3 Reference to project objective 2 success metrics

Technological result contributes to following objective 2 KPIs:

- Improved overall productivity of the manufacturing factory by ~20%
- AI based MES capability > 10 %
- 20% reduction of required data for scaling new ML models to new machines
- Self-adjusted/adapted manufacturing processes > 15 %
- Providing documented methods for dynamic application of tools to improve sustainability
- Improved performance and to show an overall 30% faster time-to-market
- Decreased energy consumption for PoC > 10%.

3.9.4 Application scenario in this project and beyond

The technology is used in UC3.

3.10 Automatic tool breakage detection in milling operations (GOIMEK, SAVVY, IDEKO, SORALUCE, MULTIVERSE)

3.10.1 Technology description

The system integrates advanced AI and signal-processing techniques to enable robust, scalable, and real-time detection of tool breakage in milling operations. It continuously acquires high-frequency vibration signals from accelerometers located near the spindle during machining. The signals are processed by a controller using Fourier Transform techniques to extract harmonic components and frequency-domain indicators at one-second intervals. The resulting features are transmitted to an edge platform, where a machine-learning model evaluates the tool condition in near real time. Deployed as a microservice-based architecture at the edge, the system identifies deviations from normal vibration patterns to detect tool breakage and other anomalies and automatically generates alerts for the operator.

3.10.2 Evaluation

Committed KPI	Achieved KPI	Validation methodology
Tool-breakage and tool-wear detection using regression and classification algorithms, achieving >95% accuracy in tool-failure detection.	Tool-failure detection model accuracy in a laboratory environment: • Milling tool-breakage detection: 98.2% accuracy on the Soraluce milling machine based on experiments with induced tool fractures. • Grooving tool-breakage detection: 98.1% accuracy on the Soraluce machine during turning operations, based on experiments conducted until the end of the tool's service life. Accuracy in the shop-floor environment: The detection accuracy on the Soraluce machine at Goimek is 70% for the specific detection of cutting-edge breakage. This increases to 90% when detecting a broader range of anomalies, including runout caused by various issues and damage to the insert-pocket edge.	Evaluation of the tool-breakage detection algorithms: Evaluate the capability of the tool-breakage detection system under laboratory conditions using two sets of experiments: (1) experiments with artificially induced tool failures, and (2) experiments continued until the end of the tool's service life
Tool-breakage detection: Tool failures detected in a	Total time: 1.5 s (from data acquisition to model output).	Tool-failure detection in a real-world environment within 3 seconds: Deploy the detection

real-world environment within 3 seconds.		system on a shop-floor machine and use data from the Savvy platform to validate its performance.
Motif Discovery Singularity Toolkit: Reduce data-segmentation time by 50% compared with the current manual part-program segmentation procedure.	Manual segmentation of operations:70 h ▪ Full customisation of segment definitions by a process expert. ▪ The NC program is manually modified. Segmentation using the Motif Discovery Toolkit: 32 h • Operations and motifs are identified mathematically. • The approach is currently limited to certain parts of the program.	Evaluation of the Motif Discovery Singularity Toolkit: Test the web application and collect user feedback on segmentation quality and ease of use. Compare the average time required for manual segmentation with the time required when using the toolkit, based on expert estimates.
FFT feature extraction for tool-breakage detection: High-frequency data processing and feature extraction at 1-second intervals.	A feature extraction module is integrated into the PAC, providing FFT-based indicators every second through data acquisition, FFT processing, and communication with the Savvy edge platform via Modbus.	Validation at the Savvy Edge platform confirms that high-frequency acceleration data are correctly transmitted to the Ingeteam PAC system, the required features are correctly calculated, and the resulting features are transmitted to the Savvy Edge device at 1-second intervals.



Figure 2. Example of detected broken cutting edge

3.10.3 Reference to project objective 2 success metrics

Technological result contributes to following objective 2 KPIs:

- 25% improvement in classification precision:
  - FFT-based harmonics analysis to ensure high-quality feature extraction from vibration signals focused on the condition of the tool
  - Motif Discovery Singularity Toolkit for automatic signal segmentation
  - Testing of several ML algorithms for comparison.

- 30% improvement in time spent on quality control: Real-time detection of tool breakage reduces the need for manual inspection, towards unattended machining.
- 30–50% Reduction in Re-Configuration Time of the AI-Based System:
  - Dockerized microservices for fast deployment and update
  - Arrowhead integration.
- 20% Reduction of Required Data for Scaling ML Models to New Machines:
  - Modular architecture of the solution: FFT-extraction embedded in PAC system, 1 s FFT feature reading at Savvy Edge and deployment of the Tool Failure detection app
  - Fine-tuning capabilities based on automatic thresholds analysis and update
  - AI-Gym training environment for model fine-tuning.
- Self-adjusted -/ adapted manufacturing processes > 15 %:
  - Fine-tuning capabilities based on automatic thresholds analysis and update
  - Configurable notification/alerting system

3.10.4 Application scenario in this project and beyond

The technology will be used in UC5 – AI Techniques at Different Layers in the Machine Tool Domain. Within the AIMS5.0 project, the proposed technology has been implemented and validated at the shop-floor level on a Soraluce milling machine. The implementation follows a multi-layer architecture, with the hardware architecture consisting of:

- A Programmable Automation Controller (PAC) directly connected to accelerometers installed on the machine
- A SAVVY Edge device, providing edge computing and orchestration capabilities.

High-frequency acceleration signals along the X and Y axes are acquired at the machine level and preprocessed locally within the PAC. Vibration-based harmonic features related to tool condition are then extracted from these signals and transmitted to the SAVVY Edge platform at a 1-second update rate.

The real-time tool-breakage detection application executes a model to determine the tool condition for each identified operation. The model takes tool-harmonic features as input and outputs the estimated tool condition. As a support tool for machine operators, a Grafana dashboard displays the values of the different tool-harmonic features together with the defined thresholds. The operator receives information about the tool condition through:

- Visual feedback on the Grafana dashboard
- Alarms or notifications when abnormal behavior or a potential tool breakage is detected by the model.

Beyond AIMS5.0, exploitation will focus on increasing the number of tools covered and extending the scope of the solution to the analysis of manufacturing operations. To this end, detection algorithms similar to those already implemented will be applied, together with the microservice-based architecture and Arrowhead implementation. For process anomaly detection, the input variables would include CNC/PLC data, enabling the use of the library developed for the medium-frequency domain.

3.11 Object detection and segmentation (EDI, BME)

3.11.1 Technology description

This technology is used to detect and segment individual bottles in RGB camera images or video frames (see figure below: *Object Detection and Segmentation*). It is designed for scenarios in

which multiple bottles may appear in the same image, including unstructured or cluttered arrangements, such as bottles lying in a pile, partially overlapping, or appearing in different orientations.

The method processes each RGB frame independently and produces two main outputs for every detected bottle:

- Object detection output: A bounding box indicating the approximate location of each bottle in the image.
- Instance segmentation output: A pixel-level mask that precisely separates each individual bottle from the background and from other nearby objects.

The processing pipeline consists of two main stages. First, an RT-DETR object-detection model is applied to the input image. RT-DETR detects all visible bottles and returns bounding boxes together with confidence scores. These bounding boxes provide an initial estimate of the location of each bottle in the frame.

In the second stage, each detected bounding box is used as an input prompt for the MobileSAM segmentation model. MobileSAM refines the detection results by generating a detailed pixel-level mask for the corresponding bottle. This enables the system to obtain a more accurate representation of the object shape than can be provided by a bounding box alone.

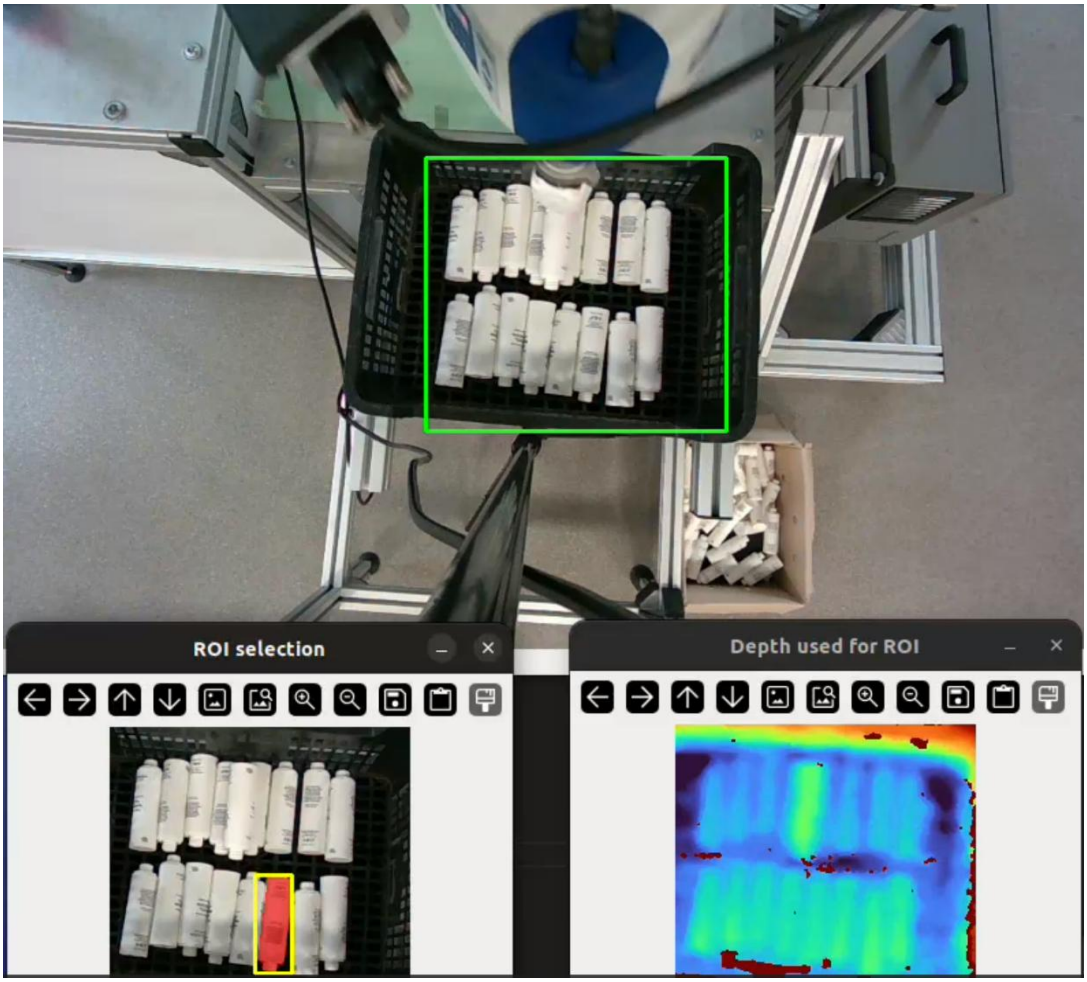


Figure 3. Object detection and segmentation.

3.11.2 Evaluation

Committed KPI	Achieved KPI	Validation methodology
Calculation time < 200ms	Calculation time < 30ms	The processing time is evaluated using images of bottles arranged in a pile. RGB images with a resolution of 640 × 480 pixels are captured using an Intel RealSense camera. The evaluation is performed on a desktop computer equipped with an NVIDIA GeForce RTX 5080 GPU.
Reconfigurability time: 3h	Reconfigurability time: 16h	To reconfigure the system for a new object, the model is fine-tuned using images of the new object. The reconfiguration time is estimated based on the time required to manually label the new images.

3.11.3 Reference to project objective 2 success metrics

Technological result "Object detection and segmentation" contributes to the following success metrics:

- Industrial process phases made more effective and precise through the integrated and tailored AI algorithms, methods and tools.
  - 20% improvement in prediction precision
  - 25% improvement in classification precision
- Building of the AI Toolbox. This technology is going to be deeply integrated (integration level explanation found here: https://aitoolbox.tmit.bme.hu/Getting-started)
- Targeted guideline documents for integration of use-case specific methods
  - Guideline documentation of this technology will be put into AI Toolbox Wiki
  - The technology will be described in the upcoming AIMS5.0 WP2 book chapters
  - The technology will be described in scientific publications.

3.11.4 Application scenario in this project and beyond

The technology is used in UC6. The resulting object pixel masks will be used further in the pipeline to estimate poses for bottles to be picked up. If depth cameras are used, the masks help to extract the part of the point cloud which belongs only to the detected object.

3.12 Single object 6DoF pose estimation (BME,EDI))

3.12.1 Technology description

This technology estimates the 6DoF pose of a detected bottle, i.e., its three-dimensional position and orientation in space, using a known 3D model of the target object (see figure below). The technology is intended for RGB-D camera-based perception, where visual information from the RGB image is combined with depth information to determine the spatial position and orientation of the bottle. The output of the pose-estimation module is the bottle pose expressed in either the camera or robot reference frame, depending on the system configuration and calibration. Currently, two methods have been developed and tested for 6DoF bottle-pose estimation.

The first method is based on the Iterative Closest Point (ICP) algorithm. It uses the output of the object-detection and segmentation module described in Section 3.11 as input. Specifically, the pixel-level mask of the detected bottle is used to extract the corresponding subset of the depth-based point cloud. This segmented point cloud represents the visible portion of the bottle

in the real scene. The ICP algorithm then aligns the observed point cloud with a reference point cloud generated from the bottle's full 3D model. The resulting alignment provides the transformation between the model and the observed object, yielding the estimated 6DoF pose of the bottle.

The second method is based on FoundationPose, a learning-based framework for 6DoF pose estimation. It uses synchronised RGB and depth frames from the camera stream as input. In addition, a 3D model of the target bottle is required, provided in .obj format and specified in the corresponding configuration files. FoundationPose estimates the pose of the known object directly from the RGB-D data and can serve as an alternative to the ICP-based pipeline, particularly in scenarios involving partial visibility, clutter, or challenging object orientations, where learning-based pose estimation may provide more robust results.

Both methods return the 6DoF pose of the bottle, consisting of:

- Position: The x, y, and z coordinates of the bottle.
- Orientation: The bottle's rotation, represented as a quaternion.

In the current implementation, the estimated pose is published in ROS 2 using the geometry_msgs/msg/PoseStamped message type. This allows the pose-estimation module to be integrated directly into the robotic perception and control pipeline, where the detected bottle pose can be used by grasp planning, trajectory planning, and real-time robot-control components. Overall, the technology provides a bridge between visual perception and robotic manipulation. The system determines the bottle's position and orientation in the scene, enabling the robot to identify suitable grasping and manipulation actions.

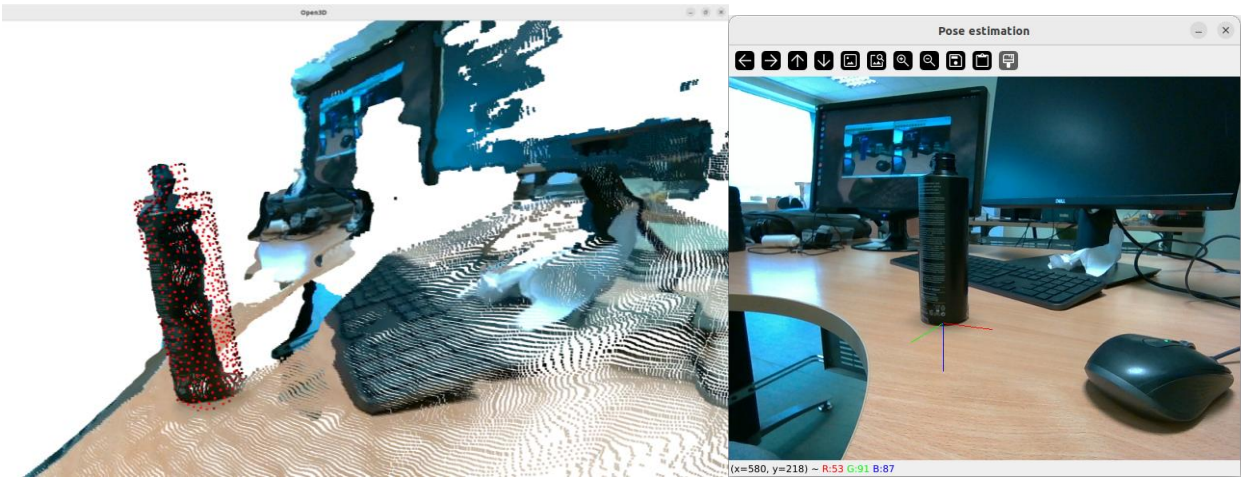


Figure 4. 6DoF pose estimation (ICP method).

3.12.2 Evaluation

Committed KPI	Achieved KPI	Validation methodology
Calculation time < 100ms	Calculation time < 30ms	The calculation time is evaluated using images and point clouds of bottles arranged in a pile, captured with an Intel RealSense camera. The evaluation is performed on a desktop computer equipped with an NVIDIA GeForce RTX 5080 GPU.

At distance 1.5m: • Angle accuracy < 10deg • Position accuracy < 5mm At distance 20cm: • Angle accuracy < 3deg • Position accuracy < 3mm	Accuracy. At distance 1.5m: • Angle accuracy < 60deg • Position accuracy < 325mm At distance 20cm: • Angle accuracy < 3deg • Position accuracy < 17mm	Accuracy is evaluated at two distances: 20 cm and 1.5 m. RGB-D images and corresponding point clouds of bottles are captured using an Intel RealSense camera. The evaluation is performed on a desktop computer equipped with a GPU.
Reconfiguration time: The system operates without reconfiguration when provided with a new 3D model of the target object.	Reconfigurability time: Validated operation without reconfiguration when a new 3D model of the target object is provided.	Reconfigurability is achieved by providing a new 3D model of the target object. It has been validated that the algorithm can automatically use the new model without requiring additional reconfiguration.

3.12.3 Reference to project objective 2 success metrics

Technological result "Single object 6DoF pose estimation" contributes to the following success metrics:

- Industrial process phases made more effective and precise through the integrated and tailored AI algorithms, methods and tools:
  - 20% improvement in prediction precision
  - 25% improvement in classification precision.
- Reduced re-configuration time of the AI-based system by ~30-50%: 6DoF pose estimation allows robots to manipulate objects of previously unknown locations, eliminating the need for reconfiguration in certain cases
- Improved overall productivity of the manufacturing factory by ~20%: 6DoF pose estimation allows robots to manipulate objects of previously unknown locations
- Self-adjusted -/ adapted manufacturing processes > 15 %
- Building of the AI Toolbox. This technology is going to be deeply integrated (integration level explanation found here: https://aitoolbox.tmit.bme.hu/Getting-started)
- Targeted guideline documents for integration of use-case specific methods:
  - Guideline documentation of this technology will be put into AI Toolbox Wiki.
  - The technology will be described in the upcoming AIMS5.0 WP2 book chapters
  - The technology will be described in scientific publications.

3.12.4 Application scenario in this project and beyond

The technology is used in UC6. The estimated position and orientation of detected bottles enable the required gripper pose to be determined for grasping. The technology can be extended to other applications where accurate 6DoF position and orientation estimation of arbitrary objects is required.

3.13 Grasp affordance estimation (EDI)

3.13.1 Technology description

This technology calculates a suitable grasp pose for a detected bottle. The grasp pose defines the position and orientation that the robot gripper should reach to grasp the bottle safely and

reliably (see Figure 5 below). As described in the previous sections, the 6DoF pose of the object to be grasped can be used directly as a target pose for motion planning in MoveIt2. Once the bottle's position and orientation are known, the system can calculate one or more suitable grasp poses based on the object geometry, the preferred grasping region, and the required gripper orientation.

Other approaches do not require the bottle pose to be explicitly estimated before generating a grasp. For example, methods such as GraspGen can directly use RGB and depth streams from depth cameras as input. In this case, the grasp pose is generated directly from the visual and depth information, without requiring a separate 6DoF object-pose estimation step in the pipeline.

Regardless of the selected method, the final output is a 6DoF grasp pose consisting of:

- Position: The x, y, and z coordinates of the desired gripper position.
- Orientation: The orientation of the gripper, typically represented as a quaternion.

The generated grasp pose is intended for use by the robot motion-planning system. In the current architecture, the grasp pose can be forwarded to MoveIt2, which plans a feasible robot trajectory from the current robot configuration to the desired gripper pose. The grasp-pose estimation module therefore serves as an important link between object perception and robotic manipulation.

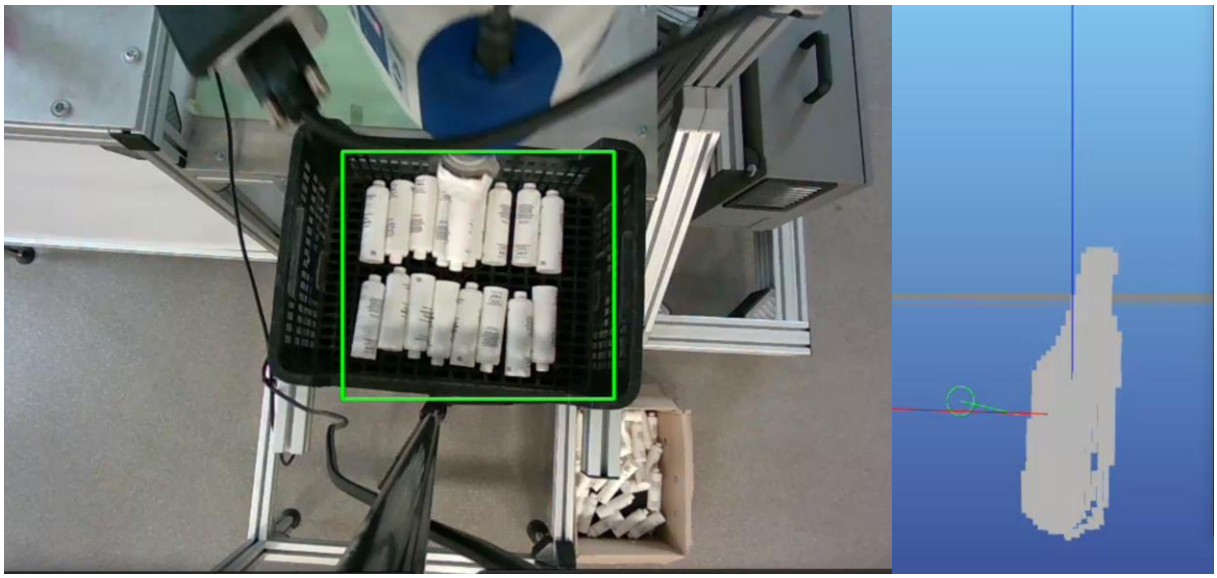


Figure 5. Grasp affordance estimation

3.13.2 Evaluation

The evaluation of the different approaches is still ongoing, as the preferred method depends on the performance of other stages in the pipeline, including the successful pick-up rate and the accuracy of the estimated bottle pose relative to the robot.

Committed KPI	Achieved KPI	Validation methodology
Calculation time < 200ms	Calculation time < 550ms using GraspGen + simplified segmentation < 100ms using ICP result + simple vector calculations for grasp pose	The calculation time is evaluated using images and point clouds of bottles arranged in a pile. The data are captured using an Intel RealSense camera, and the evaluation is performed on a desktop computer equipped with an NVIDIA RTX 3060 GPU.

3.13.3 Reference to project objective 2 success metrics

Technological result "Grasp affordance estimation" contributes to the following success metrics:

- Improved overall productivity of the manufacturing factory by ~20%
- Self-adjusted/adapted manufacturing processes > 15 %
- Building of the AI Toolbox. This technology is going to be included with shallow integration (integration level explanation found here: (https://aitoolbox.tmit.bme.hu/Getting-started)
- Targeted guideline documents for integration of use-case specific methods
  - Guideline documentation of this technology will be put into AI Toolbox Wiki.
  - The technology will be described in the upcoming AIMS5.0 WP2 book chapters
  - The technology will be described in scientific publications

3.13.4 Application scenario in this project and beyond

The technology is used in UC6. The resulting grasp pose is used in subsequent stages of the pipeline as the target pose for the robot motion-planning system. The technology can be applied to other pick-and-place tasks where the same type of gripper is used.

3.14 Grasp planning and execution (EDI)

3.14.1 Technology description

This technology calculates and executes the robot trajectory required to grasp and pick up a detected bottle. The input to the module is a target 6DoF grasp pose defining the desired position and orientation of the robot gripper. Standard MoveIt2 motion-planning algorithms are used to generate a feasible trajectory for the UR5e robot. The planning process takes into account the robot kinematics, joint limits, current robot state, and potential collision constraints within the workspace. A custom interface to the UR5e robot is used to transmit the planned trajectory to the robot controller and execute the grasping motion. The execution typically consists of moving the gripper to the target grasp pose, closing or activating the gripper, and lifting the bottle after a successful grasp.

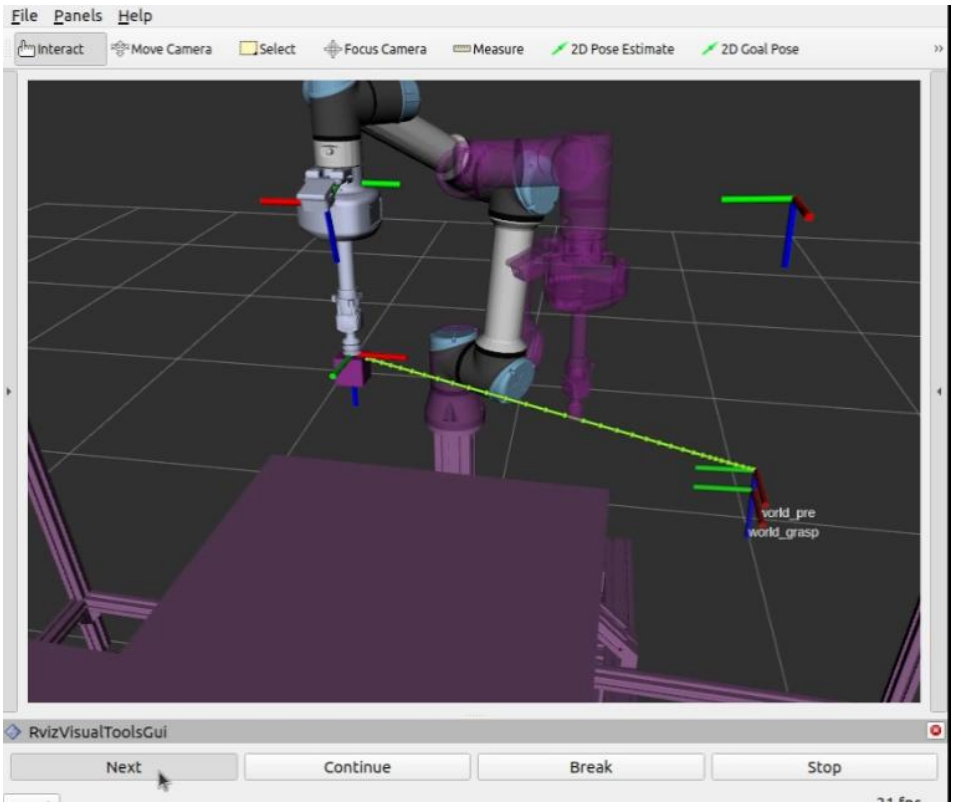


Figure 6. Grasp execution visualization in RViz

3.14.2 Evaluation

Committed KPI	Achieved KPI	Validation methodology
Planning time < 200ms	Planning time: 15-180ms depending on the distance between initial and target position	Planning time is evaluated by measuring the trajectory-planning execution time for different target positions.
Unpicked object ratio < 1/100	Unpicked object ratio: 3/100	The unpicked-object ratio is evaluated by repeatedly attempting to grasp objects from a pile and measuring the proportion of objects that cannot be successfully picked up.

3.14.3 Reference to project objective 2 success metrics

Technological result "Grasp planning and execution" contributes to the following success metrics:

- Improved overall productivity of the manufacturing factory by ~20%.

3.14.4 Application scenario in this project and beyond

The technology is used in UC6. The UR5e manipulator interface developed at EDI has already been actively used and continuously improved across multiple projects at the institute. It is expected to continue to be used in future projects requiring robotic trajectory planning, as the

interface is not limited to the UR5e arm and can also be adapted for mobile robots and other types of robotic manipulators with different gripper configurations.

3.15 Socket detection (EDI)

3.15.1 Technology description

This technology detects the centre position of sockets (see Figure below) on a moving conveyor belt and classifies each socket as either filled or empty. The detected socket position can be used by the robotic system to determine where a bottle should be inserted and to support synchronisation between perception, conveyor motion, and robot actions. Detection and classification are performed using a YOLOv8m neural network trained on manually labelled image data. The model identifies visible sockets in RGB camera frames and determines their locations in the image, including the centre point of each detected socket. At the same time, the model classifies the state of each socket, indicating whether a bottle is already present or whether the socket is available for insertion.

The dataset used for training and testing was collected using Intel RealSense D435 cameras. The cameras operated at 30 FPS with an image resolution of 1280 × 720 pixels, enabling the system to capture the conveyor belt and sockets while in motion. The output of the technology consists of the detected centre position of each socket together with its filled/empty classification. This information can be forwarded to the robot-control and task-planning modules to support bottle placement, conveyor tracking, and decision-making during automated feeding.

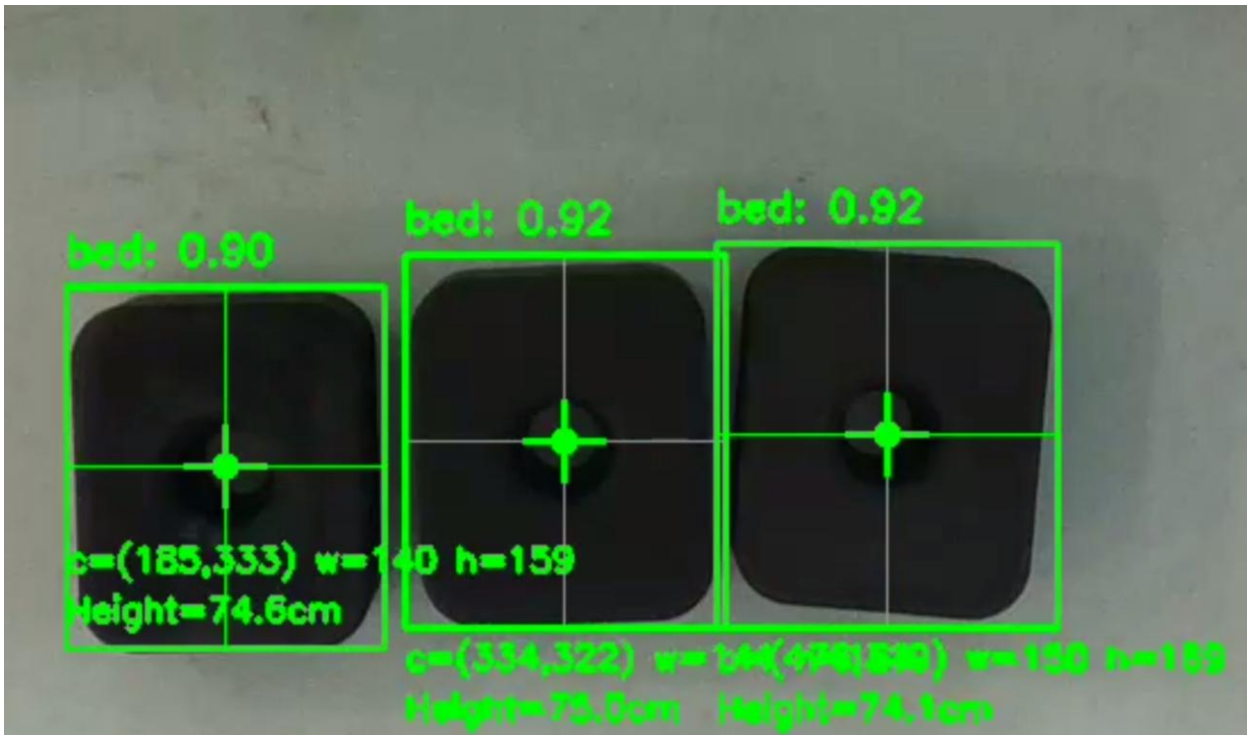


Figure 7. Socket detection.

3.15.2 Evaluation

Committed KPI	Achieved KPI	Validation methodology
Calculation time: < 100ms	Calculation time: < 15ms	The calculation time is evaluated using images of bottles arranged in a pile. Images with a resolution of 1024 × 768 pixels are captured using an Intel RealSense camera, and the evaluation is performed on a desktop computer equipped with an NVIDIA RTX 3060 GPU.
Accuracy. At distance 50cm: • mAP@0.5: > 70% • Position accuracy: <2mm Filled/empty socket classification accuracy > 99%	Accuracy. At distance 50cm: • mAP@0.5: > 70% • Position accuracy: 2mm Filled/empty socket classification accuracy 95% Socket detection accuracy 98.7%	Accuracy is evaluated at a distance of 50 cm. Socket images with a resolution of 1024 × 768 pixels are used, and the evaluation is performed on a desktop computer equipped with a GPU.
Reconfigurability time (new socket, camera angle): 3h	Reconfigurability time (new socket, camera angle): 3h	To reconfigure the system for a different camera angle or socket type, the model is fine-tuned using images representing the new configuration. The fine-tuning time is measured on a desktop computer equipped with a GPU.

3.15.3 Reference to project objective 2 success metrics

Technological result "Socket detection" contributes to the following success metrics:

- Industrial process phases made more effective and precise through the integrated and tailored AI algorithms, methods and tools:
  - 20% improvement in prediction precision
  - 25% improvement in classification precision.
- Reduced re-configuration time of the AI-based system by ~30-50%: The developed technology module will be quickly reconfigurable to new operating conditions.
- Building of the AI Toolbox. This technology is going to be deeply integrated (integration level explanation found here: https://aitoolbox.tmit.bme.hu/Getting-started)
- Targeted guideline documents for integration of use-case specific methods:
  - Guideline documentation of this technology will be put into AI Toolbox Wiki
  - The technology will be described in the upcoming AIMS5.0 WP2 book chapters
  - The technology will be described in scientific publications
  - Minimal target requirements/constraints (at project end).

3.15.4 Application scenario in this project and beyond

The technology is used in UC6. The resulting socket position is used as a target for placing bottles on the conveyor line. The technology can be applied to other industrial scenarios where the same type of sockets needs to be detected.

3.16 Robot arm feedback control (EDI, HUA)

3.16.1 Technology description

This technology implements a robot trajectory feedback-control policy for placing a grasped bottle into a moving socket. A neural network model is used to adjust the robot trajectory during execution, enabling the robot to compensate for the motion of the socket on the conveyor belt. At each robot-control time step, the model outputs target joint positions and velocities for the next step. The model takes as input the current state of the robot arm, including joint angles and joint velocities, as well as the estimated positions of the target socket and the grasped bottle.

The control model is trained using a gradient-based approach in combination with a differentiable physics simulator. This enables the system to learn how changes in robot joint commands affect the final bottle position and to optimise the insertion trajectory during training. For safety, the robot state is continuously monitored to detect potential collisions with the robot itself or with objects in the environment. This monitoring supports safer execution and can provide feedback to higher-level control modules when replanning or error recovery is required. The output of the technology is a sequence of robot joint commands that guides the robot arm during the final bottle-insertion task.



Figure 8. Robot arm feedback control.

3.16.2 Evaluation

Committed KPI	Achieved KPI	Validation methodology
Unplaced obj. rate: <2%	Unplaced obj. rate: 4%	The unplaced-object ratio is evaluated by repeatedly attempting to place a bottle into the socket. The bottle is assumed to be already held by the robot gripper. The number of unsuccessful placement attempts is recorded. A UR5e robot is used for the evaluation.

Policy step evaluation time: <10ms	Policy step evaluation time: 0.3ms	To evaluate the policy execution time, several trajectories are executed, and the time required to calculate a policy action for a single trajectory step is measured on a desktop computer equipped with a GPU.
Reconfiguration time: <22 h days	Setup time for retraining the network: ~22 h	Setup time is measured as the time required to retrain the neural network policy to operate with a different robot, object type, or conveyor configuration. Training is performed on a desktop computer equipped with a GPU.

3.16.3 Reference to project objective 2 success metrics

Technological result "Robot arm feedback control" contributes to the following success metrics:

- Industrial process phases made more effective and precise through the integrated and tailored AI algorithms, methods and tools: 20% improvement in prediction precision
- Improved overall productivity of the manufacturing factory by ~20%: Robot arm feedback control technology enables robots to perform tasks in dynamic environments where robot actions are subject to changing surroundings
- Self-adjusted -/ adapted manufacturing processes > 15 %
- Building of the AI Toolbox. This technology is going to be included in the toolbox with shallow integration (integration level explanation found here: https://aitoolbox.tmit.bme.hu/Getting-started)
- Targeted guideline documents for integration of use-case specific methods
    - Guideline documentation of this technology will be put into AI Toolbox Wiki
    - The technology will be described in the upcoming AIMS5.0 WP2 book chapters
    - The technology will be described in scientific publications.

3.16.4 Application scenario in this project and beyond

The technology is used in UC6. It is the main component responsible for placing bottles onto the conveyor line. It enables real-time robot control using a neural network and could, in principle, be applied to scenarios where conventional trajectory-generation algorithms are insufficient for controlling robot motion.

3.17 AI-based robot control script synthesis (EDI, HUA)

3.17.1 Technology description

The technology generates high-level robot-control logic from a brief natural-language task description. It also generates configuration code for simulating the task in Isaac Sim. The generated simulations can automatically randomise selected aspects of the task environment and goal criteria, such as the number and initial positions of objects in the environment and target positions for pick-and-place operations (see Figure below). During the code-generation process, the task-controller code is executed in the task simulation by an agentic code-generation assistant to test and correct the generated code. This also allows the user to visually inspect the simulation and the final state of each run to validate that the initial task description has been correctly interpreted by the code-generation agent and that the generated implementation achieves the intended task objectives. Validated task implementations are accumulated in a code repository that is accessible to the code-generation agent when

generating new task implementations, providing references for available APIs and established coding patterns.

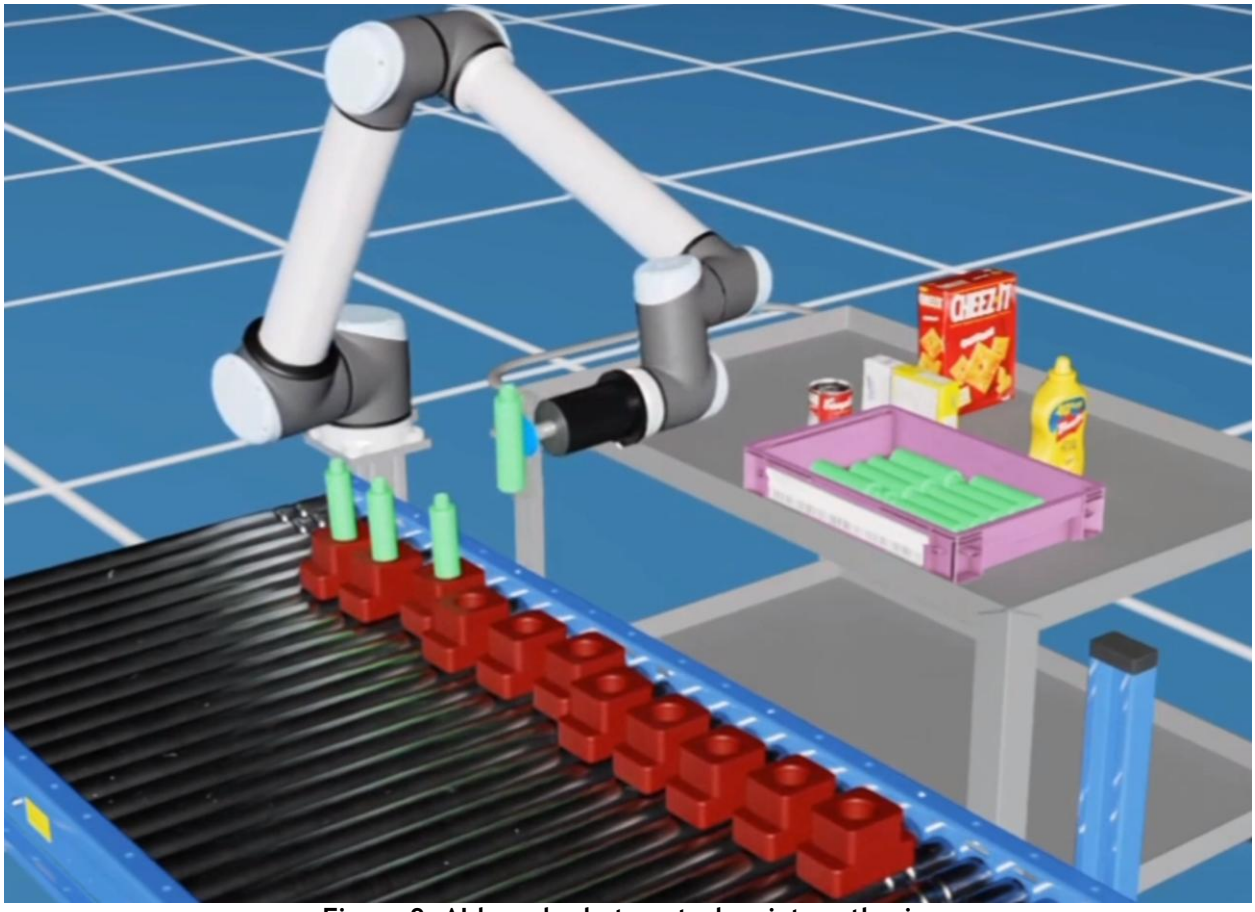


Figure 9. AI-based robot control script synthesis.

3.17.2 Evaluation

Committed KPI	Achieved KPI	Validation methodology
Synthesized script is correct >90%	For tasks that are substantially similar to previously generated tasks, success rates are > 90%.	AI-generated scripts are evaluated in simulation using automatically generated checks to verify that the simulation state satisfies the stated task objectives. The final evaluation is performed by visually observing simulation trials, inspecting the final simulation state, and manually reviewing the generated code.
Synthesis time <1h	Tasks that do not require significantly new or more complex functionality than that already available in the repository can typically be generated within 30–45 minutes. Tasks that require iterative dialogue between the user and the code-generation agent to debug or	Synthesis time is evaluated on a desktop computer equipped with a GPU and includes API calls to remote AI services. The measured time also includes automated testing in simulation and iterative improvements to the generated solution.

	refine new functionality can take 2-3 hours, depending on the complexity of the new features and the number of iterations required.	

3.17.3 Reference to project objective 2 success metrics

Technological result "AI-based robot control script synthesis" contributes to the following success metrics:

- Reduced re-configuration time of the AI-based system by ~30-50%: AI-based robot control script synthesis enables quick adaptation to new tasks
- Self-adjusted/adapted manufacturing processes > 15 %
- 70% reduction of time for developing and integrating a ML model in the digitalized industrial domain: AI-based robot control script synthesis enables quick adaptation to new tasks
- Targeted guideline documents for integration of use-case specific methods:
  - Guideline documentation of this technology will be put into AI Toolbox Wiki.
  - The technology will be described in the upcoming AIMS5.0 WP2 book chapters
  - The technology will be described in scientific publications.

3.17.4 Application scenario in this project and beyond

The technology is used in UC6.

3.18 Computer Vision (Signify, DAC)

3.18.1 Technology description

Explore and implement computer-vision-based methods for quantifying manual assembly processes, including process-step compliance, throughput time, and other relevant process metrics.

DAC contributed to UC7 by developing an MLOps-oriented semi-supervised data-annotation pipeline for industrial computer-vision tasks. The pipeline supports the preparation of training data from production-line video recordings by extracting image frames from videos and using a limited number of manually annotated examples for segmentation and object detection. Based on these initial annotations, the system generalizes the learned detections to support the annotation of previously unlabeled images.

The pipeline includes mechanisms for uploading unannotated data, performing iterative semi-supervised annotation, selecting generated annotations, and evaluating annotation quality. It is designed to reduce the manual effort required to prepare computer-vision datasets in industrial environments and to support faster development and retraining of AI models. The work is also aligned with the broader MLOps methodology, including training-process automation and machine-unlearning mechanisms.

3.18.2 Evaluation

Committed KPI	Achieved KPI	Validation methodology
Acquire quantitative data from the assembly line	The system has been installed in the factory, and data collection	A pilot trial is currently ongoing at our outdoor

during manual process execution. Automate the data-acquisition process to obtain more detailed information and enable additional functionalities.	through video recordings is ongoing. This will be followed by data analysis	luminaire factory in Valladolid, Spain.
Reduce the manual effort required to prepare annotated computer-vision datasets and accelerate the development of AI models for UC7.	The first version of the semi-supervised annotation pipeline has been implemented. The pipeline supports video-to-image conversion, manual annotation of a limited number of samples, semi-supervised generation of additional annotations for object detection and segmentation, and initial evaluation of annotation quality. Final quantitative KPI values will be determined based on the final UC7 validation dataset.	Compare automatically generated annotations against manually annotated reference samples. Evaluate annotation quality using object-detection and segmentation metrics, where applicable, such as precision, recall, IoU, and mAP. Measure the reduction in manual annotation effort and verify pipeline stability across representative UC7 video and image batches.

3.18.3 Reference to project objective 2 success metrics

AI Toolbox integration (WP2):

- Computer vision (exploration results): Shallow (sharing the explored open tools: type, how to use, how to refine); cf. reporting in WP2 deliverables
- Computer vision (implementation): Shallow (description of ongoing study with commercial tool implementation in the factory.

Industrial process phases made more effective and precise through the integrated and tailored AI algorithms, methods and tools:

- 20% improvement is prediction precision
- 25% improvement in classification precision
- 30% improvement in time spent on quality control.

Computer Vision will contribute to:

- Reduced re-configuration time of the AI-based system by ~30-50%.
- Reduced manufacturing costs in the long term (3-5 years) by ~50%.
- Improved overall productivity of the manufacturing factory by ~20%.
- AI based MES capability > 10 %
- 20% reduction of required data for scaling new ML models to new machines
- Providing documented methods for dynamic application of tools to improve sustainability.
- 70% reduction of time for developing and integrating a ML model in the digitalized industrial domain
- Improved performance and to show an overall 30% faster time-to-market.

DAC contribution additionally supports:

- reduction of required data for scaling new ML models to new machines, by enabling semi-supervised annotation from a limited number of manually labelled examples
- reduction of time for developing and integrating a ML model in the digitalized industrial domain, by automating parts of dataset preparation and model training workflow
- Targeted guideline documents for integration of use-case specific methods, through documentation of the MLOps and annotation pipeline approach

- AI Toolbox integration: shallow integration, with the possibility of deep integration if the pipeline is containerized and made available through a callable service/API.

3.18.4 Application scenario in this project and beyond

The technology is used in UC7. Signify focuses on computer vision for quantifying manual assembly and process execution, while DAC contributes an MLOps-based automated annotation pipeline to support dataset preparation for these computer-vision tasks. Beyond AIMS5.0, the pipeline can be reused in industrial scenarios where video or image data need to be annotated for object detection or segmentation, particularly where manual labelling is too time-consuming or costly.

3.19 Intelligent Scheduler (Signify, TU/e)

3.19.1 Technology description

Improve the scheduling and planning of order fulfilment in low-volume, high-variability 3D printing of polymer luminaires.

The Intelligent Scheduler is a production planning and scheduling technology developed by Signify and TU/e for low-volume, high-variability additive manufacturing of polymer luminaires. Its purpose is to improve customer order fulfilment by generating feasible and efficient production schedules under realistic shop-floor constraints. In this setting, orders can vary substantially in product characteristics, processing times, due dates, material requirements, and machine compatibility, making manual scheduling difficult, time-consuming, and prone to inefficiencies.

3.19.2 Evaluation

Committed KPI	Achieved KPI	Validation methodology
Adapt order execution at an hourly level rather than a daily level. Automate the scheduling process to enable human-free scheduling. The objective is to minimise waste while fulfilling customer orders.	Multiple trial runs have been completed, and the results have been analysed.	Evaluate the improvement in order fulfilment and increased planner efficiency through pilot trials at our 3D printing hub in Turnhout, Belgium.
Develop a next-generation scheduler that is adaptable to different factory requirements and takes into account multiple types of anomalies, building on a scientific study conducted at TU/e.	A scientific study was conducted at TU/e, and a prototype multi-objective scheduling framework was developed and evaluated through scenario-based experiments. The scheduler is designed to reschedule regularly in order to adapt to anomalies.	Validation is based on computational experiments and scenario analysis using representative production instances. Performance is assessed by comparing schedule quality across different scenarios, including their effects on tardiness, order fulfilment, material waste, and resource utilisation.

3.19.3 Reference to project objective 2 success metrics

AI Toolbox integration (WP2): Generic libraries and minor optimisations are briefly described. Industrial process phases are made more efficient and precise through the integration of tailored AI algorithms, methods, and tools, resulting in a 30% reduction in time spent on quality control.

3.19.4 Application scenario in this project and beyond.

The technology is used in UC7.

3.20 Improved overall productivity of the manufacturing factory by ~20% Energy-aware scheduling (FUH, IFD, IFAT, SYSTEMA, VIF)

3.20.1 Technology description

Scheduling approaches will be used to reduce energy consumption and improve on-time delivery performance for lots processed on batch-processing machines. The main goal is to establish new scheduling rules that enable automated decision-making in response to high energy peaks. Before the project, no scheduling rules were available, and all decisions had to be made manually. During the validation phase, the team demonstrated the impact on key performance indicators (KPIs), such as lot delays, which should be minimised.

3.20.2 Evaluation

The Use Case was validated through field tests conducted by partners FuH and IFD in IFD's manufacturing environment using real production lots.

Committed KPI	Achieved KPI	Validation methodology
Improvement by genetic programming for on-time delivery cost and for total electricity cost up to 20%	Performance assessment of scheduling heuristics for parallel machines, using randomly generated instances, demonstrated on-time delivery improvements of 16% and 11% relative to a reference heuristic, and 2% for real-world instances from IFD relative to the schedules executed at IFD. The latter result was computed using the native IFD approach as part of a limited field test.	Design, implementation, and performance assessment of energy-aware scheduling algorithms based on a large set of randomly generated problem instances for parallel machines and two-stage flexible flow shops, simulation experiments for complex job shops, and a limited field test using real-world problem instances for parallel machines at IFD.

3.20.3 Reference to project objective 2 success metrics

AI Toolbox integration (WP2): Genetic programming-based algorithms

- For anomaly detection: neural networks and data science tools (Jupyter, TensorFlow, scikit-learn, Bokeh, Pandas, HDF5, and Pickle), as well as efficient numerical processing of large volumes of data using Python libraries such as NumPy and Pandas

- For the validation framework: data visualisation using Python libraries such as Matplotlib, Seaborn, and Bokeh; neural networks based on fully connected layers implemented using TensorFlow/Keras
- Providing documented methods for dynamic application of tools to improve sustainability
- Decreased energy consumption for PoC > 10%.

3.20.4 Application scenario in this project and beyond

The technology is used in UC9.

3.21 Intelligent Energy management (FUH, IFD, IFAT, SYSTEMA, VIF)

3.21.1 Technology description

Machine learning methods will be leveraged to gain a better understanding of energy consumption. The objective is to improve the accuracy and efficiency of energy consumption monitoring across different factory buildings, cleanrooms, and hundreds of pieces of equipment.
Conventional approaches to Energy Management are cost-intensive:

- With the current infrastructure in the semiconductor industry, a breakdown of overall energy consumption at the equipment or product level is not available
- ML models trained on power, production, and weather data can help address this limitation
- This enables a detailed analysis of energy consumption at the product level, as well as forecasting of overall energy demand
- Smart meters are used for high-energy-consuming equipment, enabling automated workflows and equipment shutdowns in response to high energy peaks.

3.21.2 Evaluation

Committed KPI	Achieved KPI	Validation methodology
Potential was shown that > 750 KW/h could be saved per day in case of high energy peaks.	Use an ML model to forecast overall energy consumption, including weather conditions as input to improve forecast accuracy. A proof of concept (PoC) has been completed. High energy peaks are in the range of approximately 23 MW. Measurements demonstrated that up to 750 kW of energy demand could be saved during high-peak periods.	Data collection and analysis: smart power meters are installed on individual tools and busbars, while flow meters are used to measure gas and other media flows at the equipment level. Temperature sensors are installed in the chambers of individual tools to monitor operating conditions.

3.21.3 Reference to project objective 2 success metrics

AI Toolbox integration (WP2): Genetic programming-based algorithms. Industrial process phases are made more efficient and precise through the integration of tailored AI algorithms, methods, and tools:

- Optimisation of machine operations
- Self-adjusted/adapted manufacturing processes > 15 %

- Providing documented methods for dynamic application of tools to improve sustainability.
- Decreased energy consumption for PoC > 10%

3.21.4 Application scenario in this project and beyond

The technology is used in UC9.

3.22 MES Airbus (SYSTEMA, IFD, IFAT, IFAG)

3.22.1 Technology description

A Digital Twin of the automation solution will be created, providing a precise virtual representation of the physical system. The Digital Twin will be seamlessly connected to the actual PLC, higher-level control systems, and, where applicable, the robotics suite, depending on the specific configuration of the automation solution.

From the perspective of all connected software systems, the Digital Twin will replicate the behaviour, responses, and operating logic of the real-world automation setup. This enables engineers and developers to configure, test, and expand software solutions in a realistic environment—without requiring access to the physical automation hardware. As a result, development cycles can be accelerated, downtime risks reduced, and new functionalities validated in a safe, cost-effective manner.

3.22.2 Evaluation

Committed KPI	Achieved KPI	Validation methodology
AI based MES capability > 10 %	Anomaly detection, functional pattern available Y2, capability ~ 5 % shown.	Validation is performed by systematic testing using one year of production data from the high volume semiconductor facility at Infineon Dresden.
RAMII architecture for AI-enabling version of the Manufacturing Execution System (MES).	RAMII architecture specification done with Y2	Validation by tests in the production environment at IFD
AI Based Training for adapted / enhanced Production Scenarios > 10%	First AI based trainings performed shown with UC17	Data proven with factory experts
Development of new Methods and Processes as enabler to enhance operational efficiency > 20%. Will be shown within UC12 and UC17.	WipFlowShare new functionality to detect possible bottleneck usage of one of a kind equipment	Method was proven on existing bottleneck situations (tools, product mix, actual uptime restriction, technologies, basic product types,OEE data, etc.).

3.22.3 Reference to project objective 2 success metrics

- Improved overall productivity of the manufacturing factory by ~20%.
- AI based MES capability > 10 %

- Self-adjusted /adapted manufacturing processes > 15 %

3.22.4 Application scenario in this project and beyond

The technology is used in UC10.

3.23 PH and EC analysis (TIAG, RSA, AIT, GUT, PYT)

3.23.1 Technology description

Sensor-based pH and EC monitoring. The method aims to reduce the daily manual effort required to measure and monitor pH and EC values.

3.23.2 Evaluation

Committed KPI	Achieved KPI	Validation methodology
Improved efficiency ratio by 15% [PYT, RSA]	SOA-based implementation of autonomous elements for humidity, temperature, and air-quality monitoring.	Comparison of the daily pH and EC checks performed manually with the number of checks triggered by the sensor-based pH and EC analysis solution.

3.23.3 Reference to project objective 2 success metrics

- Reduced re-configuration time of the AI-based system by ~30-50%.
- Providing documented methods for dynamic application of tools to improve sustainability
- Decreased energy consumption for PoC > 10%
- Today, plant health is assessed manually through visual inspection once per day. The plant mass estimation algorithm could significantly reduce the time spent on quality control. If testing and validation confirm our hypothesis, the time required for the daily inspection walk could be reduced by approximately 50%.

3.23.4 Application scenario in this project and beyond

The technology is used in UC11.

3.24 Edge computing architecture and relevant SW (TIAG, RSA, AIT, GUT, PYT)

3.24.1 Technology description

Extension of time-triggered scheduling to generate schedules for containers, tasks, and network communication, ensuring end-to-end timing guarantees for container orchestration.

3.24.2 Evaluation

Committed KPI	Achieved KPI	Validation methodology
Improved end-to-end latency by 10% [TIAG]	Reduced response time in NERVE by 75% in a laboratory environment using a preemptible Linux kernel patch.	Evaluation of the measured response time using the PREEMPT_RT patch in the GUT demonstrator.

3.24.3 Reference to project objective 2 success metrics

- Reduced re-configuration time of the AI-based system by ~30-50%.
- Providing documented methods for dynamic application of tools to improve sustainability
- Decreased energy consumption for PoC > 10%
- Today, plant health is assessed manually through visual inspection once per day. The plant mass estimation algorithm could significantly reduce the time spent on quality control. If testing and validation confirm our hypothesis, the time required for the daily inspection walk could be reduced by approximately 50 %.

3.24.4 Application scenario in this project and beyond

The technology is used in UC11.

3.25 Power consumption control (TIAG, RSA, AIT, GUT, PYT)

3.25.1 Technology description

MAPE-K-based power consumption control. This method will be used to optimise the operation of devices installed in the demonstration sandbox environment.

3.25.2 Evaluation

Committed KPI	Achieved KPI	Validation methodology
Improved resource usage and output by +10% [RSA]	Deployment of a GPT-2-based LLM on an edge device as a standalone RESTful service.	Comparison of the energy consumption of various components operating in the sandbox environment, such as the edge device, microcontrollers, and ventilation fans.

3.25.3 Reference to project objective 2 success metrics

- Reduced re-configuration time of the AI-based system by ~30-50%.
- Providing documented methods for dynamic application of tools to improve sustainability
- Decreased energy consumption for PoC > 10%
- Today, plant health is assessed manually through visual inspection once per day. The plant mass estimation algorithm could significantly reduce the time spent on quality control. If testing and validation confirm our hypothesis, the time required for the daily inspection walk could be reduced by approximately 50%.

3.25.4 Application scenario in this project and beyond

The technology is used in UC11.

3.26 Visual plant mass estimation (TIAG, RSA, AIT, GUT, PYT)

3.26.1 Technology description

Combination of a plant-instance detection neural network and a statistically fitted regression-based mass estimator.

3.26.2 Evaluation

Committed KPI	Achieved KPI	Validation methodology
20% improvement in the Incident Tolerant Time Interval (ITTI) and 75% detection rate for visual plant irregularities [AIT].	Continuous data collection is ongoing; the first showcase is ready.	Dedicated benchmark test dataset with manually annotated plant irregularities. Accuracy and recall of machine learning predictions compared with the benchmark dataset.

3.26.3 Reference to project objective 2 success metrics

- Reduced re-configuration time of the AI-based system by ~30-50%.
- Providing documented methods for dynamic application of tools to improve sustainability
- Decreased energy consumption for PoC > 10%
- Today, plant health is assessed manually through visual inspection once per day. The plant mass estimation algorithm could significantly reduce the time spent on quality control. If testing and validation confirm our hypothesis, the time required for the daily inspection walk could be reduced by approximately 50 %
- AI Toolbox integration: deep Integration/Shallow.

3.26.4 Application scenario in this project and beyond

The technology is used in UC11.

3.27 Wireless Communication Enhancement algorithms for smart antenna (GUT)

3.27.1 Technology description

Technological result: Wireless Communication Enhancement Algorithm

In indoor food production (UC11), wireless communication faces significant challenges due to changing environmental conditions throughout the plant growth cycle, which can cause signal interference. Traditional omnidirectional antennas are not able to effectively address these challenges. To overcome this limitation, a beam-switching smart antenna combined with a dedicated communication enhancement algorithm has been developed. The solution improves

signal reliability and stability by dynamically adapting the antenna radiation pattern to changing environmental conditions.

3.27.2 Evaluation

Committed KPI	Achieved KPI	Validation methodology
Improvement of SNIR (signal to noise and interference ratio), 75%	8 dB (400%)	Comparison of received signal strength between a conventional omnidirectional antenna and a switched-beam smart antenna using dedicated communication enhancement algorithms

3.27.3 Reference to project objective 2 success metrics

In a new scenario, a manual or brute-force approach would require testing a large subset of the 4,096 possible antenna configurations, resulting in significant reconfiguration delays. In contrast, the AI algorithm learns the relationship between environmental conditions and beam configurations, enabling it to predict a near-optimal steering vector without searching the entire solution space. This capability substantially reduces reconfiguration time compared with conventional approaches and directly supports the KPI target of a 30–50% reduction.

3.27.4 Application scenario in this project and beyond

The technology will be used in UC11. The beam-switching antenna, known as an ESPAR (Electronically Steerable Parasitic Array Radiator), consists of one active element and twelve passive elements. By assigning binary states of 1 or 0 to the passive elements, the antenna can generate up to 4,096 different beam configurations (2^{12} = 4,096). Each configuration is defined by a steering vector, represented as a 12-bit binary pattern that determines how the antenna directs its signal. To improve connectivity between peripheral devices and the gateway station, GUT has applied advanced optimisation algorithms that automatically identify the best steering vector under changing environmental conditions.

The algorithm is designed to optimise the performance of the ESPAR antenna by evaluating different configurations and identifying the one that provides the strongest signal. It starts with a randomly selected steering vector, represented as a 12-bit binary pattern that controls the antenna configuration, and iteratively evaluates alternative settings. Each new candidate pattern is generated by flipping one bit of the current pattern, and its performance is evaluated using metrics such as RSSI (Received Signal Strength Indicator) or PER (Packet Error Rate). If the new pattern performs better, it is accepted immediately. If it performs worse, it may still be accepted with a certain probability, which decreases over time as the algorithm's "temperature" is reduced. This mechanism allows the algorithm to explore a wide range of possible solutions while becoming increasingly selective as the optimisation progresses. Once all iterations have been completed, the best-performing steering vector is stored and can be used to maintain a more stable and reliable wireless link, even when environmental conditions around the antenna change.

Another algorithm improves the steering of the ESPAR antenna by mimicking the principles of natural evolution. It starts with a small population of randomly generated steering vectors, each represented as a 12-bit binary string. Each steering vector is evaluated using real-world measurements such as RSSI or PER, and a fitness value is assigned based on its performance. The best-performing steering vectors are then selected as "parents." These parent vectors are combined through a crossover process, in which bits from each parent are recombined to

generate new steering vectors. To maintain population diversity and avoid convergence to suboptimal solutions, random mutations are occasionally introduced by flipping individual bits. With each generation, weaker solutions are replaced by better-performing ones, resulting in continuous improvement of the overall population. After a predefined number of generations or once the maximum number of evaluations has been reached, the best steering vector identified is returned and applied to the antenna to achieve a stronger and more reliable wireless link.

3.28 ThreatGet (TIAG, RSA, AIT, GUT, PYT)

3.28.1 Technology description

ThreatGet provides an automated cybersecurity framework for model-based security analysis of a given system architecture. AIT contributes to ensuring secure and resilient architectures by identifying potential cybersecurity risks and defining appropriate security measures, including monitoring activities to mitigate cyber risks and support compliance with the EU Cyber Resilience Act (CRA).

CRA establishes essential cybersecurity requirements for products with digital elements. Compliance will become mandatory for products placed on the European market from the end of 2027 onwards, covering secure engineering processes, lifecycle vulnerability management, risk-based design, and appropriate security controls.

ThreatGet was originally developed for the automotive domain, addressing UN R155 cybersecurity requirements, and was extended in AIMS5.0 to address the essential cybersecurity requirements of the CRA. The analysis workflow follows four structured steps:

- Model: Capture the system architecture, including components, data flows, and trust boundaries
- Analyse: Apply rule-based threat generation across the modelled architecture
- Attack Trees: Analyse kill-chain paths to identify potentially exploitable attack sequences
- Assess Cyber Risks: Quantify potential damage scenarios to support risk-based security decision-making.

ThreatGet's CRA coverage is focused on product-level requirements; process-level obligations, such as vulnerability management throughout the product lifecycle, are outside the scope of ThreatGet by design. A baseline level of CRA coverage was established through its automotive heritage, and the objective within AIMS5.0 is to systematically expand coverage across all applicable essential requirements. Automated compliance checking reduces the manual effort required to reassess compliance following each product update.

3.28.2 Evaluation

Committed KPI	Achieved KPI	Validation methodology
Increase in the security level of more than 15% during the exchange of large data samples [AIT].	ThreatGet analysis of the UC11 indoor farming architecture identified a structured threat landscape across all asset types, including sensors, controllers, and remote servers. A total of 103 damage scenario instances were identified across four damage categories, with Controller Failure (46) and Communication Disruption (24) representing the dominant risk	Model-based assessment: ThreatGet provides an end-to-end analysis of all connected UC11 components, identifying cybersecurity

	drivers. The main attack entry points were Shared Medium (15 threats), Network Device (11), and Remote Server (9). Overall, ThreatGet provided 58% coverage of the 15 CRA essential cybersecurity requirements under Article 13 and Annex I. Seven requirements achieved strong coverage (≥80%): risk assessment (100%); confidentiality, integrity, availability, network impact, and attack surface (90% each); and access control (80%). Four requirements achieved partial coverage (30–79%), while four remained at minimal or no coverage: secure default configuration, data minimisation, data removal, and specific vulnerability tracking. These areas represent the main targets for the next iteration.	weaknesses that could be exploited by attackers across the full IIoT stack. Regulatory alignment check: ThreatGet cybersecurity activities are mapped to the essential cybersecurity requirements of CRA Annex I, providing structured compliance evidence to support risk-based design decisions.

3.28.3 Reference to project objective 2 success metrics

The following success metrics are relevant to ThreatGet's contribution:

- Providing documented methods for the dynamic application of tools to improve sustainability: ThreatGet provides a repeatable, model-based methodology for security analysis that can be re-applied at each product iteration or architectural update, directly supporting documented and auditable security processes.
- Reduced reconfiguration time of the AI-based system by approximately 30–50%: Automation of threat generation and CRA compliance checking in ThreatGet reduces the manual effort required for security reassessment following system updates.

The remaining metrics, including energy consumption and reduction in plant health monitoring time, are not directly addressed by ThreatGet.

3.28.4 Application scenario in this project and beyond

Within AIMS5.0, ThreatGet was applied to UC11 (indoor food production/vertical farming), modelling the complete IIoT architecture, including soil and environmental sensors, sensor ECUs, actuator controllers, and remote monitoring servers. The analysis demonstrates the applicability of ThreatGet's risk-based, kill-chain methodology to cyber-physical IIoT environments, where the manipulation of agronomic parameters such as climate, irrigation, and nutrient dosing can result in crop loss, food-safety incidents, or plant downtime. The results provide direct regulatory traceability to the essential requirements of CRA Annex I, making the findings actionable for compliance-oriented design decisions. More details are presented in D6.6.

3.29 Rapid modelling (SYSTEMA, IFD, IFAT, IFAG, TUD, HSZG, WUW)

3.29.1 Technology description

Rapid modelling technology is based on the application of queueing theory, modelling a production process as a network of queues. This approach helps to understand the interactions between production capacity, inventory, and variability within a production system. It is used to reduce variability in the production process and support more effective capacity planning.

Initially, the approach is used to analyse a work centre with variable performance and identify potential areas for investigation and improvement. Subsequently, the model can be adapted by production and capacity planners to represent a process of their choice and perform rapid analyses.

3.29.2 Evaluation

Committed KPI	Achieved KPI	Validation methodology
Variability reduction in the specific cases: -30%. OEE in specific work centers: +10%	Scenario analysis for a potential factor conducted based on benchmark production data. 50% reduction in repair time variability will lead to 23% reduction in Cycle Time. OEE of ECDP work center at IFD in May 2025: 57.4% (increase of 33.8%).	For variability reduction, potential improvement actions will be discussed and evaluated together with factory experts. Only feasible actions will be considered when quantifying the achievable reduction in variability, which is calculated based on the physical laws governing the factory processes. Improvements in Overall Equipment Effectiveness (OEE) will be validated using existing bottleneck situations, considering factors such as equipment, product mix, actual uptime constraints, technologies, basic product types, and available OEE data: OEE = Availability × Performance × Quality

3.29.3 Reference to project objective 2 success metrics

Improved overall productivity of the manufacturing factory by ~20%.

3.29.4 Application scenario in this project and beyond

The technology is used in UC12.

3.30 WIP level prediction using ML (SYSTEMA, IFD, IFAT, IFAG, TUD, HSZG, WUW)

3.30.1 Technology description

The semiconductor industry already uses factory simulations in the form of short-term and long-term simulations for material flow. Short-term simulations typically cover periods of one to seven days, whereas long-term simulations are used for periods ranging from more than four weeks up to one year. However, these simulation time horizons are not sufficient to support shop-floor decisions that affect WIP variability.

The approach is to use AI models to calculate WIP based on factory data, enabling ultra-short-term simulations (i.e. predictions) that allow shop-floor personnel to make decisions within a single shift or up to one day. This type of prediction can support rapid decision-making during a shift and enable more effective balancing of WIP flows. Such simulations were previously not feasible due to the lack of suitable ML applications.

3.30.2 Evaluation

Committed KPI	Achieved KPI	Validation methodology
Explainability of AI: +10%	The current ML model performance metrics for WIP prediction are RMSE = 0.1035 and MAE = 0.0813. Although predictive accuracy remains limited, the results demonstrate the feasibility of short-term WIP forecasting and highlight the importance of data quality and feature selection	Data validated with factory experts. The WIP predictions generated by the ML models are validated using standardised metrics for regression accuracy and goodness of fit, such as root mean squared error (RMSE) and mean absolute error (MAE). A defined workflow with normalised target values ensures that different datasets and models can be evaluated consistently and reproducibly. In addition, standard regression models are compared with various ML models to assess the accuracy and robustness of the predictions.

3.30.3 Reference to project objective 2 success metrics

- AI Toolbox integration: shallow
- Improved overall productivity of the manufacturing factory by ~20%.

3.30.4 Application scenario in this project and beyond

The technology is used in UC12.

3.31 AI-based decision support (SYSTEMA, IFD, IFAT, IFAG, TUD, HSZG, WUW)

3.31.1 Technology description

ML predictions of WIP levels are presented in a way that enables operators and decision-makers to understand and trust the predictions while incorporating their own subject-matter expertise when making decisions. Therefore, suitable Explainable AI (XAI) approaches and visualisations are applied to improve transparency and support informed decision-making.

3.31.2 Evaluation

Committed KPI	Achieved KPI	Validation methodology
Acceptance of trust in AI: +10%	Pre-test experiments indicate that, so far, people are more likely to follow AI predictions when no additional explanation is provided. When a simple explanation is provided, i.e. when the AI is made more transparent, participants	The approach to AI explainability is being newly implemented and tested for applicability through the continuous involvement of, and feedback from, human decision-makers. The impact on user acceptance and trust will be evaluated by comparing different representations, visualisations, and interaction methods for presenting ML

	are less likely to follow the AI prediction. This may be because the explanation reduces the perceived "mystery" surrounding the AI recommendation. Weight on Advice (WOA): WOA_{AI} = 0.63 WOA_{XAI} = 0.58 Difference: 8.3%	predictions as decision support. Experiments will compare different user groups receiving the same ML prediction through different interventions. Users' Weight on Advice (WOA) will then be measured to determine the extent to which decision-makers accept and trust the AI recommendation and adjust their own judgement accordingly, rather than relying solely on their initial intuition. The Weight on Advice (WOA) is calculated as: WOA = (Initial human prediction − Adjusted human prediction) / (Initial human prediction − AI prediction).

3.31.3 Reference t project objective 2 success metrics

- AI Toolbox integration: shallow.
- Improved overall productivity of the manufacturing factory by ~20%.

3.31.4 Application scenario in this project and beyond

The technology is used in UC12.

3.32 Integrated transportation and production (SYSTEMA, IFD, IFAT, IFAG, TUD, HSZG, WUW)

3.32.1 Technology description

By integrating production and logistics data into a combined simulation model and incorporating a correction factor to account for dynamic effects, a more efficient planning tool is being developed to predict transport loads based on production workloads.

3.32.2 Evaluation

Committed KPI	Achieved KPI	Validation methodology
Variability reduction in the specific cases: -30%	A computation time of 2 minutes per simulated year is achieved, compared with 8 minutes to 62.5 hours for benchmark simulation frameworks, corresponding to a reduction in computation time of approximately 75% to 99.9%. This enables more efficient decision support for identifying and implementing actions to reduce process variability.	The performance of the simulation model is validated against established benchmark simulation frameworks.

3.32.3 Reference to project objective 2 success metrics

- AI Toolbox integration: none (already available online)
- Improved overall productivity of the manufacturing factory by ~**20%**.

3.32.4 Application scenario in this project and beyond

The technology is used in UC12.

3.33 Digital Kanban System (SYSTEMA, IFD, IFAT, IFAG, TUD, HSZG, WUW)

3.33.1 Technology description

Digitalisation and automation of the material delivery process through a digital Kanban system, reducing process variability and manual interventions while increasing efficiency, transparency, and productivity.

3.33.2 Evaluation

Committed KPI	Achieved KPI	Validation methodology
Variability reduction in the specific cases: -30%	Increase in the degree of digitisation (DD) of the logistics process between the cleanroom and warehouse from 20% to 80%, corresponding to a 300% relative increase, while reducing uncertainties, delays, processing time, and overall process variability	The degree of digitalisation is calculated using the methodology defined in the Planning for Digitalisation (P4D) approach. The degree of digitalisation is assessed based on the number of information flows associated with material movements in the Kanban system at the site

3.33.3 Reference to project objective 2 success metrics

Improved overall productivity of the manufacturing factory by ~20%.

3.33.4 Application scenario in this project and beyond

The technology is used in UC12.

3.34 AMHS (HOST, IFD, IFAG, IFAT, FMX, TUD, UzL)

3.34.1 Technology description

A novel hardware and software architecture for the demonstrators of high-volume automated material handling systems (AMHS), including overhead transportation (OHT) systems in Dresden and Villach, will be established to reliably predict and detect hardware malfunctions based on sensor data. The objective is to improve Overall Equipment Effectiveness (OEE) and cycle times by demonstrating predictive maintenance solutions for high-volume AMHS components and increasing the uptime of wafer transportation and storage systems.

Vision-guided object detection and classification will be used in a factory environment to identify and select objects for feeding onto a conveyor belt. The solution operates on an embedded AI hardware platform and analyses video frames captured by an industrial-grade camera within the application scenario.

3.34.2 Evaluation

Committed KPI	Achieved KPI	Validation methodology
Initial machine-learning models were deployed in Year 1 to detect hardware malfunctions.	Implementation of the new hardware at three locations in the cleanroom, enabling fast detection of defective parts in OHT vehicles and malfunctions in the AMHS.	Definition of the new infrastructure and requirements analysis, with parallel development of a strategy for monitoring and controlling vehicles within the OHT system.
Testing with defective hardware will be conducted in Y2, with validation planned for Y3.	Reduction in sensor application costs by using low-cost smartphones for hardware monitoring.	Design and demonstration of an experimental workflow involving a tracking system based on two ToF cameras and one RGB camera for monitoring a swarm of small robots. The system enables the detection and analysis of abnormal behaviour and provides insights into deviations from expected behavioural patterns
First new automation components with lower footprint installed at IFD.	Development of a new sensor platform with novel components.	Development of machine-learning algorithms for anonymising collected visual data, ensuring privacy protection through two complementary methods.
Creation of > 1000 sensor data. Separation on good and bad data ongoing	Measurement setup enabling the analysis of bearing noise at different rotational speeds.	Development of two sensor-box demonstrators integrating ToF cameras, radar, and RGB cameras. The sensor boxes were used to collect datasets, followed by data extraction, preprocessing, noise removal, visualisation, and data interpretation.
Validation of fault-detection patterns for bearings using laboratory setups.	TUD: • New version of the Android application for monitoring the AMHS established. • New visualisations developed to detect hardware malfunctions based on sensor data. • Semi-automated data analysis established to support stakeholder reporting and decision-making.	A customised dataset for tool and human segmentation was constructed by collecting images from multiple online sources, including Kaggle and Images.cv. The dataset was annotated using the Segment Anything Model (SAM 4) to generate pixel-level segmentation masks. The resulting annotated dataset was used to fine-tune deep neural networks for segmentation and detection, including YOLO26-based models. As a result, UzL developed an integrated system

	• Evaluation methodology extended. • Sensor data fusion established, combining acceleration and acoustic data.	for person and tool detection and classification, deployed on a mobile platform to address the challenges of noisy radar signals caused by platform movement.
Successful demonstration of anomaly detection on OHT vehicles using newly developed hardware systems, resulting in improved uptime of wafer transportation and storage systems		Validation of bearing fault-detection patterns using laboratory setups, with ongoing experiments to validate the results using an alternative sensor configuration to improve robustness.
Fast detection of faults in high-volume transport systems, enabling significant savings in material costs and time by avoiding the disposal of expensive components. For example, one vehicle for 300 mm wafers is valued at approximately €30,000. The solution also contributes to reducing the environmental footprint associated with wafer transport, handling, and storage. Increased level of automation in data analysis: the first new automation components with a reduced footprint have been installed at IFD.	Host: Due to the technical constraints, including a maximum conveyor rotational speed of 120 rpm and a high ambient noise level of at least 72 dB, piezoelectric structure-borne sound microphones were identified as suitable sensors.	

3.34.3 Reference to project objective 2 success metrics

- 20% improvement is prediction precision
- Improvement of uptimes of wafer transportation and storage systems by 20%
- AI Toolbox integration:
  - Edge devices
  - Machine-learning models
  - Audio-signal-processing feature extraction
  - Anomaly detection algorithms
  - Unsupervised and supervised learning
  - Predictive-maintenance models
  - Time-series analysis.

3.34.4 Application scenario in this project and beyond

The technology is used in UC13.

3.35 Tool and Person Detection with Classification for Mobile Systems (UzL, IFD)

3.35.1 Technology description

The Tool and Person Detection/Classification solution for mobile systems contributes to an autonomous mobile robot (AMR) solution for use in semiconductor fabs. The objective is to enable reliable detection and classification of humans and tools using a sensor-fusion approach. The baseline system combines a ToF camera and a radar sensor, enabling the detection of people and estimation of their distance and velocity. However, the current system is limited to stationary setups.

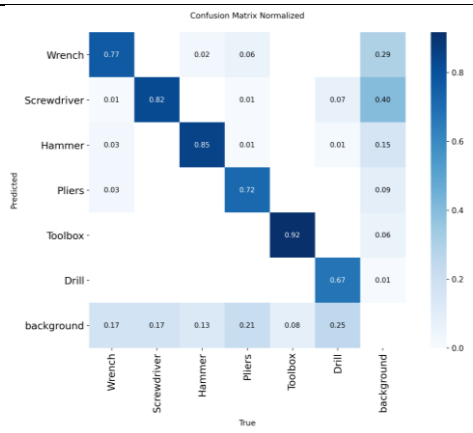
UzL extended the baseline system to enable deployment on a mobile platform. This was achieved through hardware upgrades, including replacing the original ToF camera with a wide-range camera equipped with an IMU for odometry correction and upgrading the SIMO radar to a higher-frequency mmWave MIMO radar.

To address the issue of data scarcity, UzL developed a scalable data-annotation framework supporting multiple tasks, including segmentation, detection, and classification. The framework enables fast and efficient fine-tuning and training of neural networks, particularly in scenarios where suitable datasets are not available. The framework consists of the following steps:

- A customised dataset for tool and human segmentation was constructed by collecting images from multiple online sources, including Kaggle and Images.cv
- The dataset was annotated using the Segment Anything Model (SAM 4) to generate pixel-level segmentation masks
- The resulting annotated dataset was used to fine-tune deep neural networks for segmentation and detection, including YOLO26-based models)
- The YOLO-based model segments the target classes and generates masks for the ToF and radar data, enabling the fusion of classification information with distance and velocity information.

3.35.2 Evaluation

Committed KPI	Achieved KPI	Validation methodology
Reduced re-configuration time of the Al-based system by ~30-50%. Decreased energy consumption for PoC > 10%. Providing documented methods for dynamic application of tools to improve sustainability.	Reduction in dataset collection and annotation time through the implementation of an automated labelling pipeline with direct export to a YOLO-compatible format	Baseline approach: Manual annotation or software-assisted labelling requiring continuous human supervision. Using the UzL framework, 38,418 images were generated and processed within 4 hours on an NVIDIA A100 GPU, demonstrating a substantial reduction in manual annotation effort and processing time.
20% improvement in prediction precision. 25% improvement in classification precision.	High accuracy in tool segmentation and classification was achieved while maintaining a low false-positive rate. Additional	Confusion matrix taken from the last model checkpoint:

	training runs using a refined dataset and expanded class set are currently ongoing The complete sensor-fusion approach is integrated into ROS 2 and operates at approximately 15 FPS, achieving a distance-estimation accuracy of ±5 cm in a real-world test environment on a moving platform.	 Real-world sensor-fusion tests were performed manually, with distances also measured manually.

3.35.3 Reference to project objective 2 success metrics

- 20% improvement is prediction precision
- 25% improvement in classification precision
- Reduced re-configuration time of the AI-based system by ~30-50%
- Decreased energy consumption for PoC > 10%
- Providing documented methods for dynamic application of tools to improve sustainability.

3.35.4 Application scenario in this project and beyond

The technology is used in UC13.

3.36 Virtual Commissioning using a Digital Twin (FMX, IFD, HOST, TUD, IFAT)

3.36.1 Technology description

A Digital Twin of an automation solution will be developed as a precise virtual representation of the physical system. The Digital Twin will be seamlessly connected to the actual PLC, higher-level control systems, and, where applicable, the robotics suite, depending on the specific configuration of the automation solution.

From the perspective of all connected software systems, the Digital Twin will replicate the behaviour, responses, and operating logic of the real-world automation system. This enables engineers and developers to configure, test, and extend software solutions in a realistic environment without requiring access to the physical automation hardware. As a result, development cycles can be accelerated, downtime risks reduced, and new functionalities validated in a safe and cost-effective manner.

3.36.2 Evaluation

Committed KPI	Achieved KPI	Validation methodology
Reduction of the time required for in-house	Current reduction in the time required for in-house	Comparison of the average total time required for in-house setup,

setup, commissioning, and testing of the demonstrator system by 33%.	setup and commissioning: 20%, despite the prototype status of the automation solution and Digital Twin.	commissioning, and testing of the demonstrator system with and without virtual commissioning using a Digital Twin.
Reduction in the time required for on-site setup, commissioning, and testing of the demonstrator system by 33%.	Current reduction in the time required for on-site setup and commissioning: 10%, despite the prototype status of the automation solution and Digital Twin.	Comparison of the average total time required for on-site setup, commissioning, and testing of the demonstrator system with and without virtual commissioning using a Digital Twin.
Reduction in the scheduled tool downtime required for setup and testing at a demonstrator tool by 50%.	The approach will be demonstrated using Digital Twins of other automation systems that require production tool downtime during commissioning by the end of AIMS5.0.	Comparison of the average scheduled tool downtime required for setup and testing at a demonstrator tool, with and without virtual commissioning using a Digital Twin.
Handover of the system to the customer for productive use is achieved 8 weeks earlier compared with the current project milestone plan.	The reductions in in-house and on-site commissioning times achieved to date correspond to approximately 3.5 weeks.	Comparison of the average total handover time to the customer with and without virtual commissioning using a Digital Twin.

3.36.3 Reference to project objective 2 success metrics

- Reduced manufacturing costs in the long term (3-5 years) by ~50%
- Improved overall productivity of the manufacturing factory by ~20%
- Improvement of uptimes of wafer transportation and storage systems by 20%
- Improved performance and to show an overall 30% faster time-to-market.

3.36.4 Application scenario in this project and beyond

The technology is used in UC14.

3.37 Machine Learning, deep Learning (UniPD, IFD; IFAG; IFAT, KAI, STATWOLF, LFOUNDRY)

3.37.1 Technology description

Production-ready machine-learning models will be deployed at scale in semiconductor manufacturing to enable early fault detection and ultimately support quality control. The approach includes:

- development of deep-learning forecasting models and an AI-powered machine-learning pipeline to predict OEE and support continuous improvement
- development of explainable unsupervised anomaly-detection methods to facilitate fast and informed root-cause analysis when troubleshooting problems in semiconductor manufacturing
- development of weakly supervised anomaly-detection methods that incorporate user feedback over time
- classification of defects and defect patterns in silicon-wafer images.

Main goal: significant savings in material and personnel costs, enabling greater focus on high-value-added activities and quality.

3.37.2 Evaluation

Committed KPI	Achieved KPI	Validation methodology
With DBACS, an estimated 25% reduction in the amount of data required for new ML solutions is achieved.	Faster explainability of ML solutions achieved, with AcME[2] providing explanations approximately 100 times faster than SHAP.	Reconstruction / Prediction Error / Misclassification Rate (RMSE, MAE, MCR) – relevant to sub-use cases UC15-subB, UC15-subC, UC15-subD, and UC15-subE. Description: Quantifies how accurately the developed models estimate, reconstruct, or predict target variables, such as soft-sensed or otherwise predicted physical quantities. Model performance is validated by comparing predicted and measured values on a held-out test set. Target: Achieve a low reconstruction/prediction error, typically RMSE < 10% of the signal range or R^2 > 0.9, indicating accurate estimation and good generalisation to unseen data.
Availability of an MLOps framework enabling an estimated 75% reduction in ML model development time.	Development of algorithmic approaches enabling a transition from fully supervised to weakly supervised learning scenarios.	Anomaly Detection Performance (F1-score, MCC) – relevant to sub-use case UC15-subA. Description: Evaluates the model's ability to reliably detect and classify abnormal or faulty conditions in the data. Performance is validated by comparing detected anomalies against labelled or simulated events within the same test sequence. Target: Achieve an F1-score > 0.85 and MCC > 0.80, demonstrating reliable anomaly detection with a low rate of false alarms.
Estimated 55% reduction in the time required to monitor a given industrial process.	MLOps environment established, including model versioning and monitoring, to enable scalable and efficient ML operations.	Global Feature Importance (GFI) – relevant to sub-use case UC15-subA. Description: Assesses model interpretability by verifying whether the model identifies physically meaningful and context-relevant features, such as acceleration, suspension position, and braking signals. The results are validated by comparing the ranked feature importance with domain engineering knowledge or baseline explanation models. Target: Achieve >80% alignment between the top-ranked features and known relevant physical variables, confirming interpretable and trustworthy model behaviour.
Early-warning capability for detecting production excursions and anomalies established.	Final algorithm tuning.	Local Explanation Fidelity – relevant to sub-use case UC15-subA. Description: Evaluates the consistency between local, per-sample explanations and the model's predictive behaviour, ensuring that the explanation mechanism faithfully reflects the factors influencing the model

		output. Validation is performed by measuring the overlap between features identified as influential by the XAI method and those that effectively drive the model's predictions. Target: Achieve >70% overlap or agreement with model-based sensitivity maps, demonstrating coherent and faithful local explanations.
Anomaly detection enables significant effort and cost savings in semiconductor manufacturing, particularly in the area of defect-density control.	Significant reduction in manual wafer inspection and manual decision-making efforts compared with 2023. Total amount of automated classified images increased to >10%. Several engineering success stories. High benefits for material and personal cost reduction. 12/2025.	
3% improvement in the average accuracy of defect image classification models.	Average accuracy improved from 82.4% to 89% for the Base Material 1 & 2 models.	Average accuracy measured on the validation data

3.37.3 Reference to project objective 2 success metrics

- Industrial process phases made more effective and precise through the integrated and tailored AI algorithms, methods and tools
- Reduced re-configuration time of the AI-based system by ~30-50%
- 20% reduction of required data for scaling new ML models to new machines
- 70% reduction of time for developing and integrating a ML model in the digitalized industrial domain
- Improved performance and to show an overall 30% faster time-to-market
- AI Toolbox integration:
  - Interpretable anomaly detection with diffi: Depth-based feature importance of isolation forest https://aitoolbox.tmit.bme.hu/Tools/DIFFI
  - Interpretable Data-Driven Anomaly Detection in Industrial Processes with ExIFFI https://ieeexplore.ieee.org/abstract/document/10761374.

3.37.4 Application scenario in this project and beyond

The technology is used in UC15.

3.38 Components RF design acceleration (THALES, FORTH)

3.38.1 Technology description

The design of RF components (e.g., RF MEMS) is a multiparametric and time-consuming process that requires evaluating various parameters, such as material properties and layer thicknesses, both individually and in combination. Design-related 3D electromagnetic (E/M) simulations are typically performed using specialized FEM-based software, which can be computationally intensive. Our aim is to accelerate this process through a supervised learning approach, where properly trained models can provide rapid predictions and significantly reduce the time required to redesign components to meet new specifications.

3.38.2 Evaluation

Committed KPI	Achieved KPI	Validation methodology
Reduce the re-design time (e.g. to meet new specification like the insertion loss in specific frequency or to adopt fabrication details such as interface roughness in RF MEMS) (20%).	Preliminary models trained on data from RF MEMS devices with diverse characteristics and specifications have demonstrated their ability to provide accurate predictions rapidly. Compared with the conventional 3D FEM approach, the trained models show significant potential to accelerate the redesign of RF MEMS components for new applications. A suitable methodology for a fair quantitative comparison is still under development; however, the preliminary results indicate that the KPI is likely to be fully achieved by the end of the project. The reduction in redesign time is already substantial: the conventional E/M simulation tool requires several minutes to perform parametric calculations for 5–7 representative cases, whereas the trained model can provide predictions almost instantaneously. Even including the current computational overhead, which has not yet been fully optimized, the model-based approach requires significantly less time for redesign. Based on these preliminary results, a reduction in redesign time of more than 20% is anticipated.	Using specialized 3D E/M simulation software for high-frequency electronics, RF MEMS designs covering a range of relevant specifications are generated. The resulting simulation data are then used to train ML models using methodologies established within the UC. The predictive performance of the trained models is evaluated using a dedicated test set comprising cases that were not included in the training data and therefore have not been previously seen by the models. The acceptable prediction error is application-dependent; for example, an error of 1 dB for RF MEMS isolation may be acceptable depending on the specific application and requirements. The acceleration of the redesign process will be evaluated by directly measuring and comparing the time required to redesign a device to meet new specifications using the conventional 3D E/M simulation approach and the trained ML models.

3.38.3 Reference to project objective 2 success metrics

- Improved overall productivity of the manufacturing factory by ~20%
- Providing documented methods for dynamic application of tools to improve sustainability.

3.38.4 Application scenario in this project and beyond

The technology is used in UC16.

3.39 RF components fabrication optimization (THALES, FORTH)

3.39.1 Technology description

For RF devices, every dimensional parameter is critical, as even small deviations from the intended design can significantly affect device performance and result in failure to meet the required specifications. The aim of this task is to identify potential fabrication imperfections after each processing step and correlate them with the corresponding characterization results. Using a properly trained ML model, we will assess whether a detected imperfection is likely to affect subsequent fabrication steps and determine whether it could have a critical impact on the final device performance. This activity is supported by our ISO-compliant ecosystem, the quality-control procedures applied after each fabrication step, and our advanced characterization capabilities.

3.39.2 Evaluation

Committed KPI	Achieved KPI	Validation methodology
Reduce fabrication time by 15% by identifying undesirable imperfections at an early stage of the fabrication process (e.g., photoresist thickness affecting the MEMS air gap).	No trained ML models are available yet, as data are continuously being collected during fabrication and characterization. However, identifying an imperfection at the final fabrication step of an RF MEMS device—representing the most conservative scenario for this KPI—has already demonstrated the potential to avoid approximately 25% of the fabrication/development time being spent on a design direction that would ultimately have resulted in device failure. This assessment was performed using the currently available data in the form of a lookup table to identify and predict potential future failures. The approach will be further enhanced with trained ML models once all characterization activities have been completed. Therefore, the 15% reduction target for this KPI is anticipated to be achieved.	This will be assessed by estimating the time “gained” compared with the baseline fabrication sequence and the work plan established before the project. The assessment will consider the time saved by identifying imperfections at different stages of the fabrication process. This activity is closely linked to the following activity.

3.39.3 Reference to project objective 2 success metrics

- Improved overall productivity of the manufacturing factory by ~20%.
- Providing documented methods for dynamic application of tools to improve sustainability.

3.39.4 Application scenario in this project and beyond

The technology will be used in UC16.

3.40 New manufacturing paradigm for RF components (THALES, FORTH)

3.40.1 Technology description

This activity is closely linked to the previous one, as the early identification of imperfections can support decisions such as: (i) stopping the fabrication process at an early stage, (ii) implementing corrective actions, or (iii) redirecting the device to another suitable application, where possible. In this way, the use of infrastructure and personnel resources can be optimized. In addition, this activity also benefits from the following one, as scenarios aimed at more sustainable fabrication can lead to optimized fabrication approaches in terms of both facility utilization and personnel effort.

3.40.2 Evaluation

Committed KPI	Achieved KPI	Validation methodology
Optimize infrastructure utilization and personnel effort by 15%.	Based on the above, a direct optimization of effort of at least 25% has already been demonstrated. More advanced approaches involving trained ML models are being considered for Year 3. These approaches could further optimize the overall effort by predicting appropriate corrective actions, such as those described above (e.g., using dry rather than wet release for RF MEMS).	This will be evaluated using a methodology similar to that described above, with particular emphasis on personnel effort and facility utilization. More advanced approaches, such as reusing devices that are not suitable for their original application in other applications, are currently under consideration. The resulting benefits will be assessed against the work plan established before the start of the project.

3.40.3 Reference to project objective 2 success metrics

- Improved overall productivity of the manufacturing factory by ~20%
- Providing documented methods for dynamic application of tools to improve sustainability.

3.40.4 Application scenario in this project and beyond

The technology is used in UC16.

3.41 Greener/more sustainable manufacturing for RF components (THALES, FORTH)

3.41.1 Technology description

The proposed UC16 approach, in all its aspects, directly supports more sustainable production. In addition to the specific actions described above, further activities are being undertaken to identify opportunities to reduce material consumption (e.g., noble metals) and minimize waste while maintaining high fabrication standards. It should be noted that, in nanofabrication, commonly adopted approaches such as replacing one material with another are often not feasible due to the stringent process and performance requirements.

3.41.2 Evaluation

Committed KPI	Achieved KPI	Validation methodology
Reduce consumables and chemical wastes (10%)	With regard to sustainability, a monitoring protocol has been introduced within UC16. Based on this protocol, we are working in three main directions: ▪ Assessing different fabrication scenarios to identify the most sustainable option for each case; ▪ Considering alternative fabrication approaches (e.g., electroplating) to reduce the amount of wasted noble metals, where feasible; ▪ Correlating the collected data with the approaches described in KPI #2 and KPI #3 to identify materials, chemicals, and consumables that can be avoided as a result of early warnings. Preliminary results from the first two approaches, focusing on gold utilization—which cannot currently be replaced—indicate that the 10% reduction target is expected to be achieved.	A sustainability monitoring protocol has been introduced during the project, and the assessment will be based on this protocol. As part of this activity, the quantities of materials, chemicals, and general-use consumables are currently being monitored in detail. The different fabrication approaches will then be compared to quantify the resulting reductions.

3.41.3 Reference to project objective 2 success metrics

- Improved overall productivity of the manufacturing factory by ~20%
- Providing documented methods for dynamic application of tools to improve sustainability.

3.41.4 Application scenario in this project and beyond

The technology is used in UC16.

3.42 Development of RF components capable to support diverse applications (THALES, FORTH)

3.42.1 Technology description

This is a hardware-based use case. In parallel with the AI-related activities, hardware components supporting new specifications are being continuously developed. These include GaN-based semiconductor devices, RF MEMS, as well as the previously mentioned sensors based on MEMS technology and Surface Acoustic Wave (SAW) technology.

3.42.2 Evaluation

Committed KPI	Achieved KPI	Validation methodology
This is not a KPI defined in the proposal; however, as part of the activities aimed at achieving the project objectives, the two teams (FORTH and THALES) are developing hardware and devices (RF MEMS, sensors, and HEMTs) designed to meet a range of specifications in terms of RF performance, RF power, reliability, and other relevant parameters. The specifications are defined according to the target applications and may support a range of domains, including telecommunications and avionics.	Following the initial design activities, devices are being continuously fabricated. Functional RF MEMS and HEMTs have already been fabricated, while the sensors are currently in the final fabrication phase and will be considered for on-chip monitoring. At the same time, new approaches are being developed and are also approaching their final stage, such as the use of new dielectric materials for RF MEMS.	In this case, validation against the defined specifications is performed using characterization techniques that monitor the relevant device performance parameters (e.g., S-parameters, pull-in voltage, etc.).

3.42.3 Reference to project objective 2 success metrics

- Improved overall productivity of the manufacturing factory by ~20%
- Providing documented methods for dynamic application of tools to improve sustainability.

3.42.4 Application scenario in this project and beyond

The technology is used in UC16.

3.43 Sensors and actuators for cleanroom monitoring (FhG-IISB, IFD, IFAG, IFAT, PFEIFFER, SYSTEMA)

3.43.1 Technology description

Sensor systems will be used for fully automated monitoring and control of semiconductor manufacturing cleanrooms. Static sensors based on a meander structure will be developed to detect electrical signals upon exposure to hazardous gases. These Si-based AMC sensors will contribute to maintaining stable manufacturing yields.

3.43.2 Evaluation

The technology has been fully developed (100% completion). Validation was performed directly in the cleanrooms of FhG IISB and IFD, confirming that the sensors are fully functional.

Committed KPI	Achieved KPI	Validation methodology
First results on sensor with Neander showed a good feasibility. Using the sensors we will reduce the reaction time from 8 h to less than 1 hour.	Small sensor components offer a cost-effective solution for monitoring gases in cleanrooms and enabling early detection of contamination. IFD has demonstrated the manufacturability of gas sensors as integrated circuits on 300 mm silicon wafers and has conducted initial tests of different sensor coatings. New sensor concepts have also been developed by the IFD and IFX PSS groups. In addition, an initial concept for deploying the sensors as the basis of a sensor network has been established.	Development of a demonstrator for static sensors based on a meander structure as a foundation for a prospective sensor network.
Initial tests of the Si-based sensors have shown promising results. By the end of the project, we aim to demonstrate the feasibility of manufacturing these sensors on 300 mm silicon wafers.	Design of a heuristic scheduling framework for flow-shop environments based on genetic programming.	Sensors designed to detect electrical signals upon exposure to hazardous gases.
Validation of MQTT protocols for data sharing and contamination control algorithms. Reduction of reaction time from 8 h to less than 1 h.	Pfeiffer: An appropriate algorithm enables abnormal behavior to be detected as soon as it occurs.	Construction of sensors on silicon wafers.
Deployment of pilot visualization tools and heat maps for contamination data analysis. Visualization dashboards are available.	Final algorithm tuning.	Use of ML for prediction and root cause analysis.
	Coating of novel sensors, determination of the optimal coating thickness, and testing of demonstrators for AMC detection.	Corrosion monitoring system utilizing meander probes and a plastic chamber with integrated sensors for measuring gas contamination levels, such as HCl.
	Progress in contamination monitoring strategies, advanced data analysis methods, and	Development of Si-based AMC sensors in the cleanroom: coating of novel sensors,

	concepts for wafer container logistics.	determination of the optimal coating thickness on surfaces, and testing of demonstrators for AMC detection based on conductivity changes.
	Good data availability has been achieved through the use of a data lake and KNIME.	Preparation of the system by spin coating using carefully optimized parameters for various electrode geometries and materials.
	Fabrication of sensors on silicon wafers: recovery demonstrated over three cycles through heating of the films.	
	MQTT communication protocol capability demonstrated for the contamination control tool, particularly the APA tool.	
	The first data storage application has been demonstrated and integrated into an external computer. The first demonstrator has been developed and implemented on the Infineon APA platform.	
	Online visualization of multiple sensor data streams for two pilot applications has been established. A toolset and strategy for cleanroom concepts aimed at preventing AMC contamination at IFD have been defined. The automation and digitization of tools and processes have been defined.	

3.43.3 Reference to project objective 2 success metrics

Industrial process phases made more effective and precise through the integrated and tailored AI algorithms, methods and tools:

- 25% improvement in classification precision
- 30% improvement in time spent on quality control.

3.43.4 Application scenario in this project and beyond

The technology is used in UC17.

3.44 Semantic web-based DR (IFAG, FUH, NXP, IFD)

3.44.1 Technology description

The Semantic Web is an extension of the traditional Web in which information is explicitly annotated with semantic metadata, enabling it to be interpreted by both humans and machines. Its functionality is enabled by a specific architecture, commonly referred to as the Semantic Web stack or Semantic Web Layer Cake. Various technologies work together to uniquely identify data resources, establish relationships between resources, link additional data sources to provide a more comprehensive representation of information, and incorporate logic into this representation.

For UC18, we are developing a machine-readable knowledge representation of the UC objectives to facilitate AI-based improvements. This formal, machine-readable representation of semantic relationships is referred to as an ontology. Ontologies enable AI systems to identify previously hidden relationships and knowledge, which can subsequently be used to support tailored improvements in production and supply-chain processes.

3.44.2 Evaluation

Use Case 18 has delivered expert-validated models for Economic and Ecological Alpha, ATP, customer service, and carbon handprint. The carbon handprint model adds 34 classes and one OAP use case, supporting a potential mid-double-digit percentage reduction in CO_2 emissions and approximately 30× CO_2 savings in semiconductor applications. In addition, the Economic and Ecological Alpha KPI model adds seven DR classes, targeting measurable CO_2 savings per transistor. The ATP model adds eight new classes and is expected to support a low-double-digit percentage improvement, while the customer service level model adds seven classes, with both absolute and relative improvements compared with the competition expected.

The models were validated through expert review, ontology generation, reasoner validation, data mapping, inference testing, standardization, user feedback, and continuous improvement. This process was completed for Use Case 18, resulting in a machine-readable and reusable economic and ecological operating-curve model. Beyond the model deliverables, Use Case 18 also contributed to capability building: one full-time employee was hired, two PhD students were trained, 20 modelers were trained through internships, and approximately 1,000 people were reached through the DtC Roadshow training.

Committed KPI	Achieved KPI	Validation Methodology
Potential for high two-digit percentage improvement in less CO2 per transistor function within the next 3-5 years in production.	Expert-validated semantic model with 7 DR classes delivered to represent Economic and Ecological Alpha and support assessment of CO2 improvement per transistor in production.	Validated against expert-defined ground truth through expert review, ontology generation, reasoner validation, data mapping, inference testing, formal standardization, user feedback, and continuous improvement.
As a result of improved Cpk, the defect rate (1 – Yield) has the potential to decrease by a mid-single-digit percentage.	The Yield concept has been integrated into the semantic model to represent the effects of improved Cpk and support the assessment of a	Validated through semantic mapping and inference testing, ensuring correct representation and linking of

	potential mid-range reduction in (1 − Yield).	Yield relationships within the model.
As a result of earlier problem detection, (1 − OEE) has the potential to decrease by a mid-single-digit percentage.	The OEE concept has been integrated into the semantic model to represent the benefits of earlier problem detection and support the assessment of a potential mid-range reduction in (1 − OEE).	Validated through semantic mapping and inference testing to ensure that OEE relationships are consistently represented within the model.
As a result of improvements in SPC, Yield, and OEE, overall ATP stability is expected to improve by a low-double-digit percentage, while missed ATP is expected to improve by a low- to mid-double-digit percentage.	An expert-validated ATP semantic model with eight DR classes has been delivered, integrating SPC, Yield, and OEE concepts to support the assessment of improved ATP stability and its potential for a low-double-digit percentage improvement.	Validated through expert review, ontology generation, reasoner validation, data mapping, and inference testing of ATP-related concepts, including SPC, Yield, and OEE.
As a result of improved ATP, an absolute increase in the customer service level in the low-double-digit percentage range is expected, while a mid-single-digit percentage improvement relative to the competition is expected.	An expert-validated customer service level semantic model with seven DR classes has been delivered, supporting the assessment of potential absolute low-double-digit and relative mid-single-digit improvements compared with the competition.	Validated through domain-expert review, ontology generation, reasoner validation, data mapping, and inference testing of service-level relationships linked to ATP improvements.
Manpower qualification and retention are expected to improve by a low-double-digit percentage.	Qualification and retention supported through the hiring of one full-time employee, the training and hiring of two PhD students, the training of 20 intern modelers, and the training of approximately 1,000 people through the DtC Roadshow.	Validated through enablement outcomes, including employee hiring, PhD qualification and hiring, intern modeler training, and broader training reach through the DtC Roadshow.
Potential to reduce CO_2 emissions by a low- to mid-double-digit percentage, contributing to carbon-neutrality efforts; UC18 targets approximately 30× greater CO_2 savings in semiconductor applications.	The Carbon Handprint semantic model, comprising 34 classes and one OAP use case, has been delivered, supporting the assessment of the potential for a mid-double-digit percentage reduction in CO_2 emissions and approximately 30× CO_2 savings in semiconductor applications.	Validated through expert review, ontology generation, reasoner validation, sample data mapping, inference testing, formal standardization, stakeholder feedback, and continuous improvement.

3.44.3 Reference to project objective 2 success metrics

- Potential for high two digit % improved less CO_2 per transistor function within next 3-5 years in production;
- significant reduction of being off from the middle; distribution (cp) remains constant but cpk can improve by two-digit % or more;
- result of better cpk, (1-yield) has the potential to decrease by a mid % range;
- Result of earlier detection of problems (1-OEE) has the potential to decrease by a mid % range;
- Potential to reduce mid-level two digit % of CO_2 emissions, which will contribute to carbon neutrality efforts. Target with UC18 to keep about 30 times of CO_2 savings in the applications of semiconductors.

3.44.4 Application scenario in this project and beyond

The technology is used in UC18.

3.45 Digital Reference Ecosystem (IFAG, FUH, NXP, IFD)

3.45.1 Technology description

Within the first two years of the project, Use Case 18, in collaboration with the WP4 group, extended the Digital Reference Ecosystem through the development of a semantic model comprising 48 classes, 72 edges, and 58 object properties. The model has been uploaded to the current version of the OAP.

The key problem addressed by the UC is the need for greater transparency regarding the end-to-end sustainability of a given semiconductor product and its supply chain, enabling environmentally focused optimization and improvement.

3.45.2 Evaluation

The solution approach involves the use of an end-to-end supply chain simulation model that leverages operating curve (OC) and ecological operating curve (eOC) concepts to enable companies to estimate the impact of different utilization levels on the sustainability of a given product, supplier, and customer.

Committed KPI	Achieved KPI	Validation methodology
Generation of an expert-validated semantic model for the operating curve and ecological operating curve, followed by verification and application of data to the model.	▪ 1 expert-validated semantic model; ▪ 7 DR classes.	Economic and Ecological Alpha to be integrated to demonstrate improvements in CO_2 emissions per transistor in production.
Generation of an expert-validated semantic model for ATP availability, including SPC, Yield, and OEE concepts, followed by application of the model to demonstrate improvements in ATP availability.	▪ 1 expert-validated semantic model; ▪ 8 DR classes.	ATP availability is expected to improve by a low-double-digit percentage. The integration of concepts such as SPC, Yield, and OEE will contribute to achieving this KPI.

Generation of an expert-validated model for customer service level and application of data to demonstrate a low-double-digit percentage improvement in customer service level.	▪ 1 expert-validated semantic model; ▪ 7 DR classes.	As a result of improved ATP, an absolute increase in the customer service level in the low-double-digit percentage range is expected, while a mid-single-digit percentage improvement relative to the competition is expected.
Implementation of a carbon footprint and handprint semantic model. Validation using UC-qualified synthetic data to demonstrate footprint reduction while maintaining or increasing the carbon handprint.	▪ 1 Carbon Handprint semantic model; ▪ 34 classes; ▪ 1 OAP use case.	Potential to reduce mid-level two digit % of CO2 emissions, which will contribute to carbon neutrality efforts. Target with UC18 to keep about 30 times of CO2 savings in the applications of semiconductors
Train more students with the aim of retaining them within the company or the semiconductor ecosystem, depending on business needs and student performance.	▪ 1 full-time employee hired; ▪ 2 PhD students trained and hired; ▪ 20 intern modelers trained; ▪ ~1000 trained with DtCRoadshow.	Manpower qualification and retention should be improved by a low two % digit range

3.45.3 Reference to project objective 2 success metrics

ATP availability: As a result of improvements in SPC, Yield, and OEE, overall ATP stability is expected to improve by a low-double-digit percentage, while missed ATP is expected to improve by a low- to mid-double-digit percentage.

Customer service level: As a result of improved ATP, an absolute increase in the customer service level in the low-double-digit percentage range is expected, while a mid-single-digit percentage improvement relative to the competition is expected.

3.45.4 Application scenario in this project and beyond

With the ontology now aligned with the standards used in the Digital Reference (WP4), the next steps will involve establishing secure and scalable data exchange mechanisms between the UC18 ontologies and the Open Access Platform (OAP) using the Arrowhead framework.

This integration will facilitate company-to-company querying and sharing of data on economic, ecological, and societal improvements. The use of shared class structures between the Digital Reference and UC18 will help minimize duplication and future integration efforts.

Additionally, the Digital Reference Learning Tool, developed by the IFX Supply Chain Innovation Team, enables users to provide targeted feedback on specific ontology classes, which is subsequently reviewed by industry experts. This process directly contributes to the ongoing improvement and evolution of the semantic model.

3.46 Open Access Platform (IFAG, FUH, NXP, IFD, DAC)

3.46.1 Technology description

The Open Access Platform (OAP) is another technology with which UC18 interacts in collaboration with WP4. The OAP provides functionality for visualizing the semantic model and is also expected to enable easier ontology generation through its research on semi-automated ontology generation.

DAC contributed to the implementation and extension of the Open Access Platform, focusing on data integration and access procedures, semantic interoperability, marketplace mechanisms, and integration with the Eclipse Arrowhead framework. The platform supports semantically described data exchange by combining the Digital Reference Ontology with knowledge-graph-based data representation and controlled access mechanisms.

DAC's contribution included extending the OAP functionalities developed during the first project year, supporting integration with the Digital Reference Ontology to ensure semantic consistency, adding support for marketplace mechanisms, and enabling the visualization of knowledge graphs and ontologies. In the context of UC18, DAC also supported the planning use case by implementing APIs for secure access to analytical and planning tools available through the OAP, mechanisms for monitoring plan execution, and transformations between knowledge graph representations and simulation and planning formats such as AutoSched and AnyLogic.

3.46.2 Evaluation

Committed KPI	Achieved KPI	Validation methodology
Provide reusable OAP procedures and services for semantic data integration, data discovery, and controlled data access to support UC18 and related sustainability and planning scenarios.	OAP functionalities were extended with Eclipse Arrowhead integration, marketplace mechanisms, Digital Reference Ontology integration, knowledge graph and ontology visualization support, planning use case automation, plan execution monitoring, and data transformation mechanisms for simulation and planning tools. D4.6 was completed and submitted by DAC as the responsible partner.	Validation through UC18 and OAP scenarios: verify the publication, discovery, and access of semantically described data; test transformations between knowledge graph representations and AutoSched/AnyLogic formats; verify API access to planning tools; assess plan execution monitoring; and verify Arrowhead-based service interoperability and marketplace metadata discovery without exposing sensitive raw data.

3.46.3 Reference to project objective 2 success metrics

The OAP contribution supports the following Objective 2 success metrics:

- Building a structured toolbox repository and enabling reuse of AI tools and methods across use cases
- Providing documented methods for dynamic application of tools to improve sustainability
- Reduced re-configuration time of the AI-based system by ~30-50%, by enabling standardized semantic data access and reusable integration procedures

- 70% reduction of time for developing and integrating a ML model in the digitalized industrial domain, by simplifying access to semantically described data and planning/simulation tools
- I Improved overall manufacturing productivity by approximately 20%, indirectly supported through planning automation, monitoring, and improved interoperability of AI-enabled tools.
- AI Toolbox integration: shallow integration for documentation and procedures; potential deep integration for OAP APIs and services exposed through the platform.

3.46.4 Application scenario in this project and beyond

The technology will be used in UC18 as part of the planning use case and the broader Open Access Platform activities. Within AIMS5.0, the OAP enables access to semantically described data, supports knowledge-graph-based representation, and provides mechanisms for integrating planning and analytical tools. DAC's contribution supports data transformation between the knowledge graph and simulation and planning formats, secure access to OAP tools through APIs, and plan execution monitoring.

Beyond the project, the OAP can be reused as a secure and interoperable platform for industrial data sharing, AI tool integration, planning support, and sustainability-related analysis. The approach is particularly relevant to manufacturing environments where data sovereignty, semantic interoperability, and controlled access to sensitive industrial data are required.

3.47 Production Trace Algorithm (NXP, Semaku, TU/e)

3.47.1 Technology description

The production trace algorithm can be used to extract production traces from lot-level manufacturing event data. These traces often involve lot splits and merges, introducing uncertainty that is quantified by the algorithm. The algorithm can support improved decision-making in various applications, such as root-cause analysis and risk assessment.

3.47.2 Evaluation

Committed KPI	Achieved KPI	Validation methodology
nable the extraction of production traces from datasets containing lot-level manufacturing events, including lot splits and merges. Depending on the use case, this can reduce the number of resources requiring inspection by approximately 10%.	Interaction profile showing how different factors affect the KPI/metric and whether they lead to an improvement. The analysis indicates improvements ranging from 0% to 30%.	Use discrete-event simulation to generate data for different scenarios. Apply the algorithm to the generated datasets and evaluate the resulting metrics.

3.47.3 Reference to project objective 2 success metrics

- 30% improvement in time spent on quality control
- Reduced manufacturing costs in the long term (3-5 years) by ~50%
- Improved overall productivity of the manufacturing factory by ~20%.

3.47.4 Application scenario in this project and beyond

The technology is used in UC19.

3.48 Supply Chain Simulation Model (NXP, Semaku, TU/e)

3.48.1 Technology description

The supply chain simulation model will be used to test and validate the supply chain master plans generated by both the linear programming (LP) model and stochastic optimization models. The model represents the complete semiconductor supply chain, including supply-side factors such as lead times and yield, as well as demand uncertainty. The model operates in a Python environment.

3.48.2 Evaluation

Committed KPI	Achieved KPI	Validation methodology
Validate supply chain master plans under uncertainty, with at least 75% of products covered by a statistically valid uncertainty modelling approach.	88% of high-demand, stationary products can be modelled with statistical significance.	Use historical demand data and the Kolmogorov–Smirnov (K–S) test to perform hypothesis testing.

3.48.3 Reference to project objective 2 success metrics

- 30% improvement in time spent on quality control
- Reduced manufacturing costs in the long term (3-5 years) by ~50%
- Improved overall productivity of the manufacturing factory by ~20%.

3.48.4 Application scenario in this project and beyond

The technology is used in UC19.

3.49 Linear Programming (LP) Model (NXP, Semaku, TU/e)

3.49.1 Technology description

The linear programming (LP) model is used to generate supply chain master plans and serves as a baseline for the stochastic optimization models and the master-planning heuristic. The LP model replicates the supply chain master-planning engines used by NXP. It operates in a Python environment and requires a Gurobi license to run.

3.49.2 Evaluation

Committed KPI	Achieved KPI	Validation methodology
Generate baseline master plans for a	The current runtime is 2–4 minutes. However, dual sourcing is not yet included in the model, and its	Compare the computational effort required by the ESP

product group within 5 minutes.	integration is expected to increase the computational effort required by the LP model.	engine with that of the LP model.

3.49.3 Reference to project objective 2 success metrics

- 30% improvement in time spent on quality control
- Reduced manufacturing costs in the long term (3-5 years) by ~50%
- Improved overall productivity of the manufacturing factory by ~20%.

3.49.4 Application scenario in this project and beyond

The technology is used in UC19.

3.50 Stochastic Optimization Models (NXP, Semaku, TU/e)

3.50.1 Technology description

The stochastic optimization models are being developed to increase plan stability in supply chain master planning. Two models are proposed: two-stage stochastic programming (2SP) and chance-constrained programming (CCP). 2SP models decision-making under uncertainty by dividing decisions into two stages: first-stage decisions are made before uncertainty is realized, while second-stage (recourse) decisions are made after outcomes are observed, typically minimizing expected costs across scenarios. CCP imposes constraints that must be satisfied with a specified probability (e.g., at least 95%), thereby directly controlling the risk of constraint violations rather than modelling explicit recourse actions. The key difference between the two methods is that 2SP focuses on optimizing decisions while allowing adaptive responses to uncertainty, whereas CCP focuses on ensuring feasibility with high probability, often without explicitly modelling detailed second-stage corrective actions. Both models operate in a Python environment and may require a Gurobi license.

3.50.2 Evaluation

Committed KPI	Achieved KPI	Validation methodology
Improve plan stability under uncertainty, reducing rescheduling frequency by at least 20% compared with LP-based plans.	One scenario was tested using a stochastic optimization model (two-stage stochastic programming), demonstrating an approximately 70% improvement in plan stability compared with the LP-based approach.	Calculate plan stability using the defined stability metric, taking into account the number of changes made to the plan. Compare the stability metric between the LP and stochastic optimization models. Use multiple scenarios to assess the robustness of the stochastic optimization models.

3.50.3 Reference to project objective 2 success metrics

- 30% improvement in time spent on quality control
- Reduced manufacturing costs in the long term (3-5 years) by ~50%
- Improved overall productivity of the manufacturing factory by ~20%.

3.50.4 Application scenario in this project and beyond

The technology is used in UC19.

3.51 Supply Chain Master Planning Heuristic (NXP, Semaku, TU/e)

3.51.1 Technology description

The supply chain master-planning heuristic is being developed as a computationally efficient alternative to the linear programming (LP) model. The heuristic will be used to create a rapid scenario-modelling framework that specifically addresses the complexity and stochastic nature of semiconductor supply chains. We propose a rule-based heuristic that decomposes the planning problem into sequential planning steps. The heuristic operates in a Python environment.

3.51.2 Evaluation

Committed KPI	Achieved KPI	Validation methodology
Enable rapid scenario modelling with a deviation of no more than 20% from LP model outputs and at least a 20% reduction in computation time.	In terms of solution quality, the rule-based heuristic outperforms the greedy benchmark on the resource utilization and inventory position KPIs. In terms of computational speed, we conclude that the heuristic scales linearly with problem size, whereas the LP model scales exponentially.	Compare the plans generated by the greedy heuristic and the rule-based heuristic to determine deviations. Use multiple demand scenarios to assess the robustness of the heuristic. Compare the computational effort required by the LP model and the heuristic.

3.51.3 Reference to project objective 2 success metrics

- 30% improvement in time spent on quality control
- Reduced manufacturing costs in the long term (3-5 years) by ~50%
- Improved overall productivity of the manufacturing factory by ~20%.

3.51.4 Application scenario in this project and beyond

The technology is used in UC19.

3.52 LLM for Root Cause Analysis (NXP, Semaku, TU/e)

3.52.1 Technology description

The LLM-based framework for root-cause analysis is used to query knowledge graphs using SPARQL. It is based on the SPARQL-CAR framework developed by TU/e (see Section 3.2). The LLM translates natural-language root-cause questions into SPARQL queries, while syntax and semantic checking and repair steps validate and refine the queries before execution.

3.52.2 Evaluation

Committed KPI	Achieved KPI	Validation methodology
Automate root-cause identification with at least 80% accuracy in retrieving relevant insights from knowledge graphs, while reducing engineering analysis time by at least 50%.	Achieved at the research-prototype level: SPARQL-CAR achieved an F1 score of approximately 0.67 on the QALD-9-plus open-source dataset, with syntax and semantic checking and repair.	Compare LLM-generated SPARQL queries and retrieved answers against expert-validated results. Test multiple natural-language questions to assess the robustness of the LLM. Measure the time required to retrieve answers.

3.52.3 Reference to project objective 2 success metrics

- 30% improvement in time spent on quality control
- Reduced manufacturing costs in the long term (3-5 years) by ~50%
- Improved overall productivity of the manufacturing factory by ~20%
- AI Toolbox integration: shallow integration.

3.52.4 Application scenario in this project and beyond

The technology is used in UC19.

3.53 Single & Multi-Stage Quality Control Models (NXP, Semaku, TU/e)

3.53.1 Technology description

The models focus on quality detection and prediction in wire-bonding processes. Developing these models enables improvements in product and process quality and supports operational decision-making.

3.53.2 Evaluation

Committed KPI	Achieved KPI	Validation methodology
Reduce quality incidents with unknown root causes by 20%. Reduce quality-related manufacturing delays by 10%.	Approximately 3% decrease in quality incidents and approximately 5% decrease in quality-related manufacturing delays.	Compare the number of (near-)incidents detected or predicted by the developed models with the baseline. Calculate the percentage of devices that do not need to be placed on hold.

3.53.3 Reference to project objective 2 success metrics

- 30% improvement in time spent on quality control
- Reduced manufacturing costs in the long term (3-5 years) by ~50%
- Improved overall productivity of the manufacturing factory by ~20%
- AI Toolbox integration: shallow integration (only subset of the models).

3.53.4 Application scenario in this project and beyond

The technology is used in UC19.

3.54 Test Wafer Harmonization (IFAT, IFD, IFAG, TUD)

3.54.1 Technology description

A novel unified UI will be developed to harmonize test wafer management across the 300 mm fab sites of IFD and IFAT. The goal is to reduce the total number of test wafers required, thereby minimizing costs while maintaining quality standards.

3.54.2 Evaluation

The technology is fully developed, and the results have exceeded expectations. To evaluate its impact, we measured the level of harmonization between the two fabs and the cost savings associated with test wafers.

Committed KPI	Achieved KPI	Validation methodology
7 out of 19 work centers are fully harmonized, with an average harmonization level of 45%.	BETA version of the ANKO harmonization UI completed and deployed at the Villach and Dresden sites. Initial test wafer concept specified and aligned across the sites. First rollout of Version 1.0 completed. Harmonization targets defined for 7 of 19 targeted work centers; 7 of 19 work centers fully harmonized, achieving an average harmonization level of 45%. Infineon internal Tableau reporting established and fully operational for project progress tracking. Lessons learned documented and aligned across cross-site teams, with improvement measures implemented accordingly. Enhanced Tableau reporting method deployed for detailed data analysis and KPI visualization. Relevant databases connected to support improved test wafer cost optimization. Relevant databases connected and data extraction completed, resulting in approximately €700k in savings in Year 2.	Identification of additional requirements to further improve existing software functionalities. Establishment of a BETA version of the harmonization UI. Trial and verification of the BETA ANKO harmonization UI, including extensive usage and training with harmonization teams across IFAT and IFD, following the OneVirtualFab approach. Specification of the target state by each team; activities partly completed and ongoing. Acceptance and demonstration of the defined target states through initial test runs; activities partly completed and ongoing. Preparation for sustainment, including the collection of interim lessons learned and alignment on identified challenges and risks. Development of algorithms to track material usage, process efficiency, recycling, and associated costs.

	Deviation response time reduced by approximately four production days, enabling production to return to full flow faster. Feasibility and effectiveness of the harmonization approach successfully demonstrated. First successful update of the harmonization UI completed. BETA version of the ANKO harmonization UI successfully completed for the Villach and Dresden sites. UI deployed and actively used across IFD and IFAT sites; harmonization software successfully implemented and fully operational.	Development of new database architectures for two factories. Reduction of ANKOs through measures such as phase-out and merging. Reduction of wafers per ANKO, e.g., from three to one. Reduction of ANKO frequency. Material optimization through measures such as reuse and recycling. Adjustment and optimization of logistics processes. Spin-off activities, including Rolling ANKO, Circular ANKO, and ANKO Reporting.
Approximately €700k in savings achieved in Year 2 through database integration and data-driven optimization.		
Approximately 4 production days saved in restoring production to full flow.		

3.54.3 Reference to project objective 2 success metrics

- Reduced manufacturing costs in the long term (3-5 years) by ~50%
- Improved overall productivity of the manufacturing factory by ~20%.

3.54.4 Application scenario in this project and beyond

The technology is used in UC20.

4. Expert Groups – Final Contributions and the Book

During the closing phase of AIMS5.0, the cross-WP2 Expert Groups consolidated their joint technical work into a series of book chapters, which together form the AIMS5.0 book, *Artificial Intelligence in Manufacturing Leading to Sustainability and Industry 5.0*. The chapters cover Image Processing; Robotics, Automation, and Process Management; Reinforcement Learning, Optimization, and Scheduling; Anomaly and Fault Detection and Predictive Maintenance; Large Language Models in Industry 5.0; and Machine Learning in Industry 5.0.

Each chapter brings together contributions from multiple project partners, combines a structured review of the state of the art with original industrial use cases developed within AIMS5.0, and concludes with an explicit identification of research gaps relevant to Industry 4.0 and Industry 5.0. The final contributions are summarized below.

4.1 Machine Learning Models and Techniques

The ML chapter, led by BME and AITIA, complements the other five chapters by focusing on the non-deep, "traditional" machine-learning toolkit that remains a workhorse of industrial deployments due to its interpretability, modest computational requirements, and ease of certification under emerging regulations such as the EU AI Act. It frames the transition from Industry 4.0's efficiency-driven AI to Industry 5.0's human-centric, resilient, and sustainable AI; reviews supervised, unsupervised, semi-supervised/active, and non-deep reinforcement and evolutionary learning methods; and maps them onto four cross-cutting application themes: sustainable production, resilient and adaptive supply chains, collaborative robotics, and ML for digital twins.

Thirteen AIMS5.0 use cases are surveyed, with a master table cataloguing the algorithms used by each use case.

UC03 (StreamAnalyze) deploys edge-side decision trees, gradient-boosted decision trees, TinyML, and incremental learning for predictive maintenance across a fleet of 500 robotic lawnmowers, achieving an 82.5% reduction in administrative costs and reaching TRL 6–7. UC05 (IDEKO) detects tool breakage in heavy-duty milling using accelerometer data, FFT-based harmonic feature extraction, and binary classifiers (Random Forest, k-NN, Logistic Regression, and SVM), achieving accuracy above 0.95. This is complemented by Motif Discovery and Change Point Detection for automated NC-program segmentation in unlabelled production data.

UC07 (Signify, TU/e) develops an adaptive scheduler for distributed 3D-printing hubs using discrete-event simulation, Markov decision processes with reinforcement learning, and genetic algorithms, complemented by semi-supervised, iterative self-training, and incremental-learning annotation. This approach increases mean IoU from 0.4 to approximately 0.8 and was subsequently deliberately halted as a "pilot-learn-fast/fail-fast" outcome. UC08 strengthens human–manufacturing interaction through Kalman filtering for tracking, Common Spatial Pattern features for EEG-driven exoskeleton control, and Euclidean clustering of 3D point clouds for human/obstacle separation, achieving 80% human-presence detection and 85% gesture-recognition accuracy.

UC09 (Infineon Dresden / IFD) addresses energy-aware scheduling of batch-processing wafer-fab equipment using Genetic Programming, achieving a 16% reduction in weighted tardiness and an 11% reduction in electricity costs. Constraint Programming is used for grid-load-shifting response, complemented by K-means clustering, Symbolic Aggregate Approximation, PCA, and time-series analysis of more than 700 facility tags and approximately 20 GB of cooling-system

data. The analysis identified anomaly clusters that revealed actual equipment damage and demonstrated peak-shaving potential of more than 750 kWh/day.

UC11 (GUT, TTTech) uses Simulated Annealing and binary classification on a Nerve-edge platform to control ESPAR antenna beam switching for indoor-farming connectivity, achieving an approximately 400% improvement in SNIR, with SPSA used as a comparison baseline.

UC12 (Infineon) implements XGBoost (RMSE 39.4) for WIP forecasting, benchmarked against Regression Tree and Gradient Boosted Tree models, and combines it with SHAP for transparent decision support. The approach delivers a 33.8% increase in OEE and an 8.3% improvement in operator AI acceptance, measured by Weight-on-Advice.

UC13 (Infineon) applies Isolation Forest, DBSCAN, and spectral and time-series analysis to MEMS microphone and acceleration data from OHT shuttles to support predictive maintenance of wafer-transport systems.

UC15 (LFoundry) operationalises Isolation Forest for anomaly detection in sputtering and metal-etch processes, with DIFFI, AcME-AD, and Extended B-ALIF providing fast, model-agnostic explainability and human-in-the-loop refinement, reducing monitoring time by 55%.

UC16 (Thales, FORTH) uses Linear Regression and Decision Trees to correlate fabrication parameters with the performance of RF-MEMS and SAW sensors.

UC17 contributes a Genetic Programming-based flow-shop heuristic for FOUP cleaning, an edge-side anomaly detector for Pfeiffer Vacuum equipment, ML-driven root-cause analysis using ICP-MS data, and Statistical Process Control for airborne contamination monitoring.

UC19 combines Random Forest classification for wire-bond quality, Causal Rule Mining and K-means validation on knowledge-graph-traced data, multi-stage quality prediction achieving approximately 3% fewer final-test failures, and Two-Stage Stochastic and Chance-Constrained Programming for resilient supply-chain planning, achieving up to 40% improved plan stability and approximately 20% backlog reduction.

UC20 (Infineon Villach and Dresden) harmonises the test-wafer monitoring strategy across two 300 mm fabs using Isolation Forest and DBSCAN.

DAC contributed to this Expert Group from an MLOps perspective, covering automated model-training workflows, dataset preparation, and machine unlearning. The contribution focused on practical challenges associated with deploying and maintaining AI models in industrial environments, including data versioning, annotation quality, retraining workflows, AI-component validation, and mechanisms for removing selected data from trained models when required.

The chapter's challenges, research gaps, and conclusions are still being finalised, but the use-case review already demonstrates that interpretable, certifiable, low-footprint ML—combined with explicit explainability layers and human oversight—remains a central component of the AIMS5.0 vision for Industry 5.0.

4.2 Image Processing

The Image Processing chapter is a joint contribution by EDI, UniPD, BME, Signify, KAI, DAC, and GUT, documenting deep-learning-based vision as a core enabler of Industry 5.0, including real-time quality control, robotic perception, and human-in-the-loop automation. The review covers synthetic dataset generation for industrial vision—including text-to-image generation, diffusion

models, image-to-image translation, vision-language models, simulation engines, 3D-to-2D rendering, and the corresponding fidelity, utility, and privacy metrics—as well as multi-camera, multi-target tracking systems and continual machine learning.

The original contributions span pose estimation, socket detection, semiconductor quality control, luminaire assembly action recognition, and industrial data annotation.

The hybrid 6-DoF pose-estimation method (BME) combines AI-based segmentation (RT-DETR + SAM) with ICP fitting on the resulting point cloud, achieving accuracy comparable to FoundationPose while being approximately 20× faster per frame and eliminating the need for a textured CAD model.

The socket-detection pipeline (EDI) for a cosmetics-line conveyor feeding station uses a YOLOv8m segmentation model trained on a small custom dataset, achieving an RMSE of 3.3 mm for socket-center localization under controlled conditions and highlighting the remaining gap to industrial tolerances.

The semiconductor quality-control contributions (UniPD and KAI) include defect-image classification on an imbalanced, multi-material SEM dataset using EfficientNetV2S with class-balanced resampling and moderate augmentation; a CycleGAN-based domain-adaptation method (DBACS) for SEM distribution shifts in UDA and SSDA settings, benchmarked against AdaMatch; and a supervised U-Net-based anomaly detector for aluminium corrosion, achieving 97.1% balanced accuracy with built-in interpretability through segmentation masks and outperforming unsupervised baselines, ResNet, and EfficientNet.

The luminaire-assembly action-recognition study (Signify) fine-tunes VideoMAE on factory video data and analyses the cost–benefit trade-offs and generalisation limits across product variants.

The semi-supervised industrial data-annotation pipeline (GUT), built on RT-DETR with confidence-filtered pseudo-labelling, approximately doubles the initial mean IoU on a battery-detection dataset. DAC contributed to the Image Processing Expert Group through the UC07 work on semi-supervised automated data annotation for computer vision. The contribution addressed the practical challenge of reducing manual annotation effort for industrial image and video datasets. The developed approach supports object-detection and segmentation annotation based on a small number of manually prepared examples, followed by iterative generation of additional annotations.

The identified research gaps in industrial vision include the handling of highly imbalanced and noisy data, pseudo-label error propagation in semi-supervised pipelines, domain shifts between labelled and unlabelled data, scalable human-in-the-loop annotation, and reliable confidence estimation.

4.3 Anomaly Detection, Fault Detection, Predictive Maintenance

The AD/FD/PdM chapter, led by UniPD and LFoundry, provides a comprehensive review of AI-enhanced diagnostics and prognostics in Smart Manufacturing. Anomaly Detection (AD), Fault Detection and Diagnosis (FDD), and Predictive Maintenance (PdM) are integrated into a single connected framework in which the diagnostic outputs of AD/FDD serve as inputs to prognostic Remaining Useful Life (RUL) estimation. The human-centric vision of Industry 5.0 is further supported through dedicated sections on Explainable AI (XAI) and Uncertainty Quantification (UQ).

The chapter provides a structured catalogue of benchmark datasets—including C-MAPSS and N-CMAPSS for turbofan RUL estimation; the PHM18 dataset on ion-milling tools, jointly relevant

to FDD and PdM in semiconductor manufacturing; the NASA PCoE lithium-ion battery dataset; the PRONOSTIA and IMS bearing datasets; the CWRU bearing dataset; and the Tennessee Eastman Process for FDD and XAI benchmarking. The catalogue covers dataset structure, typical labels, sampling characteristics, and the challenges they expose, including heterogeneous run lengths, multiple degradation modes, hidden failure onsets, and partial observability. State-of-the-art model-based, data-driven, and hybrid methodologies are surveyed for each task, with the scope restricted to tabular and sequential data; high-dimensional perceptual data are left to the companion Image Processing and APM chapters. Separate framework sections describe AD, FDD, and PdM, each combining theoretical background with industrial-application overviews, while a dedicated section discusses XAI and UQ as core enablers of trustworthy industrial AI.

Four AIMS5.0 use cases anchor the review and span a broad maturity spectrum. Anomaly Detection on Production Processes exemplifies the early-stage challenges of problem formulation, KPI specification, and data strategy. Edge Anomaly Detection and Predictive Maintenance demonstrates the deployment of lightweight, interpretable models—gradient-boosted decision trees and TinyML networks—directly on PLCs and embedded gateways, with specialised per-asset models and incremental on-device updates. Tool Breakage Detection in Machine Tool Operations combines FFT-based feature extraction with supervised classifiers (RF, SVM, k-NN, and Logistic Regression) and motif-based segmentation, illustrating the transition from laboratory data to noisy, unlabelled production data. Industrial Production-Ready Anomaly Detection represents a fully operationalised deployment with defined targets and a live pilot, demonstrating the end-to-end industrial AI lifecycle.

The identified research gaps—model transferability across domains and equipment, computational efficiency for edge deployment, data privacy through federated learning, and broader adoption of XAI and UQ—are identified as priorities for the next generation of human-centric, sustainable industrial AI systems.

4.4 Reinforcement Learning, Optimization, Scheduling

The Reinforcement Learning, Optimization and Scheduling chapter (UniPD, Infineon Dresden, NXP, EDI, and partners) reviews how Deep Reinforcement Learning (DRL), Evolutionary Strategies (ES), and Differentiable Physics (DP) can address decision-making problems that classical optimization methods and heuristics cannot tackle at the required scale and speed. The methodological coverage includes Markov Decision Processes and the value, action-value, and advantage functions; the DRL algorithms used in the project (DQN/Double DQN, REINFORCE, A2C/A3C, and PPO); evolutionary methods (ES, CMA-ES) for credit-assignment-free black-box optimization; and differentiable physics as an alternative to RL that backpropagates gradients through simulations, alongside a comparative table of differentiable simulators (Brax, Dojo, NimblePhysics, Taichi, and Warp).

Five industrial use cases anchor the chapter.

The Wafer Lot Scheduling use case at Infineon formalises the Semiconductor Factory Scheduling Problem as a Complex Job-Shop Scheduling Problem under strong stochasticity. It contrasts classical heuristics (FIFO), metaheuristics (ABC, ACO, GA), exact solvers (CP, MIP, LP, ASP), and DRL approaches (DQN, PPO, DDPG, P3O) with different reward, state, and action formulations, and validates CMA-ES as a scalable, credit-assignment-free alternative in a real industrial setting. An Explainable Reinforcement Learning discussion is integrated into this use case.

The Master Planning for Semiconductor Supply Chain Simulation use case addresses the capacitated production–inventory control problem with dual sourcing and shared capacities. Because rapid-event simulation makes repeated LP runs computationally infeasible, a PPO-

based policy with state discretisation and action masking is benchmarked against an (S,Q) base-stock policy across 50 trained agents. The DRL policy achieves a higher cumulative reward and lower aggregate variance and exploits the demand forecast embedded in the state to react proactively to demand shifts. Scalability strategies, including transfer learning, graph neural networks, and mixture-of-experts models, are also outlined.

The Reel Utilization use case applies index-based Deep Controlled Learning to additive manufacturing, framing the problem as an online cutting-stock problem and demonstrating that the DRL-refined index policy outperforms both random assignment and the static index policy.

The Long-Term Rapid Scenario Planning use case at NXP applies DRL to semiconductor supply-chain planning, with DCL matching or outperforming tailored dispatching-priority rules and reducing decision-making latency compared with MP models.

The Differentiable Physics for UR5 Motion Planning use case (EDI) trains a feedforward controller through Nimble Physics using a composite loss covering position, orientation, soft contact, and velocity. In the fixed-socket scenario, it achieves higher placement precision (1.21 mm vs. 4.32 mm), smoother trajectories, and a 24× lower training budget than the PPO baseline, while maintaining comparable success rates. The identified research gaps include generalisation across use cases, human–AI collaboration, and scalability under real-time constraints.

4.5 Robotics, Automation, Process Management

The Robotics, Automation, and Process Management chapter (Infineon Dresden, UniPD, EDI, HUA, and partner companies) sits at the intersection of engineering and operations management and examines how robotics, automation, and data-driven process management jointly enable the transition from highly automated Industry 4.0 environments to human-centric, ML-driven Industry 5.0 systems. The literature review covers operations management (lean production, Quick Response Manufacturing, Factory Physics, and rapid modelling), robotics and material handling (AGVs, AMRs, mobile manipulators, and cobots), automation and intralogistics, and MLOps.

The industrial applications span four areas. Production management is illustrated by the 200 mm wafer fabrication line at Infineon Dresden—one of the first lines worldwide to reach approximately 93% automation—and by SPC, APC, and FDC systems fed by an MES that handles more than one billion sensor readings per day and is extended with machine learning.

A Rapid Modeling Technology (RMT) use case uses a G/G/m queueing representation built in IMPAQT to distinguish capacity-driven from variability-driven bottlenecks at a critical PTX workstation, showing that high process-time and downtime variability, rather than insufficient capacity, drive WIP buildup. A complementary use case applies SARIMAX, XGBoost, LSTM, and CNN models to ultra-short-horizon (sub-shift) WIP forecasting, with XGBoost performing best (RMSE 0.10, R^2 0.54 at a 2-hour horizon). The approach is complemented by SHAP-based explanations and a behavioural Weight-on-Advice study (N = 71) quantifying how explanations influence operator trust.

The intralogistics contributions include a supervised-learning approach to AMR travel-time prediction within a semiconductor fab, which parses fleet logs to learn the mapping between system state and observed travel times, and a low-cost sensor platform for condition monitoring of automated material handling systems (AMHS/OHT). The latter compares consumer-grade IMUs with a reference industrial sensor (KS95) across FOUP hoist operation, FOUP door opening, AGV motion, and calibration shocks, showing that data fusion across low-cost sensors can detect imminent malfunctions cost-effectively.

A robotised pick-and-place use case for the cosmetics process industry (HUA and EDI) combines a UR5e robot, machine vision for grasp-affordance and socket tracking (YOLOv9), MoveIt2 motion planning for the pick phase, and a custom machine-learned controller for dynamic placement into moving carrier sockets. It is complemented by a generate-and-refine workflow that uses LLM-based code generation with Isaac Sim feedback to support rapid task reconfiguration. DAC contributed to this Expert Group from the perspective of process automation and workflow support in industrial AI applications.

In UC18, DAC supported the automation of the planning process, plan-execution monitoring, and the integration of planning tools through the Open Access Platform. In UC07, DAC supported the automation of dataset preparation for computer-vision workflows through the MLOps-based annotation pipeline.

The chapter also documents an MLOps-based semi-supervised data-annotation pipeline with data versioning and automated retraining. A cross-use-case comparison table consolidates these contributions, while the identified future research directions focus on interpretable hybrid models, integrated AI workflows, and tighter integration between management and engineering perspectives.

4.6 Natural Language Processing, Large Language Models

The LLM chapter, contributed by BME, AITIA, TU/e, EDI, Research Studios Austria, Infineon, Signify, Statwolf, and UniPD, examines Large Language Models and AI copilots as collaborators that can handle fragmented information, generate development artefacts, and support compliance and verification tasks within the Industry 5.0 paradigm. The related-work review covers LLMs and agents, Retrieval-Augmented Generation (RAG), code generation, and LLM application architectures.

Eight AIMS5.0 application areas are documented. The Arrowhead Copilots for system-of-systems engineering (BME) comprise a multimodal Arrowhead Expert RAG chatbot that ingests technical documents, scientific papers, and flowcharts at a semantically coherent level of granularity; an Arrowhead Management Copilot implemented as a multi-agent planning engine that decomposes user requests into subtasks, invokes OpenAPI services, evaluates intermediate results, and replans in the event of errors; and an Arrowhead Design Copilot for SysML diagram generation.

The Statwolf use case deploys RAG-based chatbots for support and after-sales services, integrating ticketing-system APIs with retrieval over manuals and FAQs and enabling context-preserving handover to human agents. Reported industrial deployments yield operational cost reductions of 5–10%, a 6% reduction in customer churn, productivity gains of 2–10%, and increases in spare-parts sales of 3–7%.

The Research Studios Austria AI Act self-assessment chatbot combines a RAG pipeline with a curated regulatory corpus—including the EU AI Act, DSA, DMA, HLEG guidelines, and bias and transparency standards—to provide AI developers with personalised compliance reports.

The Signify Engineering Portal RAG chatbot, built on Azure OpenAI “On Your Data” and Azure AI Search over the SharePoint engineering corpus, highlights several open challenges in industrial deployment: SaaS-side data confidentiality, fine-grained document-level access control (particularly in graph-based or RAPTOR-style hierarchical RAG), unified handling of multimodal and distributed content, incremental re-indexing, and prompt design for constraining factuality.

The TU/e factual-correctness pipeline for natural-language-to-SPARQL translation uses an open-source LLM to generate *k* candidate queries and a BERT-based semantic-similarity model with a small MLP to rank them against the natural-language input, with correctness verified against expert-crafted queries over knowledge graphs.

The EDI LLM-driven robot-programming subsystem for AIMS5.0 Use Case 6 is a code-generation pipeline that separately produces task-simulation configuration scripts and high-level robot-control scripts that invoke vision, motion-planning, and grasp-estimation components. Each script is verified against randomised simulations, with performance measured using simulation-generation success rate, partial-success rate, and script-generation success rate.

The Infineon contribution describes the AIMS5.0 Open Access Platform: an LLM-supported ontology and knowledge-graph environment aligned with the Digital Reference model, in which Chowlk and OpenRefine workflows are augmented with LLM-driven schema extraction and natural-language-to-SPARQL/Cypher querying, thereby lowering the technical barrier for non-expert users.

The Research Studios Austria AI-Gym combines Apache Zeppelin, the Eclipse Arrowhead Adapter, and live IIoT data streams to evaluate LLMs under realistic and secure industrial conditions. Document fragmentation, access control, factuality, and legacy-system integration are identified as key open challenges, while hybrid RAG, multi-agent, and verification architectures are highlighted as a promising path towards human-centric LLM adoption.

5. Conclusions

This deliverable consolidates the final technological results of WP2 and demonstrates the progress made in transforming the AIMS5.0 AI Toolbox from a common concept for industrial AI development into a broad and practically applicable collection of methods, tools, services, integration mechanisms, and industrial solutions. The resulting AI Toolbox spans multiple levels of AI development, from individual algorithms and reusable tools to integrated applications and methodological guidance, while supporting both deep- and shallow-integration approaches. Its purpose is not to replace existing AI ecosystems, but to make heterogeneous AI technologies easier to discover, evaluate, combine, deploy, and reuse in industrial environments.

A major result of WP2 is the breadth and maturity of the technological portfolio assembled around the Toolbox. The developments cover computer vision, language and generative AI, industrial data analytics, anomaly and fault detection, predictive maintenance, robotics and automation, optimisation and scheduling, simulation, semantic technologies, explainability, evaluation, and trustworthy AI. Rather than remaining isolated research prototypes, many of these technologies are directly connected to AIMS5.0 industrial use cases, where they are implemented, evaluated, and adapted to concrete manufacturing requirements. The Toolbox therefore acts as a bridge between AI research and industrial deployment, providing a structured environment in which technological advances can be transferred across partners, application domains, and use cases.

The reported results show that this approach can generate measurable industrial value. Across the technological developments, the project addresses objectives such as improved prediction and classification performance, reduced quality-control effort, shorter AI-system reconfiguration and development cycles, increased productivity, improved manufacturing uptime, reduced data requirements when transferring models to new machines, faster time-to-market, and lower energy consumption. At the same time, several contributions explicitly address industrial requirements that extend beyond raw model accuracy, including explainability, reliability, human oversight, interoperability, deployment on resource-constrained edge devices, semantic consistency, and integration into existing industrial infrastructures.

These characteristics are particularly important for strengthening European industrial competitiveness. European manufacturing competes not only through production volume but also through technological excellence, high-value engineering, flexibility, quality, sustainability, reliability, and the ability to manufacture increasingly complex and customised products. The AIMS5.0 results contribute to these strengths by reducing the effort required to introduce AI into industrial processes and by making advanced AI capabilities more reusable across factories, machines, engineering workflows, and industrial sectors. Shorter development and reconfiguration cycles can enable European companies to respond more rapidly to changing markets and production requirements, while improvements in quality control, predictive maintenance, planning, and process optimisation can increase productivity and reduce unnecessary costs and downtime.

The reusable and composable nature of the AI Toolbox is particularly relevant in this context. Industrial AI solutions are often costly because similar capabilities must be independently developed, integrated, validated, and maintained for individual factories or use cases. A common repository of reusable AI services, documented interfaces, deployment mechanisms, implementation guidance, and application examples can reduce this duplication of effort. It enables knowledge generated by one partner or industrial domain to be shared with others and provides a foundation for transferring successful AI solutions to new manufacturing

environments. The project therefore contributes not only individual algorithms but also an enabling infrastructure for accelerating industrial AI innovation at European scale.

A further competitive advantage arises from the strong connection between AI, sustainability, and Industry 5.0 throughout the project. Increased industrial competitiveness cannot rely solely on greater automation; future European production systems must simultaneously become more resource-efficient, resilient, adaptable, and human-centric. The technological results presented in this deliverable support this transition through methods for more efficient production planning, predictive maintenance, reduction of quality losses and scrap, intelligent process adaptation, efficient use of data, and optimisation of energy and material consumption. Trustworthy and explainable AI, human-in-the-loop approaches, and technologies designed for integration into real industrial environments further support the responsible adoption of AI, rather than treating automation as an objective in itself.

The work of the WP2 Expert Groups reinforces this long-term perspective. By bringing together expertise in Machine Learning, Image Processing, Anomaly and Fault Detection, Predictive Maintenance, Reinforcement Learning, Optimisation and Scheduling, Robotics and Automation, and Large Language Models, the project has created a cross-domain knowledge base that extends beyond individual use cases. The resulting contributions combine state-of-the-art knowledge, industrial experience, original project results, and identified research gaps, thereby helping to establish common technological directions for the next generation of industrial AI systems.

Importantly, the results also demonstrate that industrial AI does not have a single technological form. In some applications, advanced foundation models and Large Language Models provide new possibilities for knowledge access, engineering support, and automation. In others, traditional machine-learning methods remain highly competitive because of their interpretability, modest computational requirements, and suitability for certification and deployment in resource-constrained environments. Computer vision, differentiable simulation, optimisation, robotics, anomaly detection, semantic technologies, and hybrid physics- and data-driven approaches each provide complementary capabilities. The strength of the AIMS5.0 approach lies in bringing these technologies together within a common industrial framework rather than promoting a single class of AI methods.

This D2.4 deliverable therefore represents more than a collection of individual technological achievements. It demonstrates a model for collaborative European industrial AI development in which research organisations, technology providers, and industrial partners jointly develop, validate, document, and share reusable solutions. Such collaboration enables fragmented technological advances to be transformed into transferable industrial capabilities and shortens the path from research results to deployment.

The continuation of this approach beyond AIMS5.0 can provide lasting value. The AI Toolbox, its technological assets, accumulated implementation experience, and cross-partner expert networks establish a foundation on which future industrial AI solutions can be developed more rapidly and with lower integration effort. Further reuse, extension, and industrial validation of these results can increase their impact across European manufacturing ecosystems and supply chains.

Ultimately, the contribution of AIMS5.0 is not only to demonstrate that AI can improve individual manufacturing processes, but also to show how AI technologies can be systematically transformed into reusable, trustworthy, deployable, and industrially meaningful capabilities. By accelerating innovation, strengthening knowledge transfer, supporting sustainable production, and reducing barriers to industrial AI adoption, the results of WP2 contribute to a more technologically capable, resilient, and competitive European manufacturing sector. They

represent an important step towards the broader AIMS5.0 vision, in which European industry combines advanced AI with human expertise, sustainability, and engineering excellence to remain at the forefront of global industrial innovation.