Artificial Intelligence in Manufacturing leading to Sustainability and Industry5.0

Project Acronym:

AIMS5.0

Grant agreement no: 101112089

Deliverable no. and title	**D2.1 – Development and refinement plans for AI tools methods and algorithms: Preliminary descriptions of the AI tools methods, algorithms and tools**	
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	v2.0	
Date	26/11/2023	
Status	FINAL	
Dissemination level	Public (PU)	

Copyright: AIMS5.0 Project Consortium, 2023

Authors

Participant no.	Part. short name	Author name
34	BME	Pal Varga
38	EDI	Kārlis Freivalds
37	STATWOLF	Gian Antonio Susto
44	TU/e	Mike Holenderski
5	FMX	Karl-Benedikt Reith
43	SIGNIFY	Marc de Samber
39	NXP	Eric Weijers
17	KAI	Corinna Kofler
34	BME	Dániel Ficzere
33	AITIA	Tamás Bordi
46	GUT	Mateusz Growth
39	NXP	Kenneth Braakman
40	PCL	Ahmet Gozutok
40	PCL	Olga Kattan
25	IDEKO	Leire Agirretxe
32	HUA	Iraklis Varlamis
32	HUA	Christos Chronis
37	STATWOLF	David Dandolo
34	BME	Csaba Hegedűs
10	IFD	Germar Schneider

Document History

Version	Date	Author name	Reason
v0.1	21.07.2023	Pal Varga	Initial Version
v0.2	23.09.2023	Gian Antonio Susto	Added contribution for Predictive Maintenance and for Italian cluster activities
v0.3	27.09.2023	Karl-Benedikt Reith	Added Fabmatics Contribution to Section 3
v0.4	28.09.2023	Marc de Samber	Added Signify contribution in section 3 and 5
v0.5	29.09.2023	Kārlis Freivalds	Added EDI contribution to section 3
v0.6	29.09.2023	Eric Weijers	Added contribution for NXP/Semaku/TUe activities
v0.7	29.09.2023	KAI	Adding KAI and Infineon contribution
v0.8	03.10.2023	Dániel Ficzere	AI Toolbox concept
v0.9	07.10.2023	Gian Antonio Susto	Added contribution on Section 3.1
v1.0	08.10.2023	Pal Varga	Editing, merging UniPD content
v1.11	09.10.2023	Tamás Bordi	Added AITIA contribution to section 3
v1.13	11.10.2023	Mateusz Groth	Added GUT/TTTech contribution to section 3
v1.14	13.10.2023	Kenneth Braakman	Added contribution for NXP/Semaku/TUe activities in section 3.5 and 5.2
v1.15	13.10.2023	Ahmet Gozutok / Olga Kattan	Added contribution for PCL to section 3.10.
v1.16	13.10.2023	Kārlis Freivalds	Added section “5.3 Tool for robot control task execution”. Slightly updated EDI contribution in 3.4.2
v1.17	13.10.2023.	EDI	EDI section
v1.18	23.10.2023.	lagirretxe	UC5 section
v1.19	23.10.2023.	Pal Varga	BME section
v1.20	25.10.2023	Iraklis Varlamis & Christos Chronis	Added section “3.13 HUA Contribution” and section “5.4.1 Robot Control and Algorithms integration to the general pipeline “
v1.21	28.10.2023	Gian Antonio Susto & David Dandolo	Added section “5.3 5.3Unsupervised Anomaly Detection”
v1.22	06.11.2023.	Csaba Hegedűs	Added introduction and reviewed contributions
v1.23	8.11.2023	Karl-Benedikt Reith	Adjusted FMX parts
v1.24	10.11.2023	Pal Varga, Germar Schneider	Merged previously misaligned versions, edited for better cohesion including WP6 contributions
v1.25	16.11.2023	Pal Varga	Editing for review
v2.0	24.11.2023	Pal Varga	Editing for final review

Publishable Executive Summary

The deliverable D2.1 of the AIMS5.0 project describes the planned AI methods, algorithms and tools that is planned to be developed and/or used in the AIMS5.0 project, as it is envisaged at the first internal technical milestone. The objective is to share the knowledge that can be useful for general industry and academia for addressing their own related challenges. Keeping in mind the sensitive nature of the methods used by some of the AIMS5.0 industrial partners, it omits the details about the particular configuration and fine-tuning that gives the authors their competitive advantage on the market. Instead, it provides a description of the AI tools, methods and algorithms in isolation, but with sufficient detail to allow application in similar use cases.

The objective of this document is to summarize the problems to be addressed in AIMS use cases, and the machine learning solutions that are being considered for them. This is a preliminary report based on the initial analysis done by partners and solution designs are still being established.

The main objectives of the Work Package 2 (AI tools, methods and algorithms for sustainable industrial processes) are threefold. First, to provide AI tools, methods and algorithms as a Toolbox, taking the industrial needs and practices of the other technology WPs in AIMS5.0. Second, to extend the current SoTA on utilizing AI tools, methods and algorithms in various stages and levels of sustain-able industrial processes. Third, to present use-case specific integration methods that lead to actual, working prototypes in use-case WPs - and based on that, the creation of integration guidelines is also a result of this endeavour to improve better time-to-market.

In general, Work Package 2 in AIMS5.0 plans to reach their common goals by

- aggregating requirements from the use cases and structuring the requirements based on target domain, data volume and feature needs, as well as the issue to be tackled by the algorithms, methods or tools;
- collaborating with other WPs in providing widely accepted, verified and validated AI methods, algorithms and tools for sustainable industrial processes;
- covering gaps based on the aggregated and structured requirements, in relation to methods and algorithms to be either created or refined – with the identified competence of partners;
- developing and refining methods and algorithms to fit the needs of European Industry, especially those that can be demonstrated through the use cases: this may lead to pioneer approaches and best practices for various industrial domains;
- supporting the integration of algorithms into specific use-case scenarios (covering their various applications for Manufacturing-Responsive Development, Intelligent Manufacturing, Intelligent Supply Chain, Other Industrial applications, and for Facility Management 5.0) through the collected, refined and developed methods and algorithms, helping with best practices as well as concrete implementation support, as well as

presenting use-case specific results as gains reached by the applied methods and algorithms to ensure sustain-able industrial processes.

D2.1 has a simple structure. In Chapter 2, each partner or use case-grouping provide description of the requirements and mission statement. In Chapter 3, the implementation plans are presented, describing the algorithms, methods, tools and approaches taken by the partners. Chapter 4 provides and initial overview on the AI toolbox approach that will utilize the output of individual Chapter 3 contributions within use cases during the project. Finally, Chapter 5 is a summary of the in-progress use case works are provided as a table, mapping with AI/ML technology keywords to facilitate the elevation of commonalities and cohesion within WP2, as well as minimizing semi-independent parallel work on similar technology advancements that could otherwise strengthen each other.

Table of contents

1. Introduction 7
1.1 Objective of WP2 7
1.2 The generic plan for WP2 7
1.3 Objective and scope of the document 7
1.4 Public and project-internal deliverable parts 8
2. Partner and use case plans for individual and collaborative contributions in WP2 9
2.1 LFoundry - Statwolf Data Science - Università degli Studi di Padova contribution 9
2.1.1 eXplainable Artificial Intelligence (XAI) approaches for unsupervised anomaly detection 9
2.1.2 Continual Learning and scalability 10
2.1.3 Dealing with heterogenous data 10
2.2 Signify contribution 11
2.2.1 Intelligent Scheduler 11
2.2.2 Vision based process execution and product quality evaluation 11
2.2.3 Fully automated robotics product assembly using Computer Vision 12
2.2.4 Analysis of sparse data from Operations 12
2.2.5 Personalized Operator guidance using LLMs 12
2.2.6 Intelligent tendering and propositions 13
2.3 Institute of Electronics and Computer Science (EDI) contribution 13
2.3.1 Program synthesis for robot action planning 13
2.3.2 Tools for robot control task execution 14
2.4 NXP, Semaku and TU/e contribution 14
2.4.1 Using Deep Reinforcement Learning (DRL) for planning 15
2.4.2 Using GNNs for quality incident prediction 16
2.5 KAI's and Infineon's planned contribution to AI tools, methods and algorithms for sustainable industrial processes 17
2.5.1 Data-centric development approaches 17
2.5.2 Model-centric development approaches 18
2.5.3 Scalable ML pipeline development 18
2.5.4 Model monitoring concepts 19
2.5.5 Explainability 19
2.6 BME contributions 19
2.7 AITIA contributions 20
2.8 Gdansk University of Technology (GUT) and TTTech contribution 21
2.9 RSA FG contribution 22
2.10 PCL Contribution 23
2.11 IDEKO, SORALUCE, GOIMEK and SAVVY contribution 24
2.11.1 Identification of necessary components and data 25
2.11.2 Techniques for anomaly detection 25
2.11.3 Provision of results and insights 26
2.12 HUA Contribution 26
2.12.1 Code base and Experimental Environment 27
2.12.2 Composite task deaggregation and machine learning pipelines 27
2.12.3 The use of Deep Reinforcement Learning for task automation 28

2.12.4 Behavior Cloning Learning and Immitation Learning 29
2.13 Infineon Technologies Dresden (IFD) contribution in cooperation with IFAG, IFAT, FuH, Systema, HSZG, WUW, HoST, TUD, FMX, UNPD, PFV, FhG IISB, ViF, UzL 29
2.13.1 UC09 – Energy-aware Scheduling of Batch Processing Machines in Wafer Fabs (FuH, IFD, IFAT, Systema, ViF) 31
2.13.2 UC10 – MES Airbus (Systema, IFD, IFAT) 33
2.13.3 UC12 Human-Centered AI for the Optimization of Robust and Competitive Semiconductor Manufacturing Networks and Increased Sustainability AI (WUW, HSZG, Systema, IFD, IFAT, IFAG) 33
2.13.4 UC13 Intelligent Sensors improving Robustness of Automated Wafer Transportation and Storage Systems (HoST, IFD, IFAG, IFAT, FMX, TUD) 34
2.13.5 UC14 Speeding-up the Benefit of Automation by Virtual Commissioning through Digital Twins (FMX, IFD, IFAT, IFAG, TUD, HoST) 34
2.13.6 UC15 Productionized Machine Learning models at scale (IFD, IFAT, IFAG, UNPD, STATWOLF, LFO, KAI 34
2.13.7 UC16 AI enhanced production and performance of RF transceivers for civilian SATCOM and radar applications (THALES, FORTH) 35
2.13.8 UC17: Intelligent Contamination Management & Prescriptive Control Plans and Process Flows (FhG IISB, PFV, IFD, IFAT, IFAG), Systema) 35
2.13.9 UC20: Harmonization of Automated Digitized Equipment Control (IFAT, IFD, TUD) . 35
3. Preliminary descriptions of the AI tools methods and algos for WP2 38
3.1 Tool for Predictive Maintenance 38
3.2 Unsupervised Anomaly Detection 39
3.3 AI tools methods and algorithms as used in Manufacturing for Signify use case 40
3.4 AI tools methods and algorithms as used in supply chain planning and quality incident traceability for NXP/Semaku/TUe use case 40
3.5 Tool for robot control task execution 41
3.5.1 Robot Control and Algorithms integration to the general pipeline 41
3.6 Utilizing the copilot concept in AIMS5.0 43
3.7 Details about PrototAIp 45
3.8 Methods and Tools for Facility Management 5.0 (ViF, IFD, IFAT) 47
3.9 MLOps Techniques for Industry5.0 48
4. The vision of the AI Toolbox 50
4.1 Questionnaires 50
4.1.1 Questionnaire 1 50
4.1.2 Questionnaire 2 51
4.2 AI toolbox objectives 52
4.3 AI toolbox concept 53
4.3.1.1 Application scale 53
4.3.1.2 Tools scale 53
4.3.1.3 Algorithmic scale 54
4.3.1.4 Methodology scale 54
4.3.2 AI Toolbox lifecycle 55
4.3.3 AI Toolbox implementation 55
4.3.4 Edge AI concerns 55
5. Conclusions 56
6. References 59

1. Introduction

1.1 Objective of WP2

The objectives of WP2 are the following:

- Provide AI tools, methods and algorithms as a Toolbox, taking the industrial needs and practices, as well as WP1, WP3 and WP4 results into special consideration. Based on the compiled industrial requirements, the resulting toolbox take human-like reasoning and explainability into consideration, applies hybrid, configurable AI algorithms and methodologies, and enables dynamic selection and integration of AI algorithms.

- Extend the current SoTA on utilizing AI tools, methods and algorithms in various stages and levels of sustain-able industrial processes. This includes the actual application of AI tools and algorithms for manufacturing-responsive development, intelligent manufacturing, intelligent supply chain, facility management and other industrial applications, in a way that takes WP7 and WP8 results into account as well.

- Present use-case specific integration methods that lead to actual, working prototypes in WP5 and WP6 - and based on that, the creation of integration guidelines is also a result of this endeavour to improve better time-to-market.

1.2 The generic plan for WP2

WP2 plans to reach their common goals by

- aggregating requirements from the use cases and structuring the requirements based on target domain, data volume and feature needs, as well as the issue to be tackled by the algorithms, methods or tools;
- collaborating with other WPs in providing widely accepted, verified and validated AI methods, algorithms and tools for sustainable industrial processes;
- covering gaps based on the aggregated and structured requirements, in relation to methods and algorithms to be either created or refined – with the identified competence of partners;
- developing and refining methods and algorithms to fit the needs of European Industry, especially those that can be demonstrated through the use cases: this may lead to pioneer approaches and best practices for various industrial domains;
- supporting the integration of algorithms into specific use-case scenarios (covering their various applications for Manufacturing-Responsive Development, Intelligent Manufacturing, Intelligent Supply Chain, Other Industrial applications, and for Facility Management 5.0) through the collected, refined and developed methods and algorithms, helping with best practices as well as concrete implementation support, as well as
- presenting use-case specific results as gains reached by the applied methods and algorithms to ensure sustain-able industrial processes.

1.3 Objective and scope of the document

This deliverable describes the planned AI methods, algorithms and tools that will be developed and/or used in the AIMS5.0 project. The objective is to share the knowledge that can be useful for general industry and academia for addressing their own related challenges. Keeping in mind

the sensitive nature of the methods used by some of the AIMS5.0 industrial partners, it omits the details about the particular configuration and fine-tuning that gives the authors their competitive advantage on the market. Instead, it provides a description of the AI tools, methods and algorithms in isolation, but with sufficient detail to allow application in similar use cases. The objective of this document is to summarize the problems to be addressed in AIMS use cases, and the machine learning solutions that are being considered for them. This is a preliminary report based on the initial analysis done by partners and solution designs are still being established.

The structure of the document follows a simple approach. In Chapter 2, each partner or use case-grouping provide description of the requirements and mission statement. In Chapter 3, the implementation plans are presented, describing the algorithms, methods, tools and approaches taken by the partners.

Chapter 4 provides and initial overview on the AI toolbox approach that will utilize the output of individual Chapter 3 contributions within use cases during the project.

In Chapter 5, a summary of the in-progress use case works are provided as a table, mapping with AI/ML technology keywords to facilitate the elevation of commonalities and cohesion within WP2, as well as minimizing semi-independent parallel work on similar technology advancements that could otherwise strengthen each other.

1.4 Public and project-internal deliverable parts

Originally the AIMS5.0 project partners planned to deliver a public part (D2.1) and an NDA-protected project-internal part (D2.5) of this first WP2 milestone report.

Nevertheless, the partners decided that all the deliverable content can be public, meaning D2.1 covers all that would be planned for D2.5, hence there is no need of a project-internal, NDA-protected D2.5 document.

2. Partner and use case plans for individual and collaborative contributions in WP2

This section briefly summarizes the planned individual and collaborative contributions of partners and use case providing groups.

2.1 LFoundry - Statwolf Data Science - Università degli Studi di Padova contribution

The Italian cluster, composed by LFoundry (LFO), Statwolf Data Science (STATWOLF) and Università degli Studi di Padova (UNIPD) will collaborate on several research directions in the field of Machine Learning, that will be listed in the following sub-sections.

All the reported activities will provide algorithmic approaches and tools that will be leveraged in WP6, UC15, *Productionized Machine Learning models at scale'*.

2.1.1 eXplainable Artificial Intelligence (XAI) approaches for unsupervised anomaly detection

In the realm of manufacturing, the ability to swiftly and accurately identify anomalies holds paramount importance. It underpins the overarching goals of ensuring impeccable product quality, minimizing defects, and optimizing production processes. While traditional unsupervised anomaly detection techniques have proven effective, they often operate as inscrutable black boxes, leaving manufacturers yearning for transparency and interpretability. Enter Explainable Artificial Intelligence (XAI), a transformative force that is reshaping the manufacturing landscape by infusing a new level of clarity and understanding into anomaly detection. In this area, we have preliminary results developed before entering the project of an approach called ‚DIFFI - Interpretable Anomaly Detection with DIFFI: Depth-based feature importance of Isolation Forest' [Carletti2023], which is an ad-hoc XAI approach for the popular Isolation Forest [Liu2008] approach. DIFFI is a global and local explainability approach, the first one presented in the literature to provide feature importance for the Isolation Forest. We are planning in the project to expand this area by:

- Developing similar approaches for feature importance for other isolation-based approaches for unsupervised anomaly detection, like the Extended Isolation Forest [Hariri2019];
- To consider second-order interactions in classical XAI approaches.

Another preliminary result related to XAI have been achieved by the team in [Dandolo2023]. In this paper, we propose Accelerated Model-agnostic Explanations (AcME), an interpretability approach that quickly provides feature importance scores both at the global and the local level. AcME can be applied a posteriori to each regression or classification model based on tabular data. Not only AcME computes feature ranking, but it also provides a what-if analysis tool to assess how changes in features values would affect model predictions. We evaluated the proposed approach on synthetic and real-world datasets, also in comparison with SHapley Additive exPlanations (SHAP), the approach we drew inspiration from, which is currently one of the state-of-the-art model-agnostic interpretability approaches. We achieved comparable results in terms of quality of produced explanations while reducing dramatically the computational time and providing consistent visualization for global and local interpretations. In this project, we aim to expand AcME to unsupervised anomaly detection tasks.

2.1.2 Continual Learning and scalability

Continual learning, also known as lifelong learning or incremental learning, is a paradigm in machine learning and artificial intelligence that seeks to enable models to accumulate knowledge and adapt to changing data distributions over time. Unlike traditional machine learning models that are trained on static datasets, continual learning models are designed to learn incrementally from new data while retaining previously acquired knowledge. This enables them to continuously improve and adapt without the need for periodic retraining from scratch. Continual learning is crucial in applications (like manufacturing) where data evolves, such as in dynamic environments, and it holds great promise for building more adaptive and intelligent AI systems that can leverage past experiences to make better decisions in the face of evolving information.

In this context, we developed before the start of the project, we develop a continual learning framework to scale Deep Learning approaches [DallePezze2023]. As said, continual Learning aims to learn from a stream of tasks, being able to remember at the same time both new and old tasks. While many approaches were proposed for single-class classification, multi-label classification in the continual scenario remains a challenging problem. For the first time, we study multi-label classification in the Domain Incremental Learning scenario. Moreover, we propose an efficient approach that has a logarithmic complexity with regard to the number of tasks, and can be applied also in the Class Incremental Learning scenario. We validate our approach on a real-world multi-label Alarm Forecasting problem from the packaging industry.

In the project, we aim at extended the continual learning framework to Anomaly Detection. We will develop techniques that scales and compresses the original images using Super Resolution models. To validate the proposed approach, we will leverage real-world datasets of images with pixel-based anomalies, typically associated with particles and defects.

2.1.3 Dealing with heterogenous data

Dealing with heterogeneous data, encompassing a mixture of images, time series, and tabular data, presents a complex challenge in machine learning. Such scenarios are really frequent in semiconductor manufacturing. Combining and effectively utilizing diverse data types can enhance model performance and provide deeper insights, but it also introduces unique difficulties due to the dissimilar nature of these data sources.

To address this challenge in the project, we will leverage deep learning approaches, considering techniques tailored to handling heterogeneous data. These methods include multi-modal networks, feature engineering to extract relevant features from each data type, transfer learning with pre-trained models, attention mechanisms, ensemble methods, careful data preprocessing, regularization techniques, and custom loss functions. By incorporating these strategies into deep learning architectures, it will becomes possible to leverage the strengths of each data type and create more versatile and accurate machine learning models capable of handling a wide range of data sources effectively.

In this context we will also consider scenarios where data coming from different tools are considered: this is a typical scenario in semiconductor manufacturing, where we would like to have machine learning-based technologies deployed in different machines performing the same process. While in this area, transfer learning [Weis2016] has proven successful in many cases, particularly with computer vision and homogenous data inputs, it can be challenging to apply to heterogeneous data. Motivated by the need to transfer and standardize established processes to different, non-identical environments and by the challenge of adapting to heterogeneous data representations, we will leverage Deep Learning approaches to tackle such issue. In particular, we will resort to Domain Adaptation Neural Networks and expand the works done in

[Gentner2021] to scenarios where different input data are available. If successful, this approach will also have an impact in scalability (as discussed in the previous sub-section.

2.2 Signify contribution

Signify is developing and validating Artificial Intelligence tools for Operational Excellence in Manufacturing. This is in the context of the ambition to implement Digitization in the End-to-End Operations and Manufacturing. Digitization is an essential condition for quantified data analysis and intelligence. The target operational sites that will benefit from the work are all located in Europe (Belgium, Poland, Hungary) and cover a wide range of lighting product types. We aim for solutions that are platform agnostic (fit all), and for that reason we pay less attention to the platform method (Eclipse-Arrowhead) studied in WP3. While mainly contributing however in WP2 and WP1.

We have defined a long list of potential use cases that can benefit from A.I. methods, and from that long list we have selected a preliminary list based on potential impact, reason-to-believe, maturity and absorption (transfer) likelihood.

For the topic of 'intelligent scheduler' and 'Analysis of sparse data from Operations' we collaborate with partner TU/e (NL). We still foresee to team up with partners that contribute in WP2 and WP1 and that provide AI sensors, tools or offer testbed environments.

A.I. topics as prioritized (short description):

2.2.1 Intelligent Scheduler

Being able to deliver customer orders in time, despite potential disruptions (such as because of supply chain issues, downtime of equipment, human factor factors), is one of the most important KPIs in business. For Signify, in our target application spaces, we often deal with high numbers of varying orders having small volumes and diverse products. Our selected primary use case deals with 3D printed luminaires, with Operations in Belgium and Hungary.

We have at this time a scheduler that arranges order execution based on material (filament) in the Bill-of-Materials, and this scheduler is rule-based, single stage (filament as parameter only), relatively slow and requiring full recalculation in case of changes or anomalies.

In a growing market and hence higher order complexity, we aim to extend the scheduler into an intelligent and adaptive tool that can deal with multiple parameters, based on reinforcement learning algorithms. Adaptations should deal with e.g., short term on-site supply chain issues, changes in order intake, disruptions due to machine downtime or missing personnel, etc.

2.2.2 Vision based process execution and product quality evaluation

Production yield loss and quality issues due to human errors are not only productivity issues but also lead to human dissatisfaction and mental stress. Our use case is in human-based and human centric (manual) assembly of luminaries.

We want to automatically detect quasi-real time if an error is made, and provide the outcome to the human worker, allowing for his/her immediate corrective action. The approach is by using a low-cost Vision-based method to evaluate the process execution (by human, potentially

supported by machine collaboration) with an immediate feedback to the human worker. This serves also as a surrogate indicator of the correctness (hence quality) of the assembled product.

Visual data will be collected by a single camera, and AI algorithms are needed to deal with our high diversity (flexible) production, deal with low training conditions (e.g., no prior physical data for not-made-before products) and meeting the required sub-second feedback to the human (so, AI algorithms have to run on a low-cost Edge processor). We foresee that various known AI algorithms will have to be combined and interlinked.

2.2.3 **Fully automated robotics product assembly using Computer Vision**

We want to explore the boundaries, opportunities and limitations of the fully automated assembly of luminaires, keeping in mind that the throughput time and flexibility as provided by human workers needs to be matched (and outperformed) while meeting a good ROI.

A requirement (target) is to have ultimate flexibility with robotics assembly (human-less or human-low), within a product family. So, being able to automatically assemble a product first-time-right and without prior ML training.

AI based Computer Vision tools are required for fast and flexible object detection (of parts and half-fabricates), used as input to control the robots in the assembly tasks. AI is essential when dealing with not-seen-before parts, and therefore product designs and BOM parts are artificially generated starting from digital CAD data.

2.2.4 **Analysis of sparse data from Operations**

Although our Operations provide massive amounts of data, essential data might still be missing, or certain data sets might be sparse. E.g., data on the performance/quality of individual equipment might be ‘hidden’ in data that is rather related to a set of equipment.

We use here to use case of the fabrication of 3D printed luminaire parts. Groups of 3D printers in a rack set-up are used in various combinations to execute orders. Analysing the quality outcome of the production of multiple orders of products by a groups of printers should allow to draw conclusions on the performance level of the individual printers (for specific part types), giving the opportunity to do targeted maintenance or to schedule demanding print jobs to the ‘best performing, best suited’ printers.

Mostly, big data sets are analysed afterwards, as to adapt and optimize future processing/operations. This does not provide the option for immediate quality feedback towards the processes. We strive however for real-time information that allows immediate actuation (corrective actions), having a major positive impact on yield levels and prevent scrap. Therefore, we want to combine actual data with historical data in a real-time way. So, combining incoming data with stored data with ‘immediate’ data processing into information might require powerful but low power consuming AI algorithms (processed on Edge processor rather than in the cloud).

2.2.5 **Personalized Operator guidance using LLMs**

Factories experience potential yield loss and productivity loss due to declining expert knowledge and issues with labor force availability. E.g., during holiday season the experienced factory crew must be replaced with temporary workers that are less experienced, have a different language knowledge, are from a different ethnicity, are much less experienced or less technical, etc. One worker might perform based on a bullet list of instructions while a less

experienced colleague might require a full textual description including graphics, pictures or video instructions.

We want to develop and evaluate methods based on generative LLMs (potentially fine-tuned) combined with our own knowledge corpus (e.g., graphics and pictures are re-used from our documentation), as to optimize the operator instructions for high quality outcome (translation of written or spoken instructions/guidance). Instructions can be continuously adapted for each individual in an automated way, using a feedback loop based on the quantified result (better quality and yield) or emotional judgement by the human. For the latter purpose a bi-directional speech or text User Interface to the LLM is required.

2.2.6 **Intelligent tendering and propositions**

We aim at largely automating and optimizing the process of Tendering (offers). This is in our topics list an example of a front-end application for Operational Excellence. The method will be based on digitizing massive amounts of product data sheets of products, combined with all supporting documentation, information and expert knowledge. The final documentation also tuned to the specific target customer using LLMs.

This will allow quick and efficient preparation of personalized/customer targeted tenders, only requiring a final verification by a human expert before sending out to the specific customer.

2.3 Institute of Electronics and Computer Science (EDI) contribution

The Institute of Electronics and Computer Science (EDI) will contribute to developing algorithms and tools for robot control. The goal is to develop a modular and reconfigurable AI-based robotic system to automate a conveyor feeding processes for irregularly distributed bottles of different sizes, weights, and types, with a high degree of effectiveness and efficiency, such that the system and its components can be easily transferred to other operational or usage environments, thus enhancing circular reuse of the parts and replacing manual work in the production process. EDI will develop a robotic action planning module capable of high-level task instruction and context-aware task execution. Detailed results on AI algorithms, methods and tools will be refined to fit the industrial process phases, automation and supply chains.

2.3.1 **Program synthesis for robot action planning**

The goal is to come up with methods for automatic synthesis of generalized program-based robotic policies, and to deploy them in robots to perform context-aware action planning tasks. Robots operating under these policies will be more reliable, in the sense of providing higher safety and more predictable behaviour, including better generalization to unexpected circumstances. Better generalization should also facilitate the learning of new tasks from fewer demonstrations. To improve the explainability of the synthesized code, we divide the main robot task into subgoals. This approach, with functional level goals, where portions of the code can be associated with their functional meaning or immediate goal, provides a better understanding of the system and the decisions made by the robot.

A recent trend is using Large Language Models (LLM) for code synthesis, where OpenAI Codex (Chen et al. 2021) is a prominent example. It is a descendant of OpenAI's GPT-3 model, fine-tuned for use in programming applications including code auto-completion (GitHub Copilot). Such models use a natural language description of the programming task or the preceding context of the partially written program to generate the code. The basic approach does not account for functional correctness of the synthesized programs, it works like machine

translation that translates a task description to program text. There are attempts to improve the correctness of the generated programs (Shi et al. 2022, Poesia et al. 2022) but, still, the reliability has to be improved to apply the generated code for robotic tasks.

The objectives are to apply LLM based program synthesis to learn robot control, to extend the LLM code synthesis methods to robotic tasks, to develop a robot control mechanism operating with automatically learned program-based policies, and to demonstrate the achieved results on long horizon action planning tasks for object manipulation planning in UC6 and beyond.

2.3.2 Tools for robot control task execution

EDI will develop a workflow to support the integration of AI algorithms into an automated conveyor feeding robotic system being developed at WP5 and UC6. It should achieve rapid reconfigurability of the system, which will be supported by both a flexible middleware framework and adapted deep-learning-based algorithms, and shall lead to increased productivity, efficiency, and product quality. The robotic system will consist of several highly configurable AI software modules supported by respective components:

a) robot control integration, based on ROS, supporting a deep learning-based perception and control system, to provide the robots with flexible and dynamic perceiving, comprehending, and control abilities (sensory data processing, representation and interpretation, and low-level action implementation).
b) robotic simulation for learning and validation of robot control algorithms,
c) action planning module together with tools for training associated deep learning algorithms therein,
d) high-level AI task instruction with context-aware task execution,
e) tools for performance validation and quality metrics evaluation

2.4 NXP, Semaku and TU/e contribution

Machine learning is widely used for optimizing modern digital systems, addressing our inability to model these systems from rst principles due to their complexity and the high dimensional data they generate. However, machine learning models, such as neural networks, often contain millions of parameters, making it difficult to understand their behaviour and to trust their output. Moreover, they require large amounts of labelled data, which is not always available. There are machine learning approaches targeting these challenges (e.g., surrogate modelling or unsupervised learning), but they are often limited in their application. Symbolic AI approaches, on the other hand, are often explainable and can leverage expert knowledge to deal with few labelled samples, but struggle with noisy and high dimensional data. TUE plans to combine machine learning methods (for extracting information from structured and high-dimensional data from the manufacturing systems and the supply chain) and symbolic AI methods (for reasoning with the extracted information and expert knowledge provided by the industrial partners by means of knowledge graphs) to address various planning, scheduling and optimization problems in complex industrial systems, such as adapting the product specification to the available resources, adapting planning of manufacturing processes to bottlenecks along the supply chain, optimizing the manufacturing processes for low batch sizes of custom products, or improving traceability and root cause analysis in incident management.

Real-life manufacturing systems and supply chains can be very complex, leading to difficult-to-solve optimization problems. Simulation models are often used to evaluate alternative decision-making policies for complex systems. After building the simulation models and the analysis techniques that explicitly link data (or the outcomes of the data-driven prediction models) to

decisions is required, TUE plans to use simulation-based optimization techniques, and deep-reinforcement learning in particular to solve the optimization problems for operational and tactical decision making problems encountered in the use cases.

One major reason for the semiconductor crisis recently ongoing is due to a heavy bullwhip effect along OEM supply chains. When connecting those supply chains more closely and enabling collaborative development and planning, we will reduce future bullwhips in demand significantly. Additionally, when applying graph-based data structures, environmental measures can be considered in planning, which will support to fulfil the supply chain due diligence.

Once this key enabler is elaborated, AI methods can be used to leverage the knowledge incorporated in the data. In this regard, we will explore two methods:

1. Using Deep Reinforcement Learning (DRL) for planning
2. Using GNNs for quality incident prediction

The cluster composed by NXP, Semaku and TUe will collaborate on above mentioned research directions in the field of Deep Learning.

All the reported activities will provide algorithmic approaches and tools that will be leveraged in WP6, UC19, *'Extension of Digital Reference Ontology for Smart Sustainable Business Planning and incident handling '*.

2.4.1 Using Deep Reinforcement Learning (DRL) for planning

Semiconductor industry has a complex supply chain to manufacture and deliver semiconductors to a wide range of customers. Generating robust supply chain plans is challenging due to the high supply chain complexity.

Currently, NXP uses a centralized/top-down LP model to generate supply chain plans based on capacity and demand. Due to the large scale of such models, run times are long and planners often need to make decisions without reconfirmation from the planning models. In order to overcome these challenges, two deep learning methods will be developed to provide both planning insights and decision support, combined with limited computational and maintenance costs compared to LP models.

1. Decentralized planning using a multi-agent system

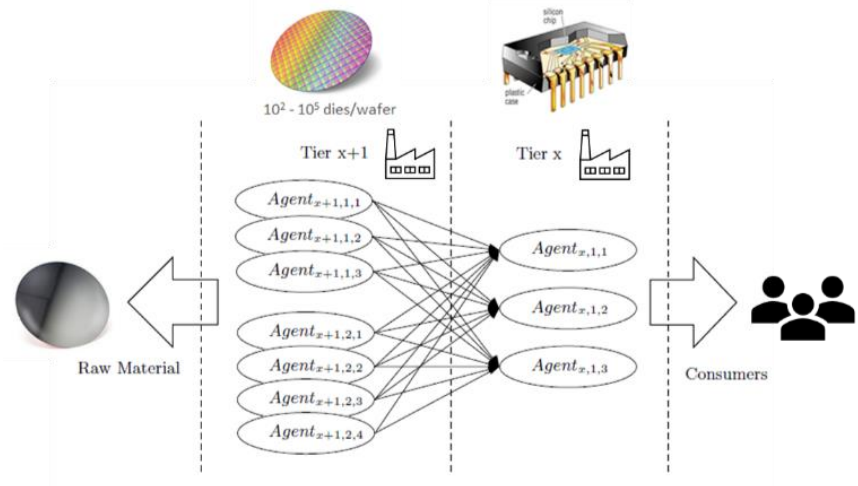

Figure 1: Decentralized planning

2. Planning using deep reinforcement learning

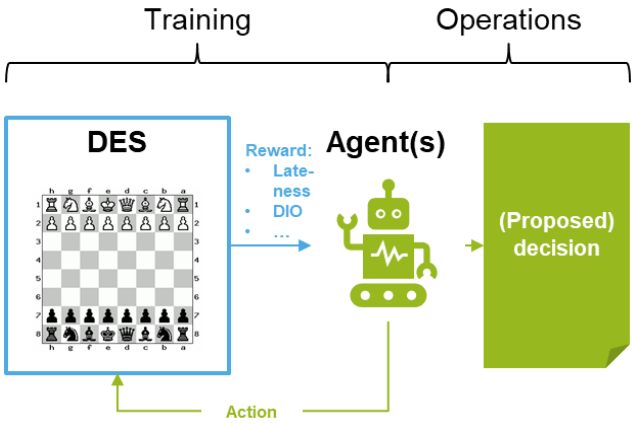

Figure 2: Planning using deep reinforcement learning

In decentralized planning, decision making is distributed over multiple intelligent agents, that can respond to their environment autonomously. The agents do so through learning capabilities. Each entity, or resource, within the system will have an agent to make responses to the environment. In planning using deep reinforcement learning, we will not assign a separate agent to each entity.
After training the agent(s), they can be used to take immediate decisions. This limits run times compared to LP models and because of the limited computational time, planners can be provided with fast reconfirmation from the planning model.

2.4.2 Using GNNs for quality incident prediction

As poor product quality is accompanied with unwanted consequences such as economic losses, wasted resources, and customer dissatisfaction it is extremely important for (semiconductor) manufacturing companies to focus on high product quality. Part of this is to learn from customer quality complaints. The Digital Reference (DR) model based on Semantic Web technologies should support case-based reasoning and risk assessments. Moreover, it should also support more efficient and effective quality incident traceability which is important to enable faster root-cause analysis. Root-cause learnings can then be used during manufacturing to proactively work on higher product quality by predicting quality incidents and preventing defect products. For this purpose, intelligent agents will be developed, for example using GNN. It should be noted that all these factors can be found in the concept of Zero-Defect Manufacturing as described by [Psarommatis 2022]. The following figure shows a generic ZDM (Zero Defect Manufacturing) framework which represents very well the goal of this use case from a high-level perspective.

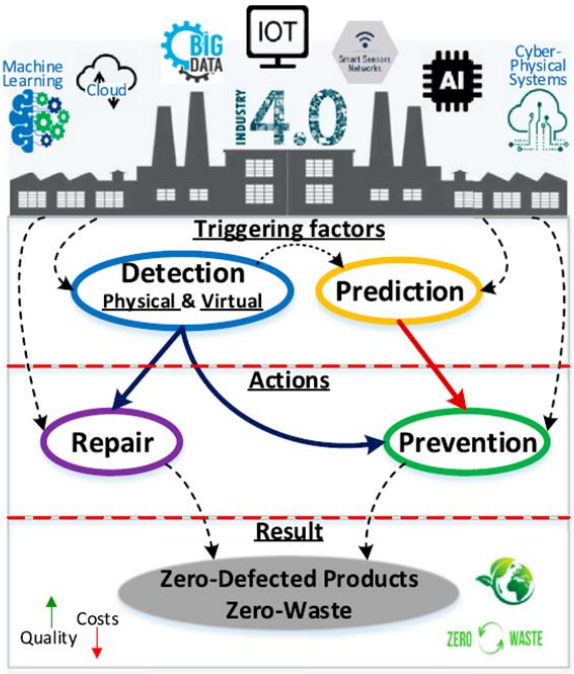

Figure 3: Zero Defect Manufacturing

Note. From "Zero-defect manufacturing the approach for higher manufacturing sustainability in the era of industry 4.0: a position paper," by F. Psarommatis, J. Sousa, J.P. Mendonca and D.Kiritsis, 2022, International Journal of Production Research, 60:1, 73-91, DOI: 10.1080/00207543.2021.1987551

2.5 KAI's and Infineon's planned contribution to AI tools, methods and algorithms for sustainable industrial processes

Within WP2, KAI will work together with Infineon Austria AG (IFAT) on data-centric as well as on model-centric AI approaches to develop single components of an ML pipeline for automated defect image classification in the semiconductor manufacturing process. The overall goal is to scale available knowledge, concepts, and processes to have a single model, comprising all defect types related to different technologies from all production layers, i.e., having a model for all defect types from every defect density inspection step along the whole semiconductor frontend production line. Next to training models, KAI will also work together with IFAT to develop a scalable ML pipeline for model deployment and model maintenance. To identify the point in time of a necessary model update due to e.g., equipment or process changes, a model monitoring concepts will be developed in WP2. Finally, explainable AI approaches and data visualization formats for the pipeline will be evaluated, to enhance the interpretability and credibility of the AI models.

All reported activities will generate algorithms and tools which are utilized in 'WP6, UC15, Productionized Machine Learning models at scale'.

2.5.1 Data-centric development approaches

Data is the language of machines, naturally implying that the better the data, the better the machine can learn. "Instead of focusing on the code, companies should focus on developing systematic engineering practices for improving data in ways that are reliable, efficient, and systematic. In other words, companies need to move from a model-centric approach to a data-centric approach." — Andrew Ng, CEO and Founder of LandingAI. KAI and Infineon will follow

this emerging and promising new approach in AI, by exploring the whole potential of data-centric approaches, focusing on the quality and quantity of data, in that case, SEM images from defects from all production layers of diverse technologies. The preparation of a high-quality data set for ML training means a lot of effort which has to be invested. KAI and IFAT will approach this task in a structured way, performing the following steps:

- Screening of available defect density data:
  - Extraction of available data
  - Assessment of class imbalances
- Data preparation:
  - Preparation of expert labelling
  - Augmentation of minority classes

Data preparation may be seen as a standard procedure, but it is not standardized, hence if it is performed in an unstructured and unguided way it may lead to non-reproducible data. In this WP, KAI and Infineon will define the appropriate workflows and tools needed to establish best working practices to ensure high quality data (in accordance with FAIR data principles) and scalable concepts.

2.5.2 Model-centric development approaches

For the model-centric approach, DL techniques like Convolutional Neural Networks (CNNs) are the first choice. If the envisaged model performance is not reached (e.g., for some defect classes), instead of “one model fits all”, single models for specific classes are generated, e.g. by using classical ML approach instead of DL. ML approaches also depend on high quality data and in addition, manual effort on feature extraction has to be invested. The concrete model-centric steps for KAI and Infineon are:

- “One model fits all”: Generation of a CNN model, comprising all defect classes among all defect layers
- If the performance is not good for some relevant defect types, the following remedies will be investigated:
  - Deeper hyperparameter tuning OR
  - Single extra models (DL or ML) for relevant defect types OR
  - Including additional data sources (e.g., images from other tools)

2.5.3 Scalable ML pipeline development

To generate business value, AI pipelines need to be deployed. For that reason, MLOps will be considered, which is a set of best practices that aims to deploy and maintain ML models in production reliably and efficiently. The implementation of the concept takes place in WP6, but it is of utmost importance that during model development the implementation setup is already taken into consideration. To support the sustainability of the whole model lifecycle management, KAI and IFAT will work on

- generalization, scalability, and transferability aspects of the developed pipeline for similar application scenarios.
- efficient model training to ensure a timely feasible re-training for model updates during productive usage.

Besides that, various tools will be utilized along the AI pipeline to support a sustainable and efficient model development, comprising e.g., HPC, data lakes, version control systems (for data, models, and code) and programming languages like Python.

2.5.4 **Model monitoring concepts**

Further, since the semiconductor industry is dynamic in the sense of continuous changes in equipment and environment, detection and early response to emerging deviations must be considered along the whole AI pipeline. Therefore, a focus will be on concepts for AI monitoring.

To achieve a robust concept, the following tasks need to be performed:

- Check and analyse the characteristics of input data distributions.
- Check the stability of the interfaces between the model and the production tools. Implement a system to trigger an alarm if the automatic classification is not working as required.
- Check the model stability after re-training with a pre-defined dataset.
- Monitor the prediction scores per class over time.
- Monitor the class distributions over time.
- Define model monitoring lots in production and analyse the model performance for these lots.
- Define a process to manually label the model monitoring lots including the responsibilities.
- Create tableau charts to compare the model predictions with the manual classifications.
- Develop and implement a strategy to trigger alarms in case of deviations between model predictions and manual classifications.

2.5.5 **Explainability**

In addition, different visualization steps will be implemented to e.g., assess results during development. Interactive Tableau charts will be created and provided to the stakeholders. These charts are handy to use and allow fast feedback to the developers.

2.6 BME contributions

As BME co-leads WP2 and leads T2.1 and Task 2.2, BME has some organizational role within WP2 besides the technical contributions. As a T2.1 leader, BME facilitates collaboration mainly with WP4, WP5, WP6 and WP7. Besides collecting requirements and facilitating matching activities withing WP2 and across WP5-WP6, BME contribute to co-design and integration of machine learning tools.

The conceptualization of the AI Toolbox is one of the main activities of BME. This includes creating the AI Toolbox repository and main pointers, and showing how the use-cases can utilize it. Besides, the connecting environment for the AI Toolbox is also important – an example for this is the AI Gym provided by WP3, hence BME contributes to these kinds of integrations as well. The initial plans for the AI toolbox can be found in Section 4.

BME plans to develop toolbox elements for Deep Learning-based modules, and tools using reinforcement learning, mostly targeting clustering, classification and prediction. BME will utilize the recent generative AI results, and facilitate the usage of copilot systems. Furthermore, the extension of the MLOps concept with the integration of the Engineering toolchain and the Digital Twin design/development/operation is also part of the BME contributions.

BME plans to create the Arrowhead copilot system, using generative AI tools. For the descriptions, see Section 5.

BME will also work on explainable AI mechanisms to improve acceptance at specific industrial domains where non-explainability would be considered a showstopper.

BME plans to contribute to the Eclipse Arrowhead AI tool integrations with AITIA – see the general plan descriptions in the AITIA contributions section.

BME is the leading responsible partner for Deliverables D2.1 and D2.2.

2.7 AITIA contributions

AITIA – as one of the main partners in developing Eclipse Arrowhead – provides comprehensive assistance to the partners in the implementation and management of AI-driven microservices. Enabling the integration of AI solutions within Eclipse Arrowhead can be accomplished through the following approaches:

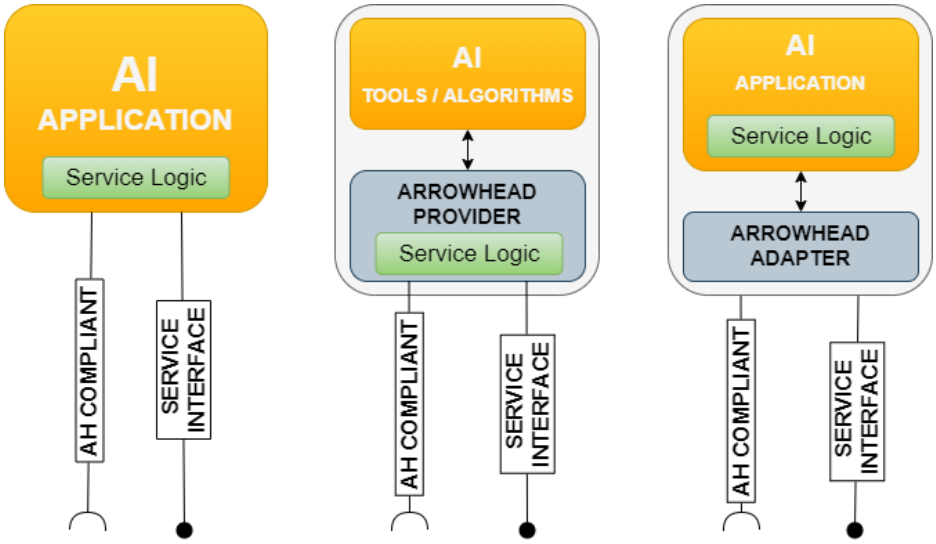

Figure 4: AI-Powered microservice integration scenarios

- Having a system with built in arrowhead capabilities and with integrated AI solutions targeting to perform a specific, well-defined service to support a complex use case.
- Having a system with built in arrowhead capabilities that performs a complex service using non-integrated AI tools and algorithms.
- Having a system with integrated AI solutions to perform a specific, well-defined service but without arrowhead capabilities. An additional adapter system is required that has arrowhead capabilities and able to expose the AI service access point within a Local Arrowhead Cloud.

To provide our partners with a broad perspective, AITIA is currently developing a proof-of-concept project aimed at demonstrating the practical application of AI-driven microservices. Within the demonstration, we monitor various areas equipped with sensors (measuring parameters like distance, temperature etc.) in addition to surveillance cameras. The camera feeds are subjected to image processing through a deep learning-based object recognition service. In instances where the image processing detects a person within a designated area, the information is collected and presented by a dashboard application alongside the sensor data. Furthermore, in a later phase, various light and sound signalling devices are planned to being activated as part of the response mechanism.

AITIA also contributes to the AI Toolbox creation, and the MLOps process definition for CPSoS in AIMS5.0.

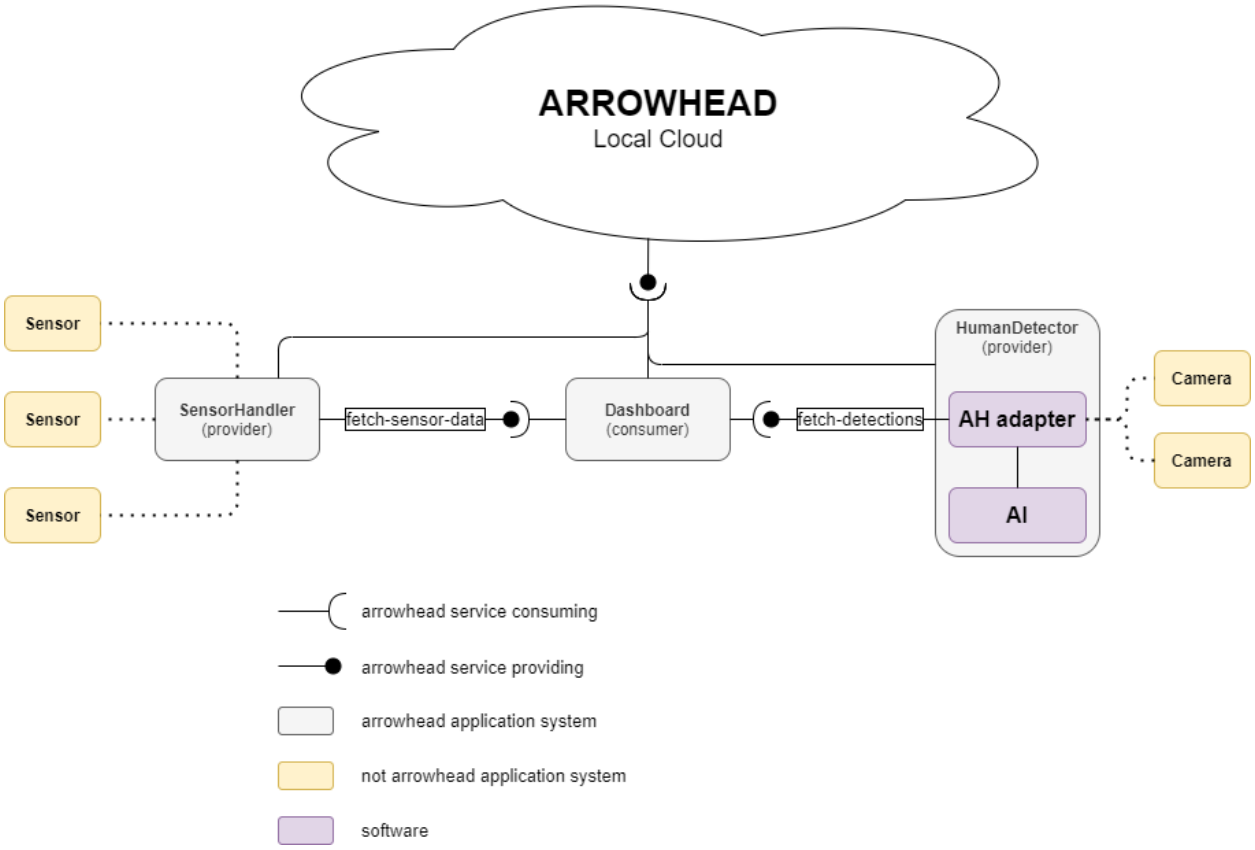

Figure 5: Demonstrator project overview

2.8 Gdansk University of Technology (GUT) and TTTech contribution

Gdansk University of Technology will collaborate with TTTech on researching and developing technology for reliable wireless communication aiming to increase the connectivity between the wireless devices within Indoor Food Production facility.

Wireless communication, in comparison to wired link, is much less robust to the influence of the environment in which it operates. Radio noise, physical obstacles or changes in the environment might decrease the quality of the communication. The wireless communication link is exposed to various types of external interference, whether unintentional, like interference from other communication signals within the same range, or intentional interference intended to purposely disrupt communication. Even though both forms of interference pose potential threats, intentional interference is more difficult to mitigate, especially when sophisticated techniques are used. To mitigate the risk, modifications to the physical layer of the communication can be implemented. An effective approach involves employing a reconfigurable antenna capable of adjusting its radiation pattern, which minimizes the impact of jamming attacks on the quality of received signals. The figure below illustrates the approach – for a traditional omnidirectional antenna, interference signal is added to the received signal in the receiver and, as a result, the received signal is noisy and unreadable. However, using a reconfigurable antenna with an optimally shaped radiation beam, the receiver

can minimize the influence of the jamming by reducing its power share in the receiver, enabling the successful reception of the transmitted message.

The utilization of switched-beam antenna with a reconfigurable radiation pattern enhances the signal-to-interference ratio in the receiver. Additionally, the implementation of a rapid control algorithm is essential to enable real-time operation of this solution.

The proposed solution introduces the practical use of a switched-beam antenna to counteract interference and jamming attacks on wireless communication. Implementing an Electronically Steerable Parasitic Array Radiator (ESPAR) antenna with switched impedances for this purpose, characterized by low power consumption and cost-effectiveness, gives the possibility to switch the radiation pattern between 4096 configurations. That way, the beam can be chosen that maximizes the signal reception and minimize the influence of interference at the same time. The main challenge is to choose the right antenna radiation pattern as fast as possible. Thus, instead of scanning one by one, sophisticated techniques need to be researched making the antenna able to maximize and maintain high signal-to-noise ratio in the situation when there is no information about the environment (ie. current positions of signal sources). Furthermore, in the absence of jamming attacks, this approach will allow the antenna's radiation pattern to adapt to changing environmental conditions, ultimately enhancing the overall quality of wireless communication.

For this purpose, ML optimization models will be researched and implemented. Binary classification with variable threshold will be the method of first choice and other models will be explored. Training data will be collected in various environments and setups. Possibilities for new implementation of the dedicated loss function will be explored, aiming strictly on signal quality indicators available in wireless communication modules.

Developed models will be integrated, implemented and tested on an industrial edge device delivered by TTTech, ensuring the necessary computational power for rapid processing and continuous reconfiguration to maintain high parameters of the communication. Solution will be validated in UC11.

2.9 RSA FG contribution

For our work in WP3 we will evaluate the usage of the PrototAIp tool and will contribute this as a cross-work package activity. PrototAIp is an open-source platform available on Github[1]. PrototAIp was developed by RSA FG in a previous research project. It is a platform for collaborative software engineering especially customized for the usage in data-science and AI projects. The platform is a Web application and, consequently, accessible via a browser. Registered users have their own workspace reserved on the platform. Common software libraries for data-science and AI experiments are already pre-installed on the platform. This relieves user from the burden of having to install the software themselves and keeping a virtual environment up to date.

As mentioned, the software required to host an on-premises installation is available on Github. It can conveniently be installed using docker and is ready in minutes.

[1] PrototAIp: https://github.com/Researchstudio-DSc/prototAIp, accessed Oct 12, 2023.

2.10 PCL Contribution

PCL is leading the AI Enhanced Value Chain project development activities. This project aims to effectively connect consumer, innovation and development groups and operations. In the this project, we put the user and manufacturing in the center, and we aim to design for them.

This target requires the following developments:

- Continuous consumer data analytics;
- Vision based product quality evaluation;
- Intelligent process control and generative process change control;
- Model Based Systems Engineering:
    - Up to date user models in the loop;
    - Up to date manufacturing data in the loop;
    - Track the design specifications to both ends;
- Generative production planning against market ask and logistics constraints.

Additionally, PCL aims to enhance the production quality and efficiency, improve the effectiveness of the design efforts, optimize the product platform creation against user variation points versus manufacturing capabilities, optimize the cost per manufacturing blue-print change request and, most importantly, enable the mass customization within PCL‘s mass production environment.

PCL contributes mainly to WP5 and WP2 with less focus on Eclipse-Arrowhead studies in WP3, to create a platform-free solution.

To achieve this ambitious goal, PCL has defined several tasks distributed among the partners. Further working with the partners is required to collaboratively prioritize the tasks and define the details per task. Short descriptions per task can be found below:

- **Model Based Systems Engineering:**

  This task aims to digitalize the development process and create active links from user data and manufacturing data towards the development activities.

    - *Design for User:*
      As part of this task, user models will be created and linked to the project models. User experience flows and user requirements will be created per user type, where the user type can represent a cluster of users or each individual user. Individually created user requirements will be tracked towards the part specifications. User data will be continuously fed through the user model(s).

    - *Design for Manufacturing:*
      Designing parts and creating product platforms for the manufacturing capabilities are already in use. However, this task aims to enhance the design for manufacturing capabilities even further. Actively tracing the production and assembly risks towards the parts and elements, modelling the process line and tracing the part CTQs to the process CTQs and support the root cause analysis at the production line. In addition, automating the CAPEX impact assessment at the product platform creation level and optimizing the decisions.

- **AI Powered Automated Supply Chain Planning:**

  Automatically translating the market sales planning into the production plans while considering the supply chain risk per material and per supplier, warehouse risks, process line quality risks and predictive maintenance at the production line along with the fluctuating sales planning and dynamic transportation conditions.

- **Vision Based Production Quality Control:**

  In the production line, automatically running the part CTQ check and tracing the part CTQ to the candidate process CTQ. The expected challenges will be the high speed of the production, capturing the part CTQs within the production line with the required sensitivity, running a quality check, and automated root cause analysis from part CTQ to process CTQ.

- **Data Based Process Risk Assessment:**

  This task aims to provide predictive process risk assessment based on the process data history. Data based risk assessment is required during the part development and part production throughout the product lifetime.

- **Customized Process Specifications within Production Line:**

  This task will be enabled with Task 3 and Task 4. It aims at automatically customizing the process parameters against production quality, used material parameters, and individual user CTQs. As an example, injection moulding process parameters can be optimized based on the used material and individual user CTQs.

- **Process Capability Prediction:**

  This task aims to automatically create a product platform using the manufacturing blueprints as input along with the user requests and build the most optimized set of platforms as output. To achieve this task, it is required to predict the capabilities of the new manufacturing blueprints and automate the platform decisions to optimize CAPEX versus Platform Capabilities versus User Value versus Maintenance Cost over time (both of R&D cost for diversity and manufacturing maintenance) versus Marketing value.

- **Predictive Maintenance:**

  This task aims to predict the required process line maintenance activities based on the data-based process quality assessment, operation costs and supply chain planning.

2.11 IDEKO, SORALUCE, GOIMEK and SAVVY contribution

UC5 cluster comprising the partners above will collaborate searching and developing Artificial Intelligence tools for machine tool domain, within the context of Industrial IoT and Manufacturing Intelligence. These developments will provide methods and algorithms that will boost various value propositions of all the partner companies involved, and that will be applied in UC5: *AI techniques at different layers in Machine Tool Domain,* focusing on specific areas of the manufacturing process and having identified at least one solvable problem:

Automatic tool breakage detection in milling operations is a problem of interest in the manufacturing industry, as a tool breakage can significantly affect product quality and increase

production costs. Milling processes can be long-term processes in which the machine works unattended, i.e. the work is carried out without the need for constant operator attention. After setting up the machine and programming the necessary instructions, the machine can run for hours without human intervention. However, there is the possibility of occurring errors during this milling process. These errors can lead to errors in the final part, leading to re-machining of the part or, in the worst cases, rejection of the machined part.

Therefore, to solve this problem, an Artificial Intelligence system capable of detecting early and automatic tool breakage would be of great benefit to reduce production costs, avoiding defective parts and wasted time. Artificial Intelligence algorithms could identify patterns in the data that indicate a fault or failure, or simply abnormal machine behaviour. These AI techniques could help improving the efficiency and reliability of milling processes.

To address this problem, different models will be trained by applying several AI techniques. In recent years, several analyses have been carried out in order to detect automatically tool breakages in milling processes.

2.11.1 Identification of necessary components and data

One of the most common techniques to address this problem is the analysis based on vibration and acceleration signals. Machine Learning and Deep Learning algorithms such as, neural networks, SVM and decision trees, can be applied to analyse the signals and detect tool breakage. In addition, pre-processing techniques like filtering and Fourier analysis have been used to improve the accuracy of the models.

Another technique that has gained popularity is the use of images of the workpiece and tool surface. The images are processed using computer vision techniques, such as segmentation and feature extraction, to identify the presence of breakages in the tool. Deep Learning techniques can be used, for example, convolutional neural networks, to improve the accuracy of the tool breakage detection. Solutions based on the fusion of multiple sensors, vibration, acceleration, temperature and power sensors, have also been analysed.

Regardless of the components on which the digitization is based, a common framework of requirements for collected data should be met (Time series):

- Univocality: all data points should be identifiable from one another. This is typically achieved by means of timestamps accompanying value points or tuples.
- High resolution: timestamps should be as granular as possible; a grain of 1ms (one millisecond) is usually considered high enough resolution.
- Continuity: there should be no gaps in a given series of data values between the starting and end points.
- Availability: data should be retrievable by an authorised actor in both „batch“ and „real-time“ fashion depending on the purpose (i.e., batches for a ML model during its training phase, real-time streams for online predictions)

2.11.2 Techniques for anomaly detection

Depending on the data collected or available to address this problem, different AI techniques can be applied. Two types of data can be gathered in the machines.

On the one hand, in high-frequency data, the data is collected at a rapid pace. During machine movements, high-frequency signals are captured until the entire process is completed and the aim of the analysis of these data is usually to detect anomalies or changes of scenario in the

behaviour of the machines. On the other hand, low-frequency real-time data is continuously generated by the machine during normal operation. The data is collected at a slower pace, but it provides a steady stream of information about the machine's performance.

Raw vibration signal and the FFT (Fast Fourier Transform) of these vibration signals can be very interesting data for tool breakage analysis. The peaks or amplitudes extracted from the FFT signals can also provide information about tool breakage. In addition, spectrograms, the result of calculating the spectrum of a signal by time windows, can be powerful in illustrating how the vibration spectrum changes. Thus, by treating the vibration signals or FFTs as images, different and useful input data can be obtained for Deep Learning algorithms.

Different Machine Learning and Deep Learning techniques can be applied for tool breakage detection. Some of these techniques are the following:

- Random Forest: a machine learning algorithm based on the combination of decision trees that is capable of handling large data sets with multiple features. It could be useful for tool breakage detection in milling, especially when there is a large amount of data.

- Support Vector Machines (SVM): commonly used for binary classification problems with the goal of constructing a hyperplane (or set of hyperplanes) in a high-dimensional space that separates the two different classes of data points. It can work well with labelled data and with sensor signals such as accelerometers and axis powers.

- Convolutional Neural Networks (CNN): neural network that is commonly used in image and time signal processing. It can be trained to recognize patterns in sensor signals and could therefore be useful for tool breakage detection, working with raw signals collected from accelerometers or time series.

- Gradient Boosting Machines (GBM): a machine learning algorithm based on a set of weak prediction models, which is useful for binary classification and regression. It could be used for tool breakage detection when having labelled data and signals from accelerometers and axis powers.

2.11.3 Provision of results and insights

Any insight, prediction or result that is be made by models developed for UC5 should be provisioned to the corresponding stakeholder by different means:

•Models that „close the loop", that is, they automatically apply the results they generate to the production process, given their habitual low-latency requirements, typically run on the Edge, as close as possible to the corresponding equipment interface – models that, due to their heavy computational requirements, are unable to run and execute their predictions in a timely fashion as a result of hardware limitations, will instead run on the Fog layers.

•Models that generate insights, either on demand or on a schedule, should have their own dedicated dashboards that authorized actors may access and consult at their convenience, at the Cloud level. Generation and mailing of reports, also on-demand and on a schedule is also factored in.

2.12 HUA Contribution

In terms of UC6, Harokopio University of Athens (HUA) will collaborate with EDI in the development of algorithms and tools for the control of a robotic arm. The overall objective is

to develop a modular and reconfigurable solution that automates the conveyor feeding processes for irregularly distributed bottles of different sizes, weights, and types. In terms of WP2 HUA will survey existing algorithms for robotic arm manipulation and computer vision tasks, which will be further analysed in order to develop an efficient solution. More specifically, HUA will develop machine learning based models for specific micro-tasks in the robotic action process, which can then be composed by the action planning module developed by EDI. For this purpose, HUA identifies several smaller tasks in the whole process (e.g., object detection, shape, size and orientation detection, arm motion planning, grip control, object picking) and is working on finding the best solution per case. The results of this collaborative work will be evaluated first in a simulated environment and consequently in a real robotic arm.

2.12.1 Code base and Experimental Environment

A very crucial element for implementing and validating the various approaches is the environment that will be used. It is common in robotic tasks to employ a simulation environment first in order to develop and test the algorithms before porting them to the real application and environment. This is very helpful when the task and its environment that has to be simulated is specific and has few degrees of freedom but can be cumbersome when the task and the environment are complicated. In the specific task of object localization and picking, the simulation environments are widely used especially in indoor environments (Fernandez-Chaves et al, 2022). Another reason is that in contrast to kinematics-based planning and other deterministic solutions, the approaches that rely on machine learning operate as decision-making black boxes and need a lot of trial-and-error approaches in order to learn. When dealing with expensive machinery like industrial robotic arms, especially where human presence is mandatory, it is essential to maintain everything under a controlled, safe and cost-free environment (Xiang et al, 2020). For this reason, we decide to use a simulation software and create a representation of the task in the simulation environment, using robotic arms, moving conveyor belts, and human interactions to emulate the real-world setting. All algorithms and approaches will be tested and evaluated based on this simulated environment.

Another important aspect, especially when porting the solution from the simulation to the real environment is to adhere to the common industry standards for application development. For this purpose, we plan to adopt widely accepted good practices such as "Test-Driven Development" (TDD), combined with principles of "Interface Segregation" and "Dependency Inversion". Also we will use CI/CD (Continuous Integration/Continuous Deployment) techniques, combined with version control. This strategy aims to develop a unified code base that will enhance cooperation between partners and result in a high-quality, expandable, and maintainable final product.

2.12.2 Composite task deaggregation and machine learning pipelines

For tasks like pick-and-place, multiple approaches can be employed. According to the literature, the principal strategies for complicated systems that perform these tasks, involving imaging sensors, other types of sensors, mechanical and electronic parts, and of course, constraints related to safety and movement, boil down to two approaches: the Modular Pipeline and the End-to-End Pipeline.

In the End-to-End Pipeline, a single model attempts to solve the problem in one go, with the entire system trained and evaluated as a unified entity (Zhou et al, 2022; Geng et al, 2023). This approach has several disadvantages, the most significant being the "black box" nature of decision-making, especially in a multi-component scenario. Other issues include limitations in generalization and expandability since the model is tailored for a specific setup and current

environmental conditions. Any new addition or modification necessitates retraining of the system to accommodate the new setup.

The Modular Pipeline approach seems more suitable for complex tasks in industrial applications. In this strategy, every part of the system is developed, trained, and evaluated independently, and then the subparts are combined to create the solution. Each module is reusable and can integrate with other systems (Chen et al, 2021). For example, we can have a module that, based on camera feedback, detects the trajectories of moving objects on the assembly line. The output of this module can be utilized as input to other modules that perform subsequent or parallel tasks, and any upgrade in the camera system solely requires the retraining of the model used by that specific module.

Given these considerations, we assert that our approach should be based on the Modular Pipeline, particularly for the task of object grasping. An initial pipeline could be broken down to specific tasks that use the respective module as follows:

Tasks	**Technologies**
Object Detection	Supervised Learning using Transfer Learning
Pose Estimation	Supervised Learning using Transfer Learning
Robotic Arm Movement	Deep Reinforcement Learning and Behaviour Cloning
Conveyor Placement	Deep Reinforcement Learning or Supervised Learning for trajectory and speed estimation

A pipeline like the one mentioned above will afford us the capability to fine-tune each module independently until we achieve the desired KPIs (Key Performance Indicators). Moreover, the modularity of this strategy will provide us with the flexibility to quickly and efficiently adapt the system to new environments or different tasks, thereby enhancing overall productivity and efficacy. It will also allow us to break a hard composite task into smaller tasks, adapt and orchestrate the robotic arm behaviour even using natural language (Zhou et al, 2023)

2.12.3 The use of Deep Reinforcement Learning for task automation

Based on literature, we have a plethora of publications related to the use of DRL (Deep Reinforcement Learning) in controlling robotic arms (Zhang et al, 2015; Amarjyoti, 2017; Saeed et al, 2021; Singh et al 2022). So far, the results in simulated environments are impressive, but there is very limited literature on the actual usage of these types of models under real conditions. Through our work, we will try to explore the performance of these models in real environments. A basic step before someone starts to train a DRL model is to define the problem in RL (Reinforcement Learning) terms, which means to define the three major parts of the task: the observation space, the action space, and the reward function (Zhao et al, 2020; Joshi et al, 2020).

- The observation space is what the model takes as input to train or make a prediction. This space could comprise a single frame or multiple frames from a video feed, data collected from sensing devices, or a combination of all these elements. For applications like those targeted by this project, the combination of both sensor data and visual inputs can potentially yield the best results. This is because most modern production lines are already equipped with multiple sensors and utilizing them in tandem with images can provide the most comprehensive information about the environment.
- The action space is associated with the available actions that the robot can perform. For instance, a robotic arm with multiple joints has a vast number of potential movements. These actions can be discrete, such as up/down movements, or continuous,

as in a joint rotating by 25 degrees. The actions, which are the outputs of the model, vary based on the application and the type of robot.

- Finally, the reward function is the most crucial part of any RL project. The model uses the reward function to self-evaluate every action based on a state of the observation space, and this function encapsulates the knowledge of what actions are permissible for the robot and also the quality of those actions. For a task like grabbing a bottle, a reward is given when the robotic arm reaches or grabs the bottle. This incentivizes specific outcomes and behaviours, guiding the system toward achieving its programmed goals efficiently and effectively.

2.12.4 Behaviour Cloning Learning and Imitation Learning

The use of DRL (Deep Reinforcement Learning) can sometimes be highly unpredictable and time-consuming in terms of training. In many cases, even simple tasks (from a human perspective), such as grabbing a bottle, can take days before a DRL model achieves a level of success. To enhance the model's performance and provide in-field programmers with better control and a more efficient method of teaching new tasks, we will combine Behavioural Cloning (Choi et al, 2020) and Imitation Learning (Fang et al, 2019) with DRL. This approach will allow a human expert to teach the system how to perform certain tasks, at least to a human-acceptable level. After this initial learning phase, the unique ability of DRL to generalize, as well as to explore and exploit its policy, is expected to lead to more efficient and faster ways to accomplish the same tasks.

By integrating the direct learning from human experts (through Behavioural Cloning, Imitation Learning or similar techniques) with the adaptive, trial-and-error learning approach of DRL, we aim to create a more robust system. This hybrid strategy not only expedites the initial learning phase, but also leverages DRL's strengths in adapting to various scenarios and optimizing performance beyond human capabilities.

2.13 Infineon Technologies Dresden (IFD) contribution in cooperation with IFAG, IFAT, FuH, Systema, HSZG, WUW, HoST, TUD, FMX, UNPD, PFV, FhG IISB, ViF, UzL

Infineon Dresden Technologies is working with its partner on 8 main use cases within WP6 using Factory Automation, IoT, Machine Learning and AI to enhance manufacturing performance, resilience, agility and sustainability in the specific environment of the semiconductor production. The Figure below gives the overview about the eight main use cases, UC09, UC10, UC12, UC13, UC14, UC15, UC17 and UC20.



Figure 6: The eight main use cases with IFX contributions within AIMS5.0 WP6 (IFX contributions in green)

Within the Deliverable D6.1 the main requirements for the eight use cases are documented in terms of sustainability, productivity, resilience and the influence on human centered work places. For the influence on the new work places the three pillars like explainable and trustable AI as well as ethic aspects are considered in the first early step of the project and summarized for all use cases with IFD contributions.

IFD is deeply involved within the task T6.3.4. for smart sensor platforms. The sensor team is working in the project on the installation of many different sensor systems to improve sustainability, quality and productivity using new methods out of the field of machine learning and AI. This work is in strong cooperation with WP1 for the sensor hardware, WP2 to implement the AI tools and establishing the specific architectures within the use cases. The Figure below gives an overview which of the tools out of the IoT will be applied within the front-end factories at Infineon at the two sites at Dresden & Villach and the backend site at Warstein.



Figure 7: Methods and Tools which will be used to establish the use cases within WP6

Below is a short description about the use of specific tools within the eight use cases with Infineon contributions.

2.13.1 UC09 – Energy-aware Scheduling of Batch Processing Machines in Wafer Fabs (FuH, IFD, IFAT, Systema, ViF)

Many different smart sensors are installed and will be also installed during the project time to get out more data about energy flows in the frontend manufacturing fabs for 200 and 300 mm plants. The goal is to define new dispatching and scheduling algorithms which could help to balance energy flows in a highly complex manufacturing environment. IFAT and ViF have started the work on digital twins for high energy consuming equipment with the goal to reduce electrical power.

FuH is working on diffusion furnaces which are among the machines with the largest energy consumption in wafer fabs [National Environment Agency Singapore 2023] we will focus on this machine group. Diffusion furnaces are batch processing machines (BPMs). This means that they can process batches, i.e. groups of jobs, at the same time. This setting is called p-batching [Fowler & Mönch 2022]. In the current stage of this research, we consider n jobs that have to be processed on m identical BPMs. Extensions to flexible flow shops are planned. Each job j has a due date d_j, a ready time $r_j \geq 0$ and a weight w_j that indicated the relative importance of the job. The jobs belong to different incompatible job families which are labeled by $f = 1,\ldots,F$. Jobs of different families cannot be processed together in a batch. The n_f jobs that belong to family f have the same processing time p_f. The family of job j is $f(j)$. The processing time of a batch is determined by the processing time of the family of the jobs that are in this batch. The maximum batch size B is measured in number of jobs. After a batch is started, it cannot be interrupted. A finite scheduling horizon consisting of $t = 1,\ldots,T$ time slots of equal size is assumed. The TEC is modeled as a piecewise constant function over the time slots of the horizon, i.e., we have $TEC = \sum_{k=1}^{m} \sum_{t=1}^{T} e(t)\, z_{kt}$ where $e(t)$ is the EC in time slot t and z_{kt} is an indicator variable which takes the value of 1 if a batch is processed in time slot t on machine k and 0 otherwise. The TWT measure is given by $TWT = \sum_{j=1}^{n} w_j \max(C_j - d_j, 0)$ where C_j is the completion time of job j. In the

three-field notation from deterministic machine scheduling theory [Graham et al. 1979] the described problem can be represented by $P\left|p-batch,incompatible,r_j\right|\lambda TWT+(1-\lambda)\,TEC$, where P denotes parallel identical machines and $p-batch,incompatible$ refers to parallel batch processing with incompatible job families. The parameter $\lambda\in[0,1]$ is used to blend the two objectives.

The scheduling problem is not regular since including idle in the schedule might be beneficial in respect to the TEC value. The two measures are therefore clearly in conflict because in order to minimize the TWT value each job should be processed as soon as possible. Since the $1||TWT$ single machine-problem, a special case of the $1|p-batch,incompatible|TWT$ single-BPM scheduling problem for $B=1$, is already NP-hard [Lawler 1977] the problem at hand is also NP-hard [T'kindt & Billaut 2002]. Hence, our goal is to design efficient heuristics to solve this problem based on list scheduling and genetic programming.

In a first step, the partners IFAT and ViF will develop a digital twin of the refrigeration network for the Villach site. This will be done in such a way that the models and methods developed can also be adapted to the site in Dresden with little effort. ViF will start to elicit the requirements by conducting interviews with relevant stakeholders. Based on that findings ViF will identify the data and information sources to be used for further analysis. This will lead to the definition of the preliminary digital twin requirements with a high-level description of the features in order to choose the right mathematical methods.

The following step is the data collection and algorithm application which is strongly related to WP6. Based on the resulting requirements ViF will specify what kind of data is needed for the application of the methods and algorithm for promising data analysis. ViF will organize several workshops with the partners in Villach and Dresden to agree on the appropriate data and data format and gather the relevant domain knowledge of the Infineon experts. This will lead to an improved understanding of the entire system as well as an increased knowledge about the examined data, in particular how was the data generated, what possible error sources exist, where to find the relevant information in the high amount of provided data. After an Exploratory Data Analysis (EDA), ViF will develop several question and hypothesis, which could be confirmed by applying the algorithms on the data.

In the related WP7 ViF will conduct qualitative interviews with the IFAT FM staff to better understand their requirements and to infer how to facilitate trust and acceptance for the new solution. This information will be fed back into the algorithm development and validation strategy to improve the use case.

Initial results

In several online meetings, the problems were described, and possible solution scenarios discussed. Subsequently, the plant was inspected in a successful on-site workshop in Villach and both partners worked out a joint approach. An inventory of the current state was made, and relevant questions were worked out. Subsequently, an approach with increasing complexity and benefit was developed.

The refrigeration network at the Villach site consists of chillers -> chilled water pumps -> cooling towers -> consumers such as make-up air units; process cooling water and all cooling systems such as fan coils, recirculating air coolers, etc. Special attention was paid to the cooling network in the new GP300 clean room area, since there are significantly more data points available here than in the legacy area and the results can be more easily applied to existing locations, but also to new fabs (e.g., Dresden or Kulim). From the sequence, the cold generation is analysed first and then the cold distribution and the cold consumption.

In order to achieve a significant optimization of refrigeration for IFAT Villach, a list of questions was developed including the following:

- How efficiently are the refrigeration machines currently working?
- Which cooling quantities are required and with which supply and return temperatures can be operated optimally?
- Does the process cooling water also run during shutdown or when there is no demand for process cooling water?
- What is the influence of changes in the current settings?
- Is the current generation efficient or too high or too low?
- Could there be an oversizing of the consumers due to too high manufacturer specifications?
- How does the demand change with additional process equipment?

A procedure in 4 steps was defined with increasing complexity and increasing demands on the mathematical procedures.

1. Performance monitoring: continuous monitoring of key performance parameters, such as cooling capacity, energy consumption, temperatures and flow rates for real-time overview of the chiller's condition to detect anomalies or performance degradation at an early stage.
2. Energy optimization: simulate and analyse scenarios to optimize energy efficiency. Test parameters such as setting the temperature difference, adjusting the refrigerant flow or using energy recovery systems to reduce energy consumption.
3. Fault diagnosis and prediction: predict malfunctions or potential failures of the chiller by analysing data from sensors and other sources. This would allow proactive maintenance actions to be taken to minimize unplanned downtime and improve system reliability.
4. Lifecycle management: track condition and performance over time, schedule maintenance intervals, predict component life, and efficiently schedule upgrades or spare parts when needed. Are the machines really running consistently and well?

2.13.2 UC10 – MES Airbus (Systema, IFD, IFAT)

This use case is one of the main bases for all of the other tasks and use cases within the world of semiconductors. The MES is the core system for factory automation, digitalization and in the future AI methods. All semiconductor partner from IFX will work together with Systema to establish a harmonized system for a factory cluster containing all components to implement digitalization, machine learning and AI methods within the MES. IFX will work in parallel on a secure IoT platform with internal cloud to connect the specific devices within the factory.

2.13.3 UC12 Human-Centered AI for the Optimization of Robust and Competitive Semiconductor Manufacturing Networks and Increased Sustainability AI (WUW, HSZG, Systema, IFD, IFAT, IFAG)

Within this use case the partner will work on different IoT applications and AI to improve the productivity and resilience within the frontend semiconductor supply chain. Methods and tools out of the IoT e.g., Arima, CNN (computational neuronal network) or CNN combined with LSTM (long short-term memory), XGBOST are already tested at IFD for one selected workcenter of furnace tools in the process flow of a power product. The first tests showed that upcoming bottlenecks could be very fast detected. The work will be spent to visualize the material flows within the complex production flows and create measures for fast reduction of bull whip effects and the possibility for line balancing. Furthermore, ML methods in combination with explainable AI methods such as SHAP will be used to identify and interpret non-linear

relationships between factors causing process variations which would then be resolved using the domain knowledge of the process experts.

2.13.4 UC13 Intelligent Sensors improving Robustness of Automated Wafer Transportation and Storage Systems (HoST, IFD, IFAG, IFAT, FMX, TUD)

Sensors are used for monitoring and controlling of the automated material handlings systems (AMHS) in the semiconductor factories at IFD and IFAT mainly on the existing systems from the company Muratec and at IFAG Regensburg on a conveyor-based material system. Within the first months of the project, sound sensors based on IFX sensors had been installed at IFD and IFAT which are collecting sound files. Machine learning applications and edge computing are started to establish predictive maintenance solutions. The main goals are fast detection of material problems on high volume AMHS components e.g., OHT shuttles and conveyor systems and to reduce down times and defect shuttles in the high-volume production. Nevertheless, the partner will work on new storage concepts for wafers in the factory which have to be stored due to market fluctuation.

2.13.5 UC14 Speeding-up the Benefit of Automation by Virtual Commissioning through Digital Twins (FMX, IFD, IFAT, IFAG, TUD, HoST)

In UC14, the project partners are working on approaches and algorithms for automating manual procedures in the commissioning process by relying on virtual twins and therefore virtual commissioning. One of the main fields tackled is the improvement of previously manually carried out “teaching procedures” of robotic arms for handling carriers. Here, a novel approach is developed to automatically identify the exact location of carriers at a variety of possible carrier ports: intelligent algorithms, e. g. from the field of “Computer Vision”, are applied for mobile, as well as for stationary handling systems. In addition to the identification of the exact location, the determination of detailed robot arm trajectories is a time-consuming task in the commissioning process, especially in storage systems with dozens of possible locations. The partners defined a novel approach to speed up this task by relying on virtual models of the robots and therefore virtual commissioning. Specifically, virtual twins of the relevant material handling systems are planned to be connected to the robotics suite using customized software, which will allow to virtually define and evaluate robot trajectories.

2.13.6 UC15 Productionized Machine Learning models at scale (IFD, IFAT, IFAG, UNPD, STATWOLF, LFO, KAI

The project partners are working on different machine learning (ML) models which will make the production more efficient and sustainable. One direction is dealing with heterogeneous data sources. Trace data e.g., time-series and quality scalar data for both the Virtual and the OEE prediction will be used to optimize OEE data. Another work is spent on enhancing scalability and productization of the ML solutions, esp. on bringing image classification to a broader application. In the first months, the team worked on considering software integration and architectures for data handling and interactions with production data sources, and for MLOps monitoring of ML solutions in production. Tools which will be applied are the development of eXplainable Artificial Intelligence (XAI) approaches for deposition processes. Furthermore, comparison with several Deep Learning architectures (GRU, CNN-GRU, TCN, CEEMDAN GRU) will be compared. The Italian cluster will discuss and share the results with the cluster in Germany, Austria and France.

2.13.7 UC16 AI enhanced production and performance of RF transceivers for civilian SATCOM and radar applications (THALES, FORTH)

Within this use case THALES and FORTH are aiming to properly apply AI techniques in a way that will offer an overall improvement in the manufacturing throughput and sustainability of RF components and system utilized in avionic radars and related high-power communication systems. This includes AI supported methodologies and fabrication protocols that will enable optimized use for the resources and the personnel effort. Among the various AI approaches the main effort will focus on supervised learning techniques as those techniques and under proper training may offer useful solutions and predictions (e.g., early warnings during fabrication) in optimizing or solving multiparametric and complicated problems such as those related to the design and fabrication of state-of-the-art RF components like the RF MEMS. Towards this end the use case partners are aiming to utilize data that will be generated by the design tasks, by the ISO monitoring protocol applied during the fabrication processes as well as from the various characterization approaches (e.g., microwave measurements). These data will serve as training validation and test datasets. Regarding the machine learning algorithms, the main aim of this use case is not to develop new algorithms but rather to introduce and utilize proper training procedures in collaboration to commonly adopted approaches and algorithms that support diverse supervised learning options such as linear regression, decisions trees and artificial neural networks.

2.13.8 UC17: Intelligent Contamination Management & Prescriptive Control Plans and Process Flows (FhG IISB, PFV, IFD, IFAT, IFAG), Systema)

The goal of this use case is to create methods based on gas analysis to prevent the semiconductor factories e.g., IFD & IFAT and the semiconductor factory for SiC at FhG IISB from any contamination within the clean room. The company Pfeiffer Vacuum is working on new chemical analytic systems to offer solutions for an intelligent contamination management and is working on new recipe management to get out more and easier data of the equipment software. IFD is working on multiple sensor systems which will be installed within the cleanroom and should be able to send multiple data to the server which is than the base for fast reactions in case of contamination. The hardware will be designed by Pfeiffer Vacuum and the company will improve the tool software to get out more relevant data for the customers. IFD is working on concepts for automated FOUP cleaning and will work on automated monitoring concepts of the contaminations level before and after FOUP cleaning.

2.13.9 UC20: Harmonization of Automated Digitized Equipment Control (IFAT, IFD, TUD)

Both front end manufacturing sites, IFAT and IFD will work on automation and digitization of the test wafers which are used to assure highest quality of the production at hundreds of tools and thousands of different process steps. Identification of many different software features is needed to harmonize the different test wafer concepts. The teams are working on many different user interfaces and especially within the MES system which is the link to UC10. A setup for the first harmonization concept is worked out until M6 with the biggest data bases e.g., MES, test wafer database, base data for process data at IFD (FTGB), and data bases for visualization e.g., an internal IFX application called DaVinici.

Most of the use cases where Infineon is involved and using IoT applications are driving hundreds of new sensors and use cases where advanced sensor applications are needed to achieve process control far beyond the existing methods. Scalability is one of the most important key success factors within those projects. There is a strong need to get faster in installing new sensors.

The huge sensor networks need to be operated 24/7 with a maximum on availability. Fixing problems as fast as possible at any time increases the productive time of the production equipment which mustn't operated with reduced process control. One part of this aspect is fast and easy onboarding and maintaining the sensor systems. But as industry doesn't start at the greed field in most cases it's getting very complex. Compared to the home automation, where discovery and onboarding of sensor and actor systems is meanwhile state of the art, industrial products seem far away from that kind of easy system handling. Lot of standardization have been done within the last couple years which arrive now the industrial manufacturing systems.

There are to mention the IO-Link and OPC-UA standard for example. But while in home automaton newer technologies like HTTP / REST communication is meanwhile state of the art, industrial products lack often on such flexible communication standards. And if they use for example OPC-UA or HTTP/REST there is no semantic harmonization between different supplier except the IO-Link standard. The IO-Link protocol addresses only the easy onboarding of sensors to the gateway but not the onboarding and discovery of the IO-Link sensor gateways. For example, if devices have an HTTP/HTTPS Interface there is no standard for the semantic of the REST services. e. g. one device requires "GET device-id" and another "GET device ID".

Another challenge is the onboarding of new sensor gateways in a complex fab network. There are about hundreds of different sub nets separated by a firewall. Not able to use low level network communication between those sub nets. It is unfortunately seldom that a device uses "Option 12 Hostname" of the DHCP Service. Without this option it is impossible to discover the provided IP from the DNS server. Due to the hundreds of different sub nets and missing broad cast functionality it is impossible to onboard the device without the network engineer with access to the switch which is logging the connecting activities using the mac address.

Another point is that lot of devices still are delivered with fix IP addresses like 192.168.x.x. So set the address to the needed address range requires the knowledge of an expert, reading the Supplier manual for knowing the default address and how to be setup a one-to-one connection to a PC. Finally, each supplier provides different software to change the IP address range.

Infineon is looking within the AIMS5.0 project for answers and methods for making device onboarding and maintaining as easy as in-home automation systems and as safe and reliable as industry requires. One additional upcoming aspect for maintaining sensor gateways with web servers (especially PLC's) is the certificate management. Modern PLC's provide HTML5 based web visualization which is also used at Infineon. But more and more https is forced which requires the installation of a certificate. The issue is certificates have a limited lifetime and need to be renewed on a regular base. There are several certificate management systems on the market, but they need to be installed manually on PLC's due to the installation procedure is not harmonized. There is a strong need for automation also in this sector of device operation.

Within the next months of the project Infineon will work together with its partner on the eight use cases and especially within the task T6.3.4. on smart sensor integration to benefit from the digitalization and AI improving productivity, create resilient, agile supply chains and make production more sustainable. Furthermore, Infineon will establish better working conditions and with cooperation with WP7 the teams will work on the three pillars and had already analysed the main contributions according the three pillars, see Figure 8 below.

	Trustable AI	Explainable AI	Ethics in AI
UC09	Scheduling algorithms break factory	Energy simulation not comprehensible	Employee surveillance based on performance
UC10	MES results break production	MES algorithms too complex to comprehend	Productivity improvements results in stress Employee surveillance based on performance
UC11	Algorithms prioritize efficiency over health or quality	Greenhouse control too complex	Algorithms prioritize efficiency over health or quality
UC12	Results are distorted due to biased data	Manufacturing decisions provided without explanations	Unclear accountability for decisions Employee surveillance based on performance
UC13	AI enabled material handling is inefficient	Material flow decisions provided without explanations	AI integration results in too stressful work tasks Employee surveillance based on performance
UC14	Digital twins doesn't reflect real-life counterpart	Digital twins are modeled as black boxes	Automation results in too mundane work tasks
UC15	Engineers don't have trust in ML systems	ML model decisions provided without explanations	AI integration results in too stressful work tasks
UC16	Wrong training data results in unusable results Engineers don't have trust in AI design ideas	AI design ideas provided without explanations	-
UC17	Engineers don't trust alerts. Alerts are not recognized	Contamination alters not traceable	AI surveys more than production line
UC18	Training data is incomplete	Source of knowledge is unknown	Sensitive data is shared between partners
UC19	Acceptance of AI solution within company not high	Planning decisions not explained/comprehensible	Sensitive data is shared between partners
UC20	Data analysis produces biased results Engineers don't have trust in AI systems	Monitoring alerts are not explained/comprehensible	Realtime monitoring of employees

Figure 8: Overview about the three pillars trustable, explainable AI and ethics for the use cases within WP6

3. Preliminary descriptions of the AI tools methods and algos for WP2

3.1 Tool for Predictive Maintenance

Predictive Maintenance (PdM) is one of the main AI-based technologies to enhance sustainability and more effective operations in manufacturing. Effective maintenance management helps reduce costs related to defective products and equipment downtime. A well-planned maintenance strategy improves reliability, prevents unexpected outages, and lowers operating costs. For such reasons, PdM is one of the research direction of major interest in AIMS5.0 and a dedicated library in the AIMS5.0 AI Toolbox has been developed.

CeRULEo, which stands for 'Comprehensive utilitiEs for Remaining Useful Life Estimation methOds', is a Python package designed to train and evaluate regression models for predicting remaining useful life (RUL) of equipment. RUL estimation is a process that uses prediction methods to forecast the future performance of machinery and obtain the time left before machinery loses its operation ability. The RUL estimation has been considered as a central technology of PdM [Christ2018]. PdM techniques can statistically evaluate a piece of equipment's health status, enabling early identification of impending failures and prompt pre-failure interventions, thanks to prediction tools based on historical data [Susto2014]. CeRULEo offers a comprehensive suite of tools to help with the analysis and pre-processing of preventive maintenance data. These tools also enable the training and evaluation of RUL models that are tailored to the specific needs of the problem at hand.

While PdM models can be used in a variety of contexts with different data sources and errors, there are some general techniques that can be applied, such as time-series validation, imputing missing values, handling homogeneous or non-homogeneous sampling rates, addressing values, range and behaviour differences across different machines and the creation of run-to-failure-cycle-based data. CeRULEo addresses these issues by providing a comprehensive toolkit for preprocessing time series data for use in PdM models, with a focus on run-to-failure cycles. The preprocessing includes sensor data validation methods, for studying not only missing and corrupted values but also distribution drift among different pieces of equipment. In addition to preprocessing, it enables the iteration of machine data for use in both mini-batch and full-batch regression models and is compatible with popular machine learning frameworks such as scikit-learn and tensorflow.

The library also includes a catalogue of successful deep learning models from the literature and a collection of commonly used RUL datasets for quick model evaluation. The acceptance of PdM technologies is pivotal in Industry 5.0 for successful implementation, but hesitations or reluctance by decision-makers can still pose significant barriers. One effective approach to foster acceptance and understanding is through explainability, which plays a crucial role in PdM. As such, CeRULEo incorporates explainable models capable of providing additional information about the predictions, enhancing comprehension: one that can select the most relevant features for the model, and a convolutional model that provides post-hoc explanations of the predictions to understand the reasoning behind the predicted RUL.

Moreover, CeRULEo provides tools for evaluating and comparing PdM models based on not only traditional regression metrics, but also on their ability to prevent errors and reduce costs. In many cases, the costs of not accurately detecting or anticipating faults can be much higher than the cost of inspections or maintenance due to reduced efficiency, unplanned downtime, and corrective maintenance expenses. In PdM, it is particularly important to be accurate about the RUL of equipment near the end of its lifespan, as an overestimation of RUL can have serious

consequences when immediate action is required. CeRULEo addresses this issue by providing mechanisms for weighting samples according to their importance and asymmetric losses for training models, as well as visualization tools for understanding model performance in relation to true RUL.

Additional details can be found at in the original paper [Lorenti2023], while the code for the package could be retrieved at the following link: https://github.com/lucianolorenti/ceruleo

3.2 Unsupervised Anomaly Detection

Anomaly/Outlier Detection (AD) is one of the most important tasks in unsupervised learning as it aims at detecting anomalous behaviors with respect to historical data; in particular, multivariate Anomaly Detection has an important role in many manufacturing applications thanks to the capability of summarizing the status of a complex system or observed phenomenon with a single indicator (typically called 'Anomaly Score (AS)') and thanks to the unsupervised nature of the task that does not require human tagging. In particular, black-box approaches, in contrast with physics inspired model-based ones, are receiving increasing attention in many applications where the derivation of a reliable model of the underlying phenomenon is impossible or non-trivial.

While black-box AD approaches have proven to be effective in many applications, several factors still limit the applicability and efficiency of such approaches in many scenarios. In this project we aim at advancing black-box AD methodologies by tackling 2 important aspects, namely: i. interpretability, and iii. the weakly-supervised (WS) scenario. Particular attention will be provided to isolation-based approaches due to the proven effectiveness of the Isolation Forest algorithm and its variants.

i. Interpretability - One aspect, that may be considered as a general problem affecting any technology based on Machine Learning (ML) and it is mainly due to two issues: (i) lack of confidence/trust from the users in AD algorithm outcomes and (ii) not immediate association between AD algorithm outcomes and root causes. The first issue arises from the lack of labelled data points (that, on the other hand, is one of the main reasons why AD algorithms are appealing in the first place), which makes it impossible to set up adequate testing procedures. This leads either to blindly trust the algorithm or not to use it at all, both the cases being undesirable. The second question, instead, investigates the possibility to gain additional knowledge about the task at hand, which may translate into actionable insights for troubleshooting or root cause analysis. The aforementioned issues can be addressed following the principles of eXplainable Artificial Intelligence (XAI), whose objective is to make so-called black-box ML models.

ii. Weakly supervised AD - As previously stated AD techniques belong to the class of unsupervised algorithms since they do not rely on labels provided by domain experts: as argued above, this is one of the major strengths and reasons for the popularity of such approaches. However, thanks to the diffusion of decision support systems with AD capabilities, the scenario where a small portion of labeling is available is becoming more frequent than in the past. The so-called, weakly supervised (WS) scenario is however ratherly addressed in the literature, with few works that have tackled the issue [Pang2023].

In this project is to develop algorithms able to take advantage of the WS scenarios, following two directions for investigation:

a. Exploiting the possibility of having interactive labeling of data, variants of the unsupervised AD methods could be devised with the aim of achieving higher accuracy in terms of anomalies detectability and/or simpler structures;
b. A second research direction would be to develop a self-supervised tool, employing an unsupervised performance metric developed for this task. The algorithm might give more importance to models that maximize this metric, instead of blindly guessing the abnormality level of an instance. This might be very useful for ensemble models where a collection of models is trained to enhance the detection accuracy of weak learners. Using the proposed approach, the ensemble might contain fewer or stronger models, leading to stronger and computationally more efficient AD.

Both research directions, if successfully tackled, will lead to the development of novel AD algorithms: we will develop open-source Python libraries with the developed approaches, that will be integrated in the AIMS5.0 toolbox.

3.3 AI tools methods and algorithms as used in Manufacturing for Signify use case

Signify is using for now mainly available data capturing methods and A.I. algorithms. It is clear that the used methods are not yet optimized and having access to expert knowledge from AIMS5.0 partners might lead to new (other choices). In particular, in order to fulfil the ambitions in, e.g., single camera assembly process tracking while safeguarding privacy, combinations of (open access) algorithms might be highly recommended.

We mention here a number of methods and algorithms as used at the start of the AIMS5.0 project, as an example of the status.

1. For our ambitions in intelligent scheduling of Operations in 3D printing factory, we now make use of a rule-based method, which solution is single stage, is quite slow and needs regular full recalculation. The step to an intelligent and adaptive multi-stage and multi-factorial method will require major functional additions and tools.
2. For vision-based process and product quality evaluation we are using open-source ML/A.I. tools such as YOLO, combined with mmpose method (Keypoint) to create privacy-compliant observation and quality evaluation of human activities, human-machine interaction and human-component/product interaction.
3. In our ambition to provide personalized operator guidance using LLMs, initially pre-trained open-source or commercial language models can be used. However, it is essential to augment this with data retrieval and potential fine-tuning to optimize guidance quality, as to prevent hallucinations and build upon company-specific historical data sources.

3.4 AI tools methods and algorithms as used in supply chain planning and quality incident traceability for NXP/Semaku/TUe use case

In general, TUe will develop the machine learning models for the various use cases using Python and the SciKit Learn and PyTorch libraries.

At the moment, several AI initiatives are present throughout NXP. These initiatives are in general decentralized and therefore there is coming more focus on reusability and robustness. As the use case will focus on supply chain planning and quality incident traceability, we will discuss AI initiatives focused on these topics as used at the start of the AIMS5.0 project.

Supply chain planning

1. Forecasted demand is an input parameter for the supply chain planning. Next to statistical methods, regression trees are used to further improve forecast accuracy.
2. In the ambition to reduce the planner's workload, an ongoing project investigates the use of AI to map correlation between planning input parameters and certain changes in the plan.

Quality incident traceability

1. ML techniques are used for image recognition within quality incident traceability. The goal is to categorize the images in different groups when having a specific failure mode to enhance further root-cause analysis.
2. Besides further developing root-cause analysis ML techniques are also used to gain greater quality control during manufacturing. For example, by using sensor data to predict physical characteristics of wafers.

3.5 Tool for robot control task execution

This tool implements the workflow to support the integration of AI algorithms into an automated conveyor feeding robotic system being developed at WP5 and UC6. It should achieve rapid reconfigurability of the system, which will be supported by both a flexible middleware framework and adapted deep-learning-based algorithms, and shall lead to increased productivity, efficiency, and product quality. The robotic system will consist of several highly configurable AI software modules implementing the following features:

a. robot control integration, based on ROS, supporting a deep learning-based perception and control system, to provide the robots with flexible and dynamic perceiving, comprehending, and control abilities (sensory data processing, representation and interpretation, and low-level action implementation).
b. robotic simulation for learning and validation of robot control algorithms,
c. action planning module together with tools for training associated deep learning algorithms therein,
d. high-level AI task instruction with context-aware task execution,
e. tools for performance validation and quality metrics evaluation

3.5.1 Robot Control and Algorithms integration to the general pipeline

In WP2, the integration of various AI tools and methodologies is crucial for the seamless functioning of robotic control within our operational pipeline. We utilize a suite of advanced tools and algorithms to ensure efficient task execution, adaptability, and precision in real-time scenarios.

Following the analysis that we performed so far and is detailed in section 3.13, from an algorithmic and methodological view we plan to employ:

- Coppelia Simulator and Nvidia Isaac. These environments will be the backbone of our experiments and will play a crucial role in the initial development and testing phases. By providing a virtual environment, it ensures that the algorithms are fine-tuned and debugged before being deployed on real machinery.
- Robot Operating System (ROS). Our pipeline leverages ROS for seamless integration and reliable interoperability between complex robotic components. At the same time, ROS is widely adopted by the industry and researchers due to its robustness and high reliability in various robotic applications.
- Proximal Policy Optimization (PPO) and Trust Region Policy Optimization (TRPO) RL algorithms, which are advanced Reinforcement Learning methods that will be central to

the training and policy optimization of our robots. Their inherent capacity to strike a balance between exploration and exploitation makes them ideal for such intricate tasks.

- Imitation Learning and Behavior Cloning. These techniques are employed to replicate human-like behaviors in robots. Leveraging these methodologies allows robots to learn from human demonstrations, accelerating their learning curve and ensuring the replication of human-optimized maneuvers.
- Reward Sketching. This is a novel technique that will be instrumental in shaping the reward structure of our reinforcement learning models, guiding the robot towards desired behaviors.
- Mask R-CNN. The model will be integrated within the Object Detection Module. Mask R-CNN is employed for instance segmentation, which is crucial for identifying and differentiating individual objects in an environment. This granularity is necessary for precise object interaction, especially in cluttered settings commonly found in industrial scenarios.
- PyTorch and TensorFlow: The deep learning models used in various pipeline stages, including the Pose Estimation and Object Grasp modules, are developed with PyTorch. Its flexibility and efficiency facilitate rapid prototyping and experimentation, which are crucial for the iterative nature of our AI model development. Additionally, the use of TensorFlow, and especially its capability to convert models to TensorFlow Lite, offers the ability to transfer the model into a version suitable for devices with limited computational power. This will enable us to execute predictions on edge devices.

With this integrated approach, we aim to build a robust, adaptive, and efficient robotic system that can operate seamlessly in a dynamic environment, taking full advantage of the cutting-edge tools and techniques at our disposal, as shown in the following Figure:

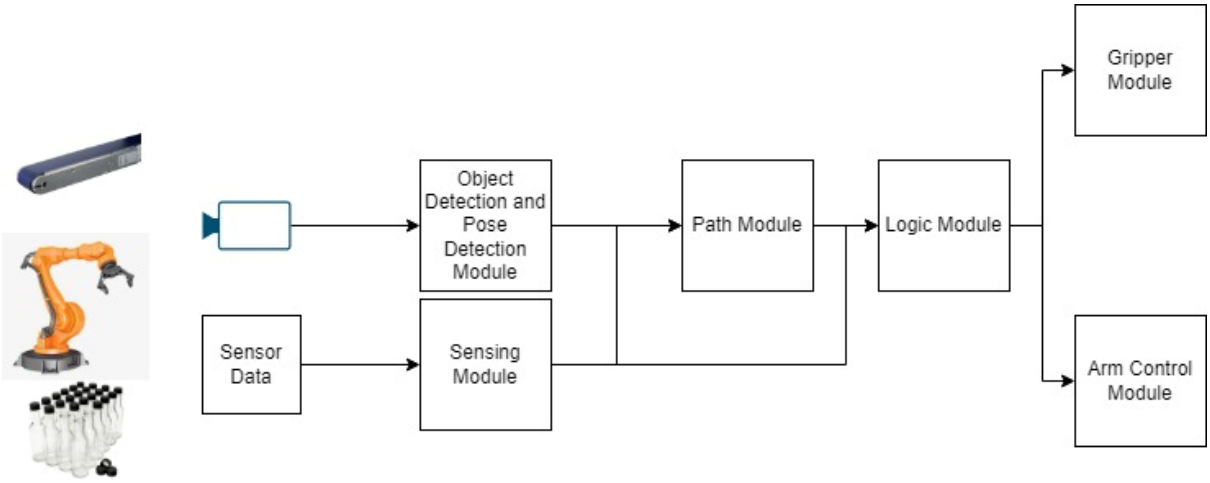

Figure 9: The overview of the proposed modular approach

The provided figure presents the overall control flow and data processing path of our robotic system.

- Sensor Data: Represented by an array of sensors, this data stream is critical for real-time feedback and environmental interaction. The sensory inputs are then processed and analysed by the Sensing Module.
- Object Detection and Pose Detection Module: This module plays an important role in perceiving the environment. It receives visual input from the camera, identifies objects, and determines their pose or orientation. Such detection is often powered by deep learning models like Mask RCNN.

- Path Module: Once the object and its pose are identified, the path module computes the optimal trajectory and sequence of movements needed for the robotic arm to approach and manipulate the object.
- Logic Module: Serving as the brain of the operation, this module integrates the data from the previous modules and orchestrates the overall strategy. It ensures that the tasks are performed efficiently and safely.
- Arm Control Module and Gripper Module: These are the effector modules that execute the decisions of the Logic Module. The Arm Control Module, in conjunction with ROS and a inverse kinematic model, manages the movements of the robotic arm. In contrast, the Gripper Module, using a Pytorch or a TensorFlow model, controls the grasp and release mechanisms.

3.6 Utilizing the copilot concept in AIMS5.0

One forward-thinking way of utilizing Large Language Models (LLM) and other, potentially multi-modal generative AI technologies in a value adding and production-grade use in industry lies in the concept of Copilots. While most knows Copilot as GitHub's product, the underlying concept can be generalized into essentially building chatbot-like interactive interfaces into various platforms, utilizing their internal APIs and services, to replace GUI-based workflows.
Microsoft has been spearheading the concept of Copilots into their Office365[2], Power Platform[3] and Windows[4] platforms. There are various use cases to embed LLM and other generative and non-generative AI into these platforms, among others:

- Text or document summarization or rephrase by an LLM
- Picture, document, slide, website or UI (re)design by generative AI
- "Ask your database or document", providing business insights by AI models

These use cases can extend in usability, if the LLM or genAI algorithms have the ability to (i) interact with various services via (e.g., REST) API-based integrations, (ii) have short and long term memory using e.g. vector databases, (iii) learn from few-shot examples, (iv) be equipped with persona prompts. These capabilities are provided by AI orchestration platforms, such as Semantic Kernel (from Microsoft)[5] or Langchain[6].

In these AI orchestration platforms, the main goal is to create chatbots that can understand complex tasks, plan and execute an appropriate workflow on their own, understand the responses from external services, all based on the input from users. Essentially, these platforms provide:

- managing *Contexts*,
- providing *Connector*s to external services, databases,
- implementing *semantic skills*, i.e., parametrized prompt templates for multiple generative AI models (picture, sound or text generation),
- *native skills*, e.g., accessing and understanding the Microsoft Azure Active Directory Graph API for user information,
- providing long term *global or user-specific searchable memory*, using e.g., vector databases and blob storages,
- *Planner capabilities* that can use e.g., LLM to deduct the necessary workflow needed to complete the task at hand.

[2] https://blogs.microsoft.com/blog/2023/03/16/introducing-microsoft-365-copilot-your-copilot-for-work/
[3] https://learn.microsoft.com/en-us/power-apps/maker/canvas-apps/ai-overview
[4] https://www.microsoft.com/en-us/windows/copilot-ai-features?r=1
[5] https://learn.microsoft.com/en-us/semantic-kernel/overview/
[6] https://python.langchain.com/docs/get_started/introduction

Based on this, we can establish Copilots for CPSoS using the Arrowhead framework. The purpose of this would be the ability to design and manage a SoS via natural language and analyze its status and monitoring data using the potentially multi-modal output of the Copilot(s). We can conceptualize three types of Copilots in Arrowhead, with increasing complexity and value-added:

- **Arrowhead Fundamentals Copilot** providing a chat-based entry point into Arrowhead, feeding on the already existing documentation and publications around the ecosystem, answering design and integration related questions. This can be embedded in the Arrowhead Framework Wiki as a chatbot.
- **Arrowhead Design Copilot** interacting with engineering toolchains to design SoS deployment and underlying industrial automation processes and infrastructure. This Copilot can be integrated into the Arrowhead Engineering Toolchain, interacting with the design tools for CPSoS.
- **Arrowhead Management Copilot** interacting with various Arrowhead Core Systems to analyse and understand the deployment and potentially execute changes via the Arrowhead governing middleware over the SoS. This tool can be embedded as a widget to the Arrowhead Management Tool.

This will demonstrate the industrial Copilot capabilities; however, continuous learning and testing will be necessary to ensure that the evolutionary nature of the orchestration engines, the LLM and generative AI does not decay the Copilots' abilities over time. Therefore, a *new paradigm of ML Ops* will need to be established to support Copilots, that can evaluate the integrations, the skills and the plans executed by Copilots for certain scenarios. This is currently a very new aspect of DevOps, and only a few tools allow such test automation, for example Semantic Kernel seems to, at least, with the feature called 'Prompt flow'[7].

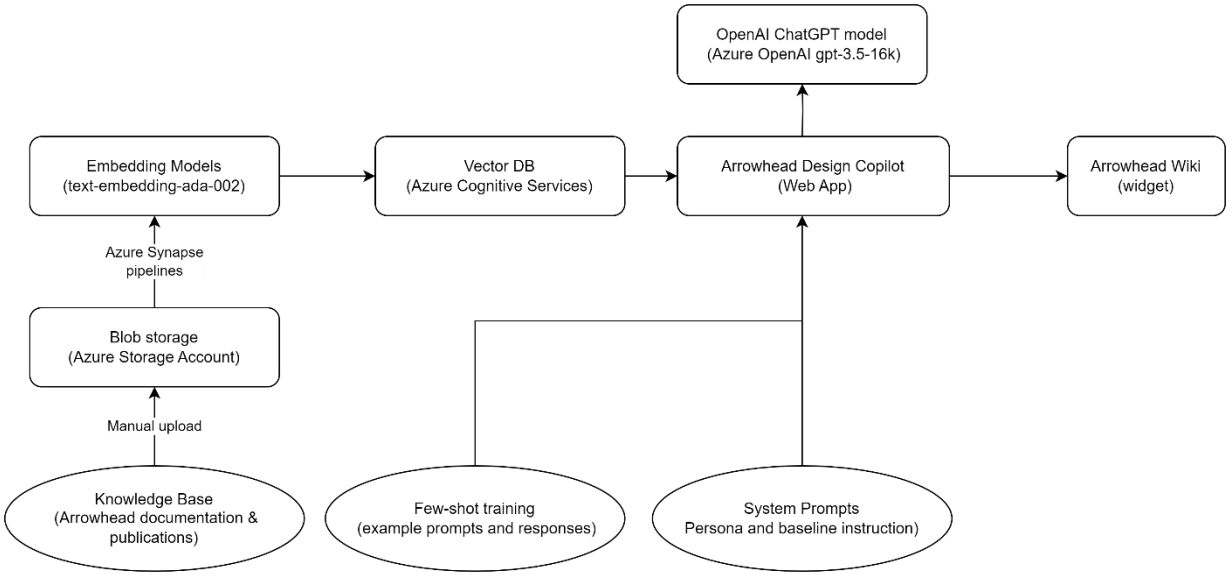

Figure 10: Arrowhead Design copilot in Azure

[7] https://learn.microsoft.com/en-us/semantic-kernel/ai-orchestration/planners/evaluate-and-deploy-planners/

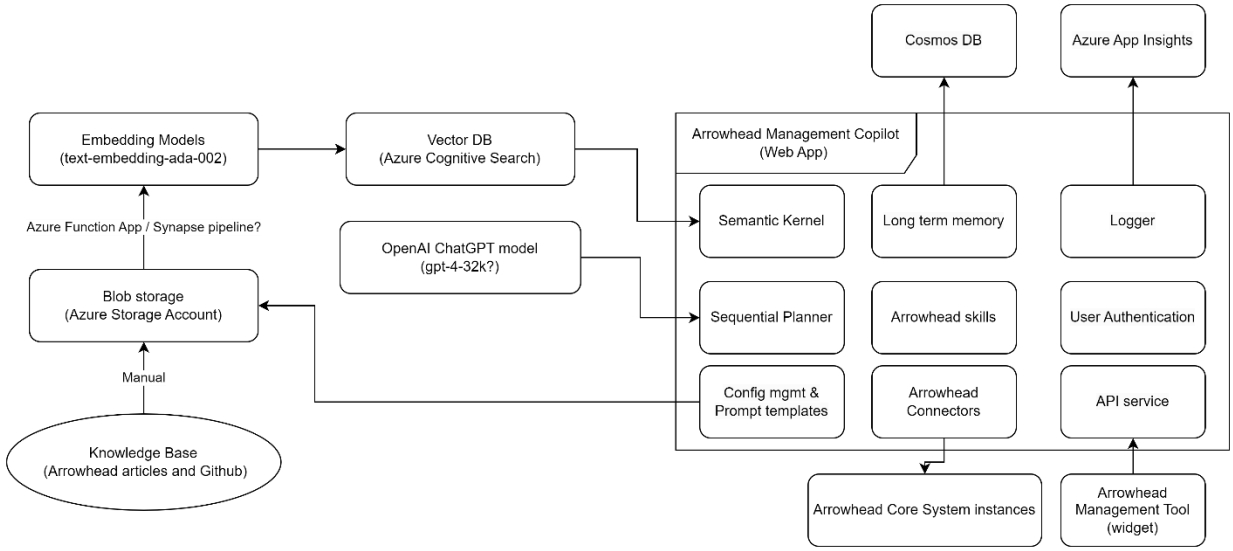

Figure 11: Arrowhead Management copilot in Azure

3.7 Details about PrototAIp

PrototAIp is an open-source platform accessible on Github[8], originally developed by RSA FG during a prior research endeavour and serves as a collaborative software engineering platform, tailored particularly for applications in data science and AI projects. More specifically, it is a specialized platform designed for data scientists, machine-learning engineers, data analysts, data visualizers, and similar professionals. It offers a seamless and adaptable environment for working on projects. PrototAIp has a web-based interactive programming platform based on Jupyter Notebook that is integrated into the final version of the PrototAIp web-based application can be accessed through a browser, granting registered users their own dedicated workspace. Essential software libraries for data science and AI experiments are pre-installed, sparing users the hassle of manual installation and upkeep of virtual environments. It is underpinned by JupyterHub, allowing it to furnish users with Jupyter Notebooks.

The server solution, "JupyterHub" from Project Jupyter, is adept at managing a substantial user base and can be expanded in capacity as needed. Being open-source, it provides the flexibility to tailor it to the specific needs of the PrototAIp project. This hub solution seamlessly integrates administrative functions, a database system, and authentication processes to generate new notebooks for users (Figure below). This decision to utilize an established tool was a prudent one, as it spared us the effort of creating something from scratch, ensuring a reliable and steady platform for our users.

[8] PrototAIp: https://github.com/Researchstudio-DSc/prototAIp, accessed Oct 12, 2023.

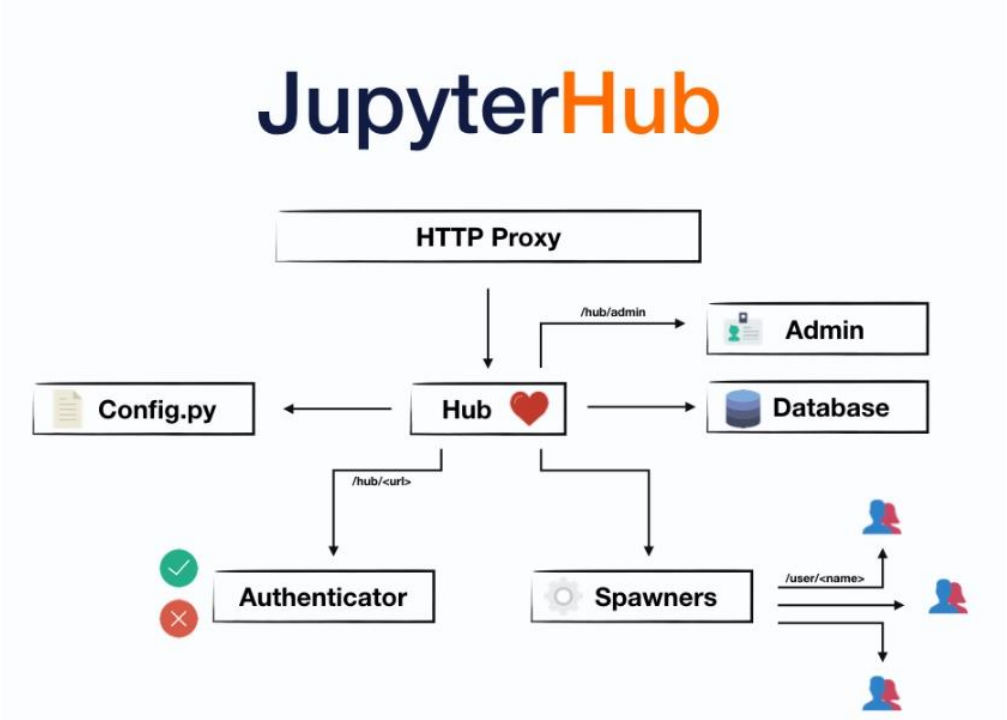

Figure 12: JupyterHub's Architecture

Adhering to our principle of "Batteries included," we've preloaded our notebooks with the most important libraries for data science. This means no waiting time for our users; they can get started straight away. To determine the most popular libraries, we conducted a systematic search on the internet, looking for websites and scientific publications specifically addressing Python libraries for data science. In the end, we identified 11 websites (the complete list can be found at the end of this post) and referenced 2 papers (Nagpal & Gabrani, 2019; Stančin & Jović, 2019) on the subject.

The outcome was an extensive list of intriguing libraries, totalling 43 in all. Below is a list of the top 10 identified libraries, ordered by mentions:

- NumPy: 13
- Pandas: 13
- Matplotlib: 12
- SciKit-Learn: 12
- TensorFlow: 11
- SciPy: 10
- Keras: 9
- Seaborn: 9
- Plotly: 7
- PyTorch: 5

The remaining libraries cover a wide range of data science topics. There are other toolboxes for machine learning and data visualization, but also highly specialized libraries. We recommend taking a closer look at all of them to get the best possible overview:

Library	Mentions	Library	Mentions	Library	Mentions
NumPy	13	Bokeh	3	Keras	1
Pandas	13	Gensim	3	Kivy	1
Matplotlib	12	GGplot	2	mlpy	1
SciKit-Learn	12	OpenCV	2	mlxtend	1
TensorFlow	11	PyBrain	2	XGBoost	1
SciPy	10	Caffe	1	PyCaret	1
Keras	9	Caffe2	1	Pydoop	1
Seaborn	9	Cython	1	pydot	1
Plotly	7	Dask	1	PyQt	1
PyTorch	5	Dumbo	1	PyTables	1
BeautifulSoup	4	Flask	1	Shogun	1
NLTK	4	h5py	1	Spark	1
Scrapy	4	Hadoop	1	Tabel	1
Statsmodels	4	Hadoopy	1	Theano	1

Table 1: Full list of libraries with score

Additionally, the necessary software for setting up an on-premises installation is readily accessible on Github. This can be easily deployed using Docker and can be up and running within a matter of minutes.

3.8 Methods and Tools for Facility Management 5.0 (ViF, IFD, IFAT)

In accordance with our 4-step plan with increasing complexity, we are starting with a regular Exploratory Data Analysis (EDA) in the performance monitoring step. Based on these findings an easy usable dashboard could be a helpful tool to involve experts in the field of cooling systems, leading to more interaction and knowledge exchange for the subsequent tasks like anomaly detection and identification of performance degradation.

In order to tackle the challenge of detecting anomalies in the available data, mostly time series data, different algorithms are considered. As statistical methods mostly rely on the assumption that data comes from a particular distribution and clustering-based methods are not capable of including the time related correlations we will follow the trail for other areas of algorithms, like the field of deep learning methods. Research on Generative Adversarial Networks (GANs) [Goodfellow2014], also in combination with Transformer [Vaswani2017] show some promising results and VIF will be trying to apply it to find anomalies in the cooling system [Xu2022].

In the next step, data and insights obtained in the preceding work will be used to have a look at energy optimization potentials, especially with the possible help of different forms of visualization and simulation tools or approaches. The specific types will turn out during the process.

After looking at the past, the focus shifts to possibilities to predict certain events in the future. Algorithms like Autoencoders [Kramer1991], LSTM Autoencoders [Bampoula2021] and XGBoost [Chen2016] will be considered to predict malfunctions or potential failures for example.

The final step is to use all the gathered data and information to support lifecycle management, especially predicting component life. The above-mentioned algorithms also have potential to support this step and deliver some meaningful results.

3.9 MLOps Techniques for Industry5.0

Although using AI/ML techniques in the Industry5.0 era, there are only a few ML initiatives make it to a production-grade implementation, mostly due to lacking proper Continuous Integration and Delivery framework and MLOps (Machine Learning Operations) practices. This is especially true for industrial use cases, where the trust and reliability of ML applications are mission critical. Most of these applications fail during the final stage of the development lifecycle, i.e., acceptance testing and validation of the ML application, while being integrated into Cyber-Physical System of Systems (CPSoS).

The primary purpose of the study by BME - based on the AIMS and research literature surveys -- is to tailor and customize Machine Learning Operations (MLOps) practices to meet the specific requirements and challenges posed by Industry 5.0. In this initial study by, in line with WP2 plans, we explored the key requirements for deploying ML applications in industrial scenarios, emphasizing the critical role of Digital Twins, edge AI, and responsible-explainable AI techniques in ensuring efficient and responsible operations. Building upon previous models, we proposed two process models: (i) the Olympics model for MLOps-coupled CPS engineering and (ii) the MLOps engineering toolchain for industrial applications. The main ideas are described below, whereas the details can be found in [Hegedus2023].

As the proposed process goes, the whole MLOps lifecycle needs to be augmented with the requirements driving the Digital Twin and Systems Engineering design flow. As described earlier, the challenge is bringing the Arrowhead Tools Engineering Process and other CPSoS engineering standard development processes together with the ML-Dev-Ops cycles of the organization. Building on these methods and approaches, we propose the extension of the DevOps and MLOps process with two additional aspects: managing the Digital Twin lifecycle and integration of the DevOps aspects with the industrial Systems Engineering process.

This newly proposed process model - resembling the five circles of the Olympics - is illustrated below. This newly proposed “Olympics model for MLOps-coupled CPS engineering” are described in detail within [Hegedus2023], that is being part of the deliverable.

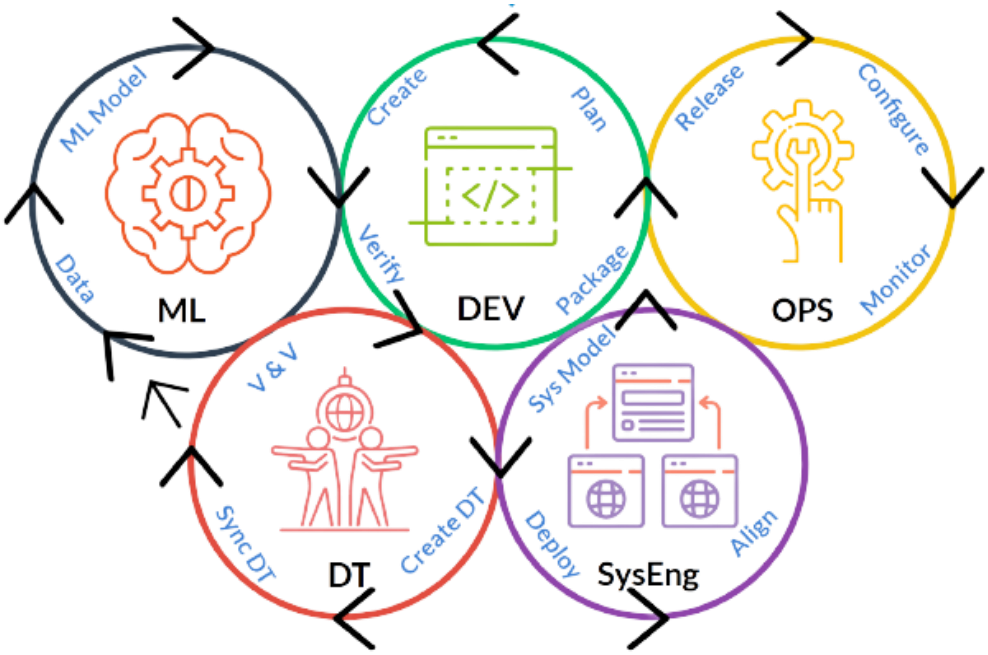

Figure 13: Olympics model for MLOps-coupled CPS engineering

Besides, the “MLOps Engineering Toolchain for Industrial Applications” is also proposed in [Hegedus2023].

Based on the AIMS5.0 requirements, design adaptation is required for the "de-facto standard" MLOps solution architecture. The proposed solution is presented by the below figure. This takes into account the current state-of-the-art, as well as considerations made for Digital Twins, ML Ops practices, the involvement of SaaS solutions, and the current engineering process for CPSoS. In this figure, blue colored components represent the main integration points (i.e. source and destination systems) of the CPSoS; the "observability" components are highlighted with yellow color (i.e. Digital Twin), while the data platform components (green color), the ML Ops CICD (orange) and the ML infrastructure (red) are also depicted in one continuous cycle of data (and model) flow. In this proposed toolchain, the "de-facto" lifecycle is continuously executed on:

- the ML Infrastructure,
- the Cyber-Physical Production Systems elements,
- the Data Platform,
- the Digital Twin,
- the AI Gym.

Further details can be found in [Hegedus2023], as part of this deliverable.

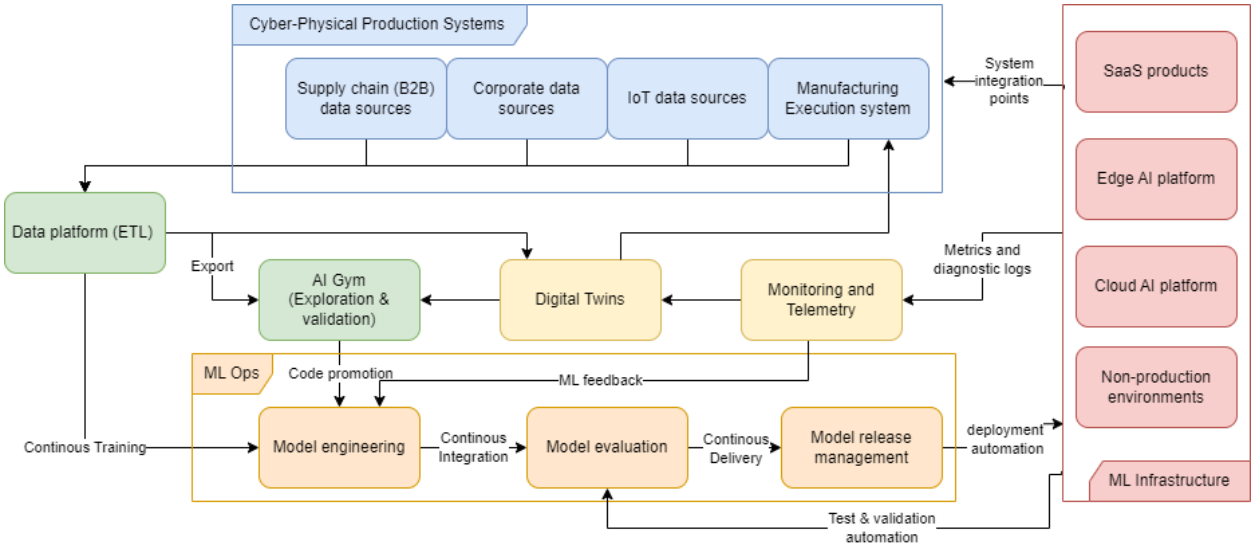

Figure 14: Design adaptation for “de-facto standard” MLOps architecture

4. The vision of the AI Toolbox

Even though by 2023 AI assisted industrial applications are spreading rapidly, there are still several challenges ahead with their deployment and integration. The Arrowhead Framework is helping the industrial systems to be designed loosely coupled, however, integrating AI tools into the framework promises a shorter time-to-market in intelligent industrial services. We created 2 questionnaires to gather information for the formulation of the AI toolbox concept. Based on the results we formulated the main objectives, scales, implementation challenges and life-cycle of the AI Toolbox concept. This topic has been published as a conference paper on the 19th International Conference on Network and Service Management.

4.1 Questionnaires

4.1.1 Questionnaire 1

The goal of Questionnaire 1 was to discover collaborating partners, identify potential future collaborations, find data owners and data sharing potential. We got 22 responses from 17 different partners. This section presents the main conclusions of the Questionnaire 1.

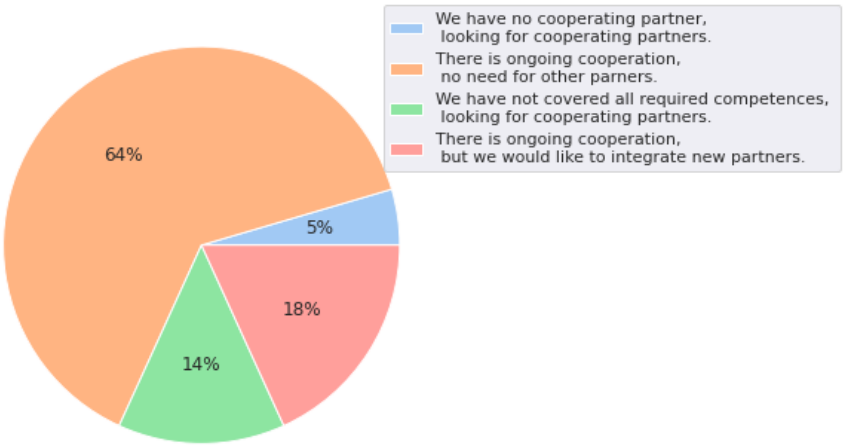

Figure 15: Collaboration potential of the partners

Figure 15 highlights the collaboration need of the partners, which shows that 64% of the partners (who filled the questions) do not need cooperating partners.

Data sharing potential and data origins are highlighted in Figure 16Figure 17. Unfortunately, more than half of the partners do not have data or cannot share it. For those who have data, the main data source is direct sensor data, but there are significant data source from system log files, monitoring data and supply chain reports. Besides, there were some contradictions in the answers as follows:

- One state that they have no data but they give answer to the data origin question.
- One state that they have no data but it is confidential.
- One state that they have no data but they can share it.

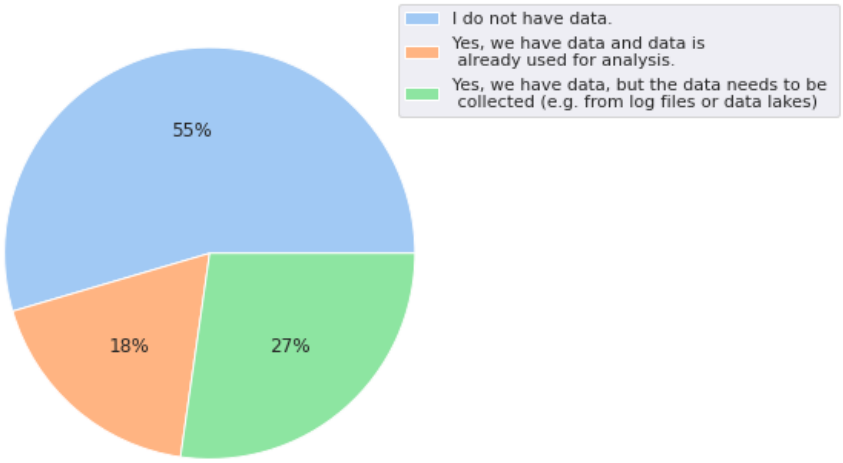

Figure 16: Data sharing potential

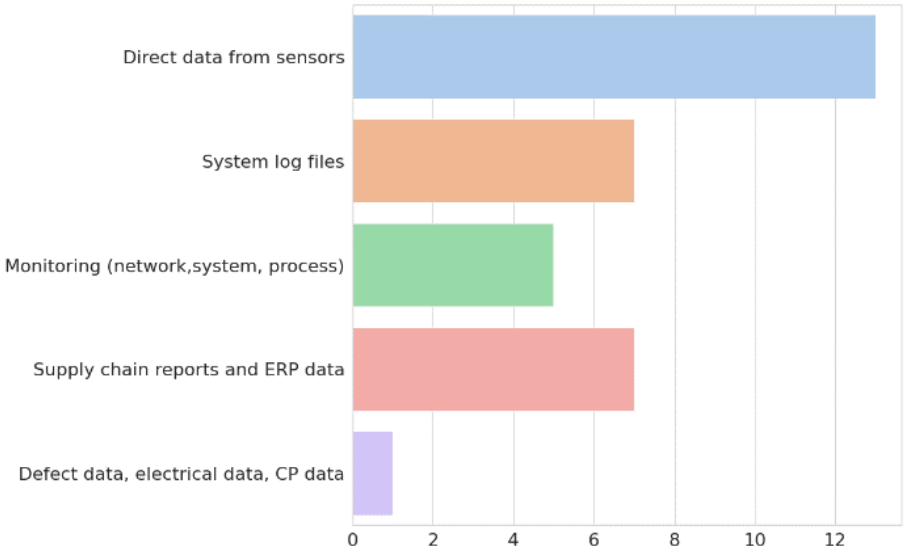

Figure 17: Possible data origins

4.1.2 Questionnaire 2

Questionnaire 2 consists of questions regarding the use-cases, AI competence and the AI toolbox. The goal of Questionnaire 2 was to identify concrete AI-related task regarding the use-cases and enquire how the partners envision the AI toolbox. We got 26 responses from 24 different partners. This section presents the main conclusions of the Questionnaire 2.

Figure 18 highlights the main industrial domains regarding the use case according to the partners. The figure shows that manufacturing, operation and management, logistics and consulting are the main industrial domains. The second part of the questionnaire presents how the partners envision the AI toolbox. The figure shows that most partners define the AI Toolbox as a collection of prewritten sources of functions and algorithms.

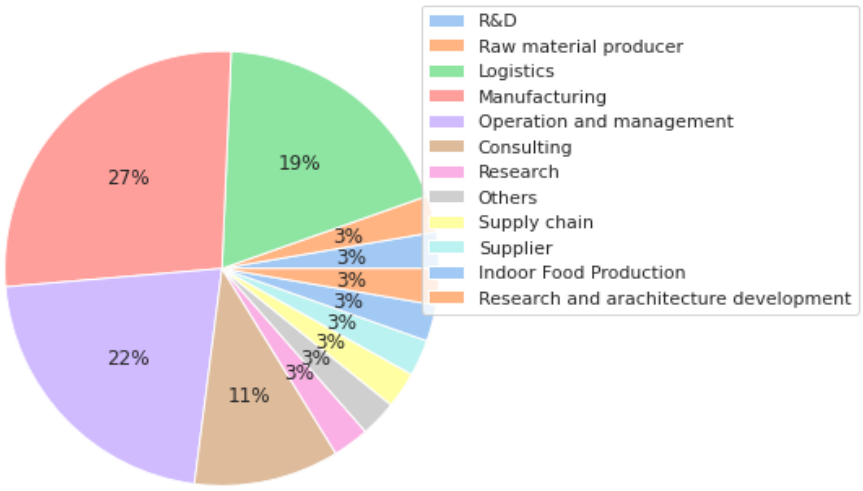

Figure 18: Industrial domains of the use cases

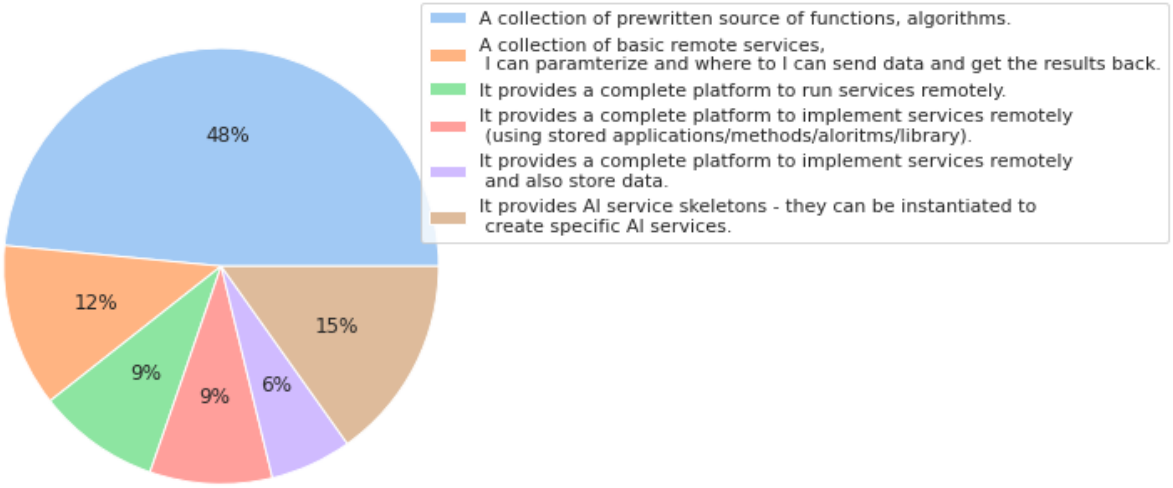

Figure 19: The envisioned definition of the AI Toolbox according to the partners

4.2 AI toolbox objectives

The main purpose of the AI Toolbox is to provide interchangeable, re-usable AI services to facilitate the implementation of intelligent services in the Arrowhead Framework. The objectives of the AI Toolbox can be summarised as follows:

- **Decrease the time-to-market**: Support fast development of intelligent services. This requires re-usable services and easy deployment.
- **Reliability and security:** Since AI tools handle potentially sensible data, the AI Toolbox needs to provide means to prevent unauthorized access to data.
- **Scalability:** The computing needs of AI algorithms are so high that scalability of the algorithms is a great concern.
- **Upgradeability**: Modern AI algorithms change rapidly, or the existing algorithms learn from new data, so the state-of-the-art models must be continuously deployed, especially in the Edge AI paradigm.

While Arrowhead Framework specifies a couple of lightweight requirements against application services, like loose coupling, and late binding, it states no rigid bounds to implement application services. However, for AI services, some additional requirements can be stated based on the objectives of the AI Toolbox presented earlier.

- **Re-usability:** To decrease development time, AI services should be implemented in a specific granularity which makes the services general enough to be utilized in different use cases but specific enough to implement complex tasks.
- **Composeability:** Rapid development of industrial use cases requires that the AI services use common data models and interfaces to combine them to implement more complex tasks easily.
- **Upgradeability:** Changes in objectives or learning from new data result in new models, which should be deployed seamlessly and easily.
- **Heterogeneity**: Heterogeneity should be required in multiple senses against AI services. Heterogeneity in platforms, implementations, and all other factors related to reliability shall be considered while designing AI services.

4.3 AI toolbox concept

The implementation of AI services to extend the Eclipse Arrowhead framework has different scales or entry levels. There are a couple of trade-offs between different levels, regarding scalability, granularity, re-usability, and development time. E.g., a specific use-case can be implemented using different existing AI solutions -- e.g., TensorFlow models -- as an Arrowhead service. However, this service will provide only functions to fulfil a specific use case, preventing re-usability. Also, providing clear and standard services ease the problem of vendor lock-in phenomena by providing interchangeable services.

The different AI Toolbox scale levels can be seen in Figure 20, where the concept of the Arrowhead AI Toolbox is depicted. There are four levels of AI Service scale, differing primarily in granularity. The three implementation-based levels are supported by the fourth -- mostly theoretic -- methodology scale. In the following, the four levels are presented more deeply and compared to the objectives and requirements.

A detailed version of this concept, further describing how the Eclipse Arrowhead framework can be extended with the AI toolbox, is described in [Ficzere2023], as part of D2.1.

4.3.1.1 Application scale

Application scale is the highest scale level of AI services. Application scale means that AI services provide solutions for complete use cases, e.g., a whole predictive maintenance service specifically for railway switches. The service is implemented as an Arrowhead application service and can be accessed through a specified, custom interface. However, services at this scale can hardly fulfil the requirements for re-usability since every use case needs its own service to be implemented using standard tools. Of course, there can be quite general use cases -- such as geofencing -- where it is easier to transform the input data to according to the application service interface requirements and use the Application scale implementation rather to implement the whole application.

4.3.1.2 Tools scale

Tools scale is a compromise between algorithmic scale and app scale. Tools basically provide applied AI methods, which are commonly one or more AI models combined with some inner logic to solve a recurring problem. E.g., localization of assets in a factory consists of a couple of small steps, but the problem arises in numerous use cases. Tools scale can provide excellent reusable scale and support rapid development. Also, careful design of service interfaces can provide good composability.

4.3.1.3 Algorithmic scale

The level of Algorithmic scale is granular since well-known AI and ML algorithms are implemented as AI services. Algorithmic scale is mostly about choosing the best service interfaces to provide composable services. However, using pure algorithms -- while being highly re-usable -- requires great effort to implement complex use cases. Also, using pure AI algorithms requires great knowledge of AI methods and techniques, which is against rapid development.

4.3.1.4 Methodology scale

Being on a theoretical scale, the methodology provides specifications, requirements, and best practices for implementing AI services. However, while the Methodology scale can also be considered an independent level of AI service scales, it is best to imagine it as a basis of all other service scale levels since it can provide means to fulfill all the objectives and requirements of AI Toolbox services.

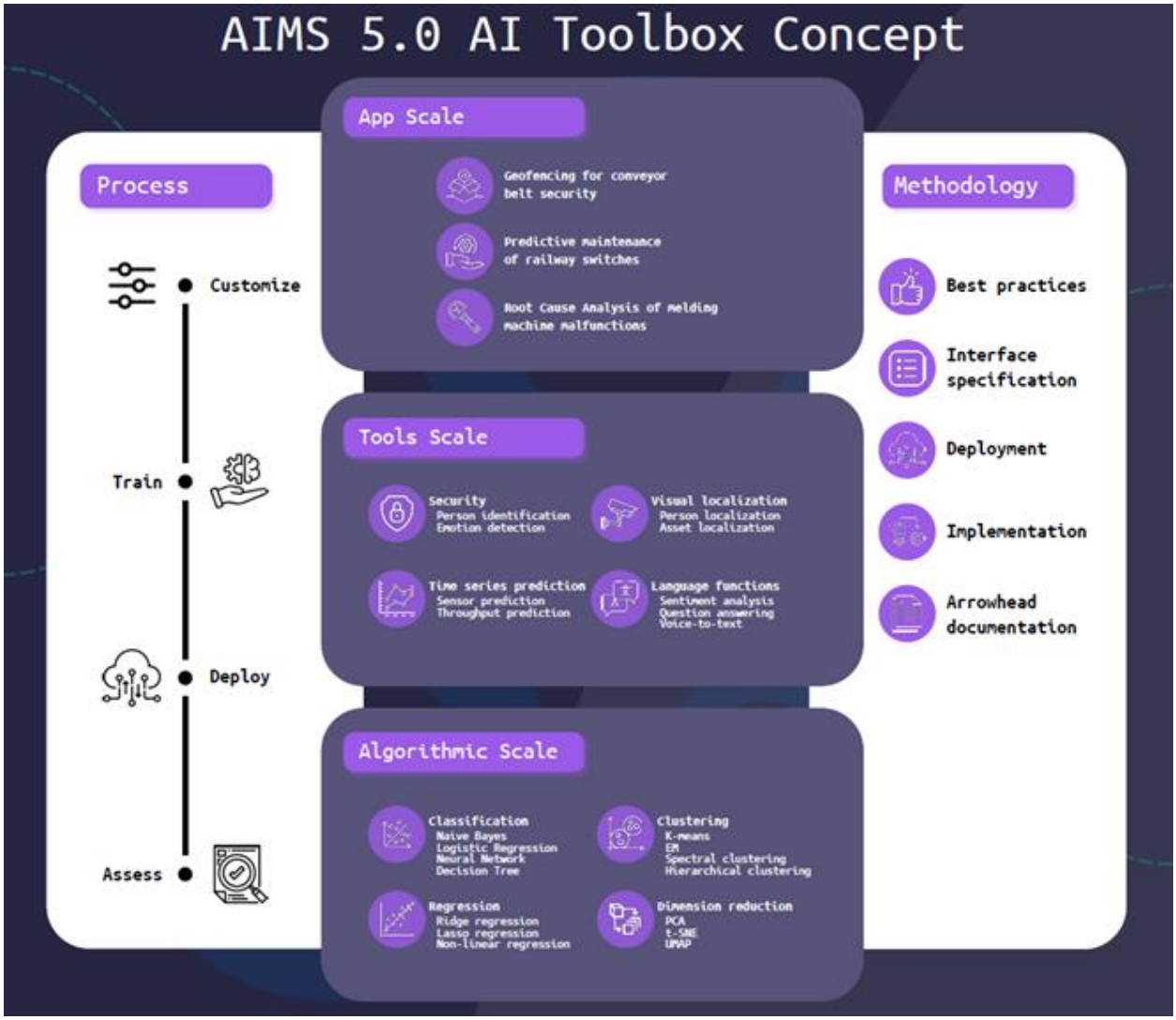

Figure 20: AI Toolbox concept

4.3.2 AI Toolbox lifecycle

The lifecycle of AI services shares the very same four steps (see Figure 20). To provide re-usable services, the customization step makes it possible to tailor the service to special needs. This includes, e.g., the settings of the parameters. The next step is optional, however, most algorithm requires training to be able to perform specific tasks, e.g., detecting uncommon objects in an image. Training can be complex and shall support widely spread frameworks to efficiently accomplish fine-tuned models. Deployment is a crucial point in the lifecycle, since models require variable resources or perhaps massive parallelization. Also, real-time upgradeability needs the necessary tools to replace old models with new models on-the-fly. After deployment, the AI service is assessed to prove it fulfils its function, meeting the expected metrics in the application context. The lifecycle might start from the beginning, or it can be periodic, as models are fine-tuned from time to time.

4.3.3 AI Toolbox implementation

While having different AI Toolbox scale levels, it does not mean that one of them must be chosen. In the implementation of the AI Toolbox, all scale levels can be implemented side by side to provide different services for different needs. Also, methodology shall be provided at all scale levels to provide integrated and unified interfaces and promote heterogeneity while implementing AI services. However, hierarchical implementation (i.e., task scale uses algorithm scale services) is not always recommended because of performance issues.

The most promising implementation scale is the tools scale: it provides the best re-usability while requiring little knowledge of AI methods. However, to satisfy composability, standardized and clean service interfaces shall be designed. Here, the methodology can provide guidance to best practices of interface development. Also, unified and independently designed interfaces make it possible to create interchangeable services with different implementations -- helping to avoid vendor lock-in and ensure reliability.

4.3.4 Edge AI concerns

In our discussion Edge AI refers to AI models running on edge infrastructure. State-of-the-art studies show how to create, deploy and operate these AI models, investigating algorithm performance, cost-effectiveness, privacy, reliability, and efficiency challenges. Also, there are recent research on domain-specific Edge AI usage, utilizing AI for mobile networks, especially for future 6G networks.

State-of-the-art Edge AI solutions pose challenges to the design of the Arrowhead AI Toolbox, namely, three following concerns can be formulated. First, Edge AI solutions extensively support real-time on-the-fly model deployment to provide the most up-to-date AI models at the edge. These models are mainly trained in the cloud but deployed onto edge devices automatically through numerous available platforms. However, Arrowhead lacks these features yet, and it is not decided how to implement on-the-fly model upgrade. Second, edge devices have limited resources and use special platforms to run AI models. Also, training and deployment of edge AI models split between the cloud and the edge device. While Arrowhead supports devices with limited capabilities, interoperability with edge AI platforms shall be investigated. Third, distributed learning and -- in particular -- federated learning is specific to Edge AI. Distributed and federated learning helps to employ the resources of numerous edge devices, converting limited resources into a serious computing capability. Especially federated learning also helps to realize the security objective -- no other device gets information from data handled by edge devices. However, the implementation of distributed learning is also an area to be investigated in Arrowhead Framework.

5. Conclusions

The below table summarizes the partner contributions to the AI toolbox in WP2:

Partners	Use case	Topic	AI related domains	Technology advancement
Lfoundry Statwolf Università degli Studi di Padova	WP6 UC15	Anomaly detection	• Explainable Artificial Intelligence (XAI) • AI continuous learning • Deep learning from heterogeneous data	DIFFI: Depth-based feature importance of Isolation Forest • ACME: Accelerated Model-agnostic Explanations • SHAP: SHapley Additive exPlanations • Domain Incremental Learning in multi-label Alarm Forecasting scenario • Transfer learning with custom strategies • Domain Adaptation Neural Networks
Fabmatics	UC14	Intelligent mobile transport systems fleet control	• Route and utilization optimization • Computer vision for navigation • Digital Twin	• Auto-teaching of carrier vehicle positions and robot arm positions • Location recognition based on vision system • Creating digital twin model with integration to CAD
Signify TU/e		Operational Excellence	• Manufacturing schedule optimization • Computer vision for part identification and QC • Retrieval augmented generation using LLM	• Combining historical and real time data streams for continuous learning • Fine-tune of generative LLM • Privacy compliant evaluation of human activities in computer vision
EDI HUA	WP5 UC6	Robot control	• Robot control and action planning algorithms • Robotic computer vision • Explainable AI • Code synthesis using codex LLM • ML Ops in robotics and ROS	• Context- and policy-aware action planning • Behaviour cloning and imitation learning • Explainable AI for robotics • Task description based "prompt to action" • Test-driven development in ML Ops (TDD) • Multi-model ML orchestration for robotic tasks
NXP Semaku TU/e	WP6 UC19	Planning in OEM supply chains, QC incident predicction	• Deep Reinforcement Learning (DRL) • Digital Twin for production	• Embedding reinforcement learning in distributed decision making and agents

KAI **IFAT**	WP6 UC15	Defect image classification Remaining Useful Life estimation	• AI model evaluation techniques • Data preparation in corporate environment • Data-centric for computer vision • Unsupervised learning for anomaly detection	• Human-in-the-loop data pipelines for collecting defect density data and expert labelling • Model monitoring, explainability and evaluation frameworks • Interpretability and weakly-supervised learning for RUL estimation
BME		AI toolbox Arrowhead Copilot	• AI techniques and best practices • Context-aware and plugin-enabled LLM-based application development • Copilots in CPSoS engineering	• Consolidate best practices for industrial use
AITIA		AI applications in CPSoS and Arrowhead	• AI application packaging and serving	• AI integration into CPSoS engineering
Gdansk University TTTech	UC11	Wireless communication resilience in production environments	Wireless comms noise reduction using AI • Classification models	technology validation and verification
RSA		PrototAIp	ML Ops	• AI development sandbox practices
PCL		Consumer data analytics Intelligent process control and QC Predictive maintenance	• Computer vision for QC • Optimalization algorithms (supply chain)	• technology validation and verification
IDEKO **SORALUCE** **GOIMEK** **SAVVY**	UC5	Predictive maintenance Anomaly detection	• Classification and forecast algorithms	• High-frequency time series data use in ML
FuH **IFD** **IFAT** **Systema** **ViF**	UC9	Machine utilization optimization	• Genetic algorithms for scheduling	• technology validation and verification
IFAT **IFD** **Systema**	UC10	MES		• digitalization, machine learning and AI methods within the MES

WUW HSZG Systema IFD IFAT IFAG	UC12	Semiconductor manufacturing network optimization	• LSTM, XGBOOST, Arima explainability	• Explainability of CNN algorithms
HoST IFD IFAG IFAT FMX TUD	UC13	Wafer transportation and storage	• Predictive maintenance, supported by edge AI	• Detection of machinery wear or faults • Fast detection of material problems on high volume components
FMX IFD IFAT IFAG TUD HoST	UC14	Virtual commissioning with Digital Twins	• Models for automating manual procedures • Computer vision	• Digital Twins • Virtual commissioning
FMX, IFD, IFAT, IFAG, TUD, HoST	UC15	Productionized Machine Learning at scale	• MLOps toolchains • Explainable AI	• Explainability of CNN algorithms • Comparison of DL architectures (GRU, CNN-GR, CEEMDAN GRU) to find best fit for given tasks
Thales Forth	UC16	SATCOM and radar RF transceivers	• Supervised learning for RF design aid	Utilizing the appropriate training procedures in collaboration to commonly adopted approaches and algorithms that support diverse supervised learning options
FHG IISB PFV IFD IFAT IFAG Systema	UC17	Gas analysis in semiconductor factories	• Data preparation, fusion and analysis	• Sensor data fusion and analysis

6. References

[Carletti2023] Carletti, Mattia, Matteo Terzi, and Gian Antonio Susto. "Interpretable anomaly detection with diffi: Depth-based feature importance of isolation forest." Engineering Applications of Artificial Intelligence 119 (2023): 105730.

[Christ2018] Christ, M., Braun, N., Neuffer, J., & Kempa-Liehr, A. W. (2018). Time series feature extraction on basis of scalable hypothesis tests (tsfresh–a python package). Neurocomputing, 307, 72–77. https://doi.org/10.1016/j.neucom.2018.03.067

[DallePezze2023] Dalle Pezze, D., Deronjic, D., Masiero, C., Tosato, D., Beghi, A., & Susto, G. A. (2023). A multi-label Continual Learning framework to scale deep learning approaches for packaging equipment monitoring. Engineering Applications of Artificial Intelligence, 124, 106610.

[Dandolo2023] Dandolo, D., Masiero, C., Carletti, M., Dalle Pezze, D., & Susto, G. A. (2023). AcME—Accelerated model-agnostic explanations: Fast whitening of the machine-learning black box. Expert Systems with Applications, 214, 119115.

[Gentner2021] Gentner, N., Carletti, M., Kyek, A., Susto, G. A., & Yang, Y. (2021). DBAM: Making virtual metrology/soft sensing with time series data scalable through deep learning. Control Engineering Practice, 116, 104914.

[Hariri2019] Hariri, S., Kind, M. C., & Brunner, R. J. (2019). Extended isolation forest. IEEE transactions on knowledge and data engineering, 33(4), 1479-1489.

[Liu2008] Liu, Fei Tony, Kai Ming Ting, and Zhi-Hua Zhou. "Isolation forest." 2008 eighth ieee international conference on data mining. IEEE, 2008.

[Lorenti2023] Lorenti et al., (2023). CeRULEo: Comprehensive utilitiEs for Remaining Useful Life Estimation methOds. Journal of Open Source Software, 8(88), 5294, https://doi.org/10.21105/joss.05294

[Psarommatis 2022] Foivos Psarommatis, João Sousa, João Pedro Mendonça & Dimitris Kiritsis (2022) Zero-defect manufacturing the approach for higher manufacturing sustainability in the era of industry 4.0: a position paper, International Journal of Production Research, 60:1, 73-91, DOI: 10.1080/00207543.2021.1987551

[Susto2014] Susto, G. A., Schirru, A., Pampuri, S., McLoone, S., & Beghi, A. (2014). Machine learning for predictive maintenance: A multiple classifier approach. IEEE Transactions on Industrial Informatics, 11(3), 812–820. https://doi.org/10.1109/tii.2014.2349359

[Weis2016] Weiss, Karl, Taghi M. Khoshgoftaar, and DingDing Wang. "A survey of transfer learning." Journal of Big data 3.1 (2016): 1-40.

[Amarjyoti2017] Amarjyoti, S. (2017). Deep reinforcement learning for robotic manipulation-the state of the art. arXiv preprint arXiv:1701.08878.

[Chen2017] Chen, Q., Heydari, B., & Moghaddam, M. (2021). Leveraging task modularity in reinforcement learning for adaptable industry 4.0 automation. Journal of Mechanical Design, 143(7), 071701.

[Choi2020] Choi, J., Kim, H., Son, Y., Park, C. W., & Park, J. H. (2020, October). Robotic Behavioral Cloning Through Task Building. In 2020 International Conference on Information and Communication Technology Convergence (ICTC) (pp. 1279-1281). IEEE.

[Fang2019] Fang, B., Jia, S., Guo, D., Xu, M., Wen, S., & Sun, F. (2019). Survey of imitation learning for robotic manipulation. International Journal of Intelligent Robotics and Applications, 3, 362-369.

[Fernandez-Chaves2022] Fernandez-Chaves, D., Ruiz-Sarmiento, J. R., Jaenal, A., Petkov, N., & Gonzalez-Jimenez, J. (2022). Robot@ VirtualHome, an ecosystem of virtual environments and tools for realistic indoor robotic simulation. Expert Systems with Applications, 208, 117970.

[Ficzere2023] Ficzere, D., Hollósi, G., Frankó, A., & Varga, P. (2023, October). AI Toolbox Concept for the Arrowhead Framework. In 2023 IEEE International Conference on Network and Service Management (NOMS), AnServApp WS.

[Geng2023] Geng, Y., An, B., Geng, H., Chen, Y., Yang, Y., & Dong, H. (2023, May). RLAfford: End-to-End Affordance Learning for Robotic Manipulation. In 2023 IEEE International Conference on Robotics and Automation (ICRA) (pp. 5880-5886). IEEE.

[Hegedus2023] Hegedus, C., & Varga, P. (2023, October).Tailoring MLOps Techniques for Industry 5.0 Needs. In 2023 IEEE International Conference on Network and Service Management (NOMS), AnServApp WS.

[Joshi2020] Joshi, S., Kumra, S., & Sahin, F. (2020, August). Robotic grasping using deep reinforcement learning. In 2020 IEEE 16th International Conference on Automation Science and Engineering (CASE) (pp. 1461-1466). IEEE.

[Pang2023] Pang, G., Shen, C., Jin, H., & van den Hengel, A. (2023, August). Deep weakly-supervised anomaly detection. In Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (pp. 1795-1807).

[Saeed2020] Saeed, M., Nagdi, M., Rosman, B., & Ali, H. H. (2021, February). Deep reinforcement learning for robotic hand manipulation. In 2020 International Conference on Computer, Control, Electrical, and Electronics Engineering (ICCCEEE) (pp. 1-5). IEEE.

[Singh2022] Singh, B., Kumar, R., & Singh, V. P. (2022). Reinforcement learning in robotic applications: a comprehensive survey. Artificial Intelligence Review, 1-46.

[Xiang] Xiang, F., Qin, Y., Mo, K., Xia, Y., Zhu, H., Liu, F., ... & Su, H. (2020). Sapien: A simulated part-based interactive environment. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 11097-11107).

[Zhang2015] Zhang, F., Leitner, J., Milford, M., Upcroft, B., & Corke, P. (2015). Towards vision-based deep reinforcement learning for robotic motion control. Australasian Conference on Robotics and Automation, ACRA 2015, Canberra, Australia.

[Zhao2020] Zhao, W., Queralta, J. P., & Westerlund, T. (2020, December). Sim-to-real transfer in deep reinforcement learning for robotics: a survey. In 2020 IEEE symposium series on computational intelligence (SSCI) (pp. 737-744). IEEE.

[Zhou2022] Zhou, Q., Li, X., He, L., Yang, Y., Cheng, G., Tong, Y., ... & Tao, D. (2022). TransVOD: end-to-end video object detection with spatial-temporal transformers. IEEE Transactions on Pattern Analysis and Machine Intelligence.

[Zhou2023] Zhou, Y., Sonawani, S., Phielipp, M., Ben Amor, H., & Stepputtis, S. (2023). Learning modular language-conditioned robot policies through attention. Autonomous Robots, 1-21.

[Bampoula2021] Bampoula, X., Siaterlis, G., Nikolakis, N., & Alexopoulos, K. (2021). A deep learning model for predictive maintenance in cyber-physical production systems using lstm autoencoders. Sensors, 21(3), 972.
[Chen2016] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. arXiv 2016. arXiv preprint arXiv:1603.02754, 11.

[Goodfellow2014] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. Advances in neural information processing systems, 27.

[Kramer1991] Kramer, M. A. (1991). Nonlinear principal component analysis using autoassociative neural networks. AIChE journal, 37(2), 233-243.

[Vaswani2017] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. Advances in neural information processing systems, 30.

[Xu2022] Xu, L., Xu, K., Qin, Y., Li, Y., Huang, X., Lin, Z., ... & Ji, X. (2022). TGAN-AD: Transformer-Based GAN for Anomaly Detection of Time Series Data. Applied Sciences, 12(16), 8085.