Dissertation

Knowledge-based Maintenance Framework for Smart Manufacturing

Advancing Efficient Planning and Cognitive Assistance to Enhance System Availability

ausgeführt zum Zwecke der Erlangung des akademischen Grades eines

Doktors der technischen Wissenschaften

an der TU Wien

unter der Leitung von

Univ.-Prof. Dr.-Ing. habil. Fazel Ansari

TU Wien, Fakultät für Maschinenwesen und Betriebswissenschaften, Forschungsbereich Produktions- und Instandhaltungsmanagement

Begutachtet von

Univ.-Prof. Dr. habil. Dr. h. c. Michael Henke

TU Dortmund, Fakultät für Maschinenbau, Lehrstuhl für Unternehmenslogistik

und

Prof. John Ahmet Erkoyuncu

Cranfield University, Centre for Digital Engineering and Manufacturing

von

Ing. Dipl.-Ing. Linus Kohl, BSc.

Matrikelnummer: 1028804

Wien, März 2025

Linus Koh

TECHNISCHE UNIVERSITÄT WIEN

Ich habe zur Kenntnis genommen, dass ich zur Drucklegung meiner Arbeit unter der Bezeichnung

Dissertation

nur mit Bewilligung der Prüfungskommission berechtigt bin.

Ich erkläre weiters Eides statt, dass ich meine Dissertation nach den anerkannten Grundsätzen für wissenschaftliche Abhandlungen selbstständig ausgeführt habe und alle verwendeten Hilfsmittel, insbesondere die zugrunde gelegte Literatur, genannt habe.

Weiters erkläre ich, dass ich dieses Dissertationsthema bisher weder im In- noch Ausland (einer Beurteilerin/einem Beurteiler zur Begutachtung) in irgendeiner Form als Prüfungsarbeit vorgelegt habe und dass diese Arbeit mit der vom Begutachter beurteilten Arbeit übereinstimmt.

Wien, März 2025

Linus Kohl

Acknowledgments

I would like to express my sincerest gratitude to my supervisor, Prof. Dr.-Ing. Fazel Ansari, for his unwavering guidance, insightful feedback, and constant encouragement throughout this journey. His expertise and mentorship have been invaluable in shaping this work.

I am also deeply grateful to my family for their endless support, love, and belief in me, allowing me to pursue this path. Further, I owe a debt of gratitude to my partner, Hanna, for her patience, understanding, and encouragement, which have been instrumental in fortifying my resolve.

Moreover, I would like to extend my thanks to my colleagues at Fraunhofer Austria and TU Wien, whose collaborative spirit, stimulating discourse, and camaraderie have significantly enriched this research endeavor. Your contributions have made this journey both rewarding and memorable.

I extend my deepest gratitude to everyone who was a part of this journey.

List of Abbreviations

Abbreviation	Full Name
AI	Artificial Intelligence
APC	Advanced Process Control
BERT	Bidirectional Encoder Representations from Transformers
BFS	Breadth-First Search
BoW	Bag-of-Words
CAS	Cognitive Assistance System
CM	Condition Monitoring
CPPS	Cyber–Physical Production Systems
CPS	Cyber–Physical Systems
DAS	Digital Assistance System
DFS	Depth-First Search
DL	Deep Learning
DNN	Deep Neural Network
DSB	Digital Shift Book
DT	Digital Twin
ERP	Enterprise Resource Planning
FD	Fault Documentation Ontology
FU	Functional Unit
GA	Genetic Algorithm
GPT	Generative Pre-trained Transformer
HI	Human Interface
HR	Human Resources
I40CO	Industry 4.0 Context Ontology
IAI	Industrial Artificial Intelligence
INFOQUAL	Information Quality
INTERQUAL	Interface Quality
IoT	Internet of Things
IT	Information Technology
KG	Knowledge Graph
KPI	Key Performance Indicator
MES	Manufacturing Execution System

Abbreviation	Full Name
ML	Machine Learning
MSE	Mean Squared Error
MTBF	Mean Time Between Faults
MTTR	Mean Time To Repair
NLP	Natural Language Processing
OEE	Overall Equipment Efficiency
OEM	Original Equipment Manufacturers
PdM	Predictive Maintenance
PHM	Prognostics and Systems Health Management
PM	Preventive Maintenance
PSSUQ	Post-Study System Usability Questionnaire
RAMS	Reliability, Availability, Maintainability, and Safety
RUL	Remaining Useful Lifetime
SLA	Statistical Learning Algorithm
SMCPS	Semiconductor Production System Ontology
SYUSE	System Usefulness
TF-IDF	Term Frequency-Inverse Document Frequency
TM	Text Mining:
TMP	Text Mining Pipeline
UML	Unified Modeling Language
WOW	World of Work Ontology

Abstract

The industrial sector faces several challenges, including the increasing complexity of manufacturing systems, a shortage of skilled labor, and the underutilization of unstructured data. These factors contribute to maintenance inefficiencies, resulting in high costs due to unplanned downtime, inefficient fault resolution, and ineffective documentation practices. Despite the advancements in predictive maintenance (PdM) using structured data, industries often neglect the significant potential of unstructured data, such as technician logs and shift reports. In this context, the integration of artificial intelligence (AI) technologies, such as machine learning (ML), natural language processing (NLP), and knowledge graphs (KG), has the potential to transform the maintenance landscape, address the aforementioned challenges, increasing asset availability, and enhance decision-making processes.

The prevailing PdM solutions are predominantly structured data–dependent, overlooking the vast repository of unstructured knowledge present in textual sources. Moreover, maintenance processes are characterized by a paucity of competence-based planning, inadequate support for technicians in fault diagnosis, and inefficient documentation methods. Existing tools address some PdM aspects but lack comprehensive cognitive support integrating structured and unstructured data.

To address these gaps, this thesis presents an architecture for a cognitive maintenance framework (ARCHIE) that enhances competence-based maintenance planning and execution, and augments human capabilities. ARCHIE integrates real-time sensor data with contextual insights from unstructured textual sources, thereby creating a holistic maintenance environment. Its modular architecture encompasses five functional units, supporting fault detection, task recommendation, semi-automatic documentation, and human-centered planning. ARCHIE was applied and validated within semiconductor manufacturing environments, demonstrating significant improvements. Its key outcomes included a 20% reduction in mean time to repair (MTTR) and enhanced task allocation. By providing precise, competence-aligned recommendations and leveraging virtual sensors to process unstructured data, ARCHIE enabled proactive maintenance strategies and improved overall equipment efficiency. Moreover, the framework emphasized usability, promoting user acceptance through its human-centered design. ARCHIE represents a robust contribution to the industrial maintenance field, addressing critical workforce and technological challenges. Its scalability and adaptability position it as a valuable tool for broader industrial applications, paving the way for more intelligent, data-driven maintenance solutions. The findings highlight the potential for AI to revolutionize maintenance practices, reduce costs, and support sustainable production systems.

Kurzfassung

Gegenwärtig sieht die Industrie mit verschiedenen Herausforderungen konfrontiert, darunter die Zunahme an Komplexität von Fertigungssystemen, der Mangel an Fachkräften sowie die unzureichende Nutzung unstrukturierter Daten. Diese Faktoren tragen zu Ineffizienzen bei der Instandhaltung bei, was zu hohen Kosten aufgrund ungeplanter Ausfallzeiten, ineffizienter Fehlerbehebung und ineffektiver Dokumentationspraktiken führt. Trotz der Fortschritte bei der vorausschauenden Instandhaltung unter Verwendung strukturierter Daten wird das erhebliche Potenzial unstrukturierter Daten, wie z. B. Instandhaltungsbereichte, oft vernachlässigt. In diesem Zusammenhang wird die Integration von künstlicher Intelligenz, wie maschinelles Lernen, natürlicher Sprachverarbeitung und Wissensgraphen, als ein vielversprechender Ansatz angesehen, um die Instandhaltungslandschaft zu transformieren.

Die vorherrschenden Instandhaltungs-Lösungen nutzen überwiegend strukturierte Daten und vernachlässigen unstrukturierte Daten. Darüber hinaus sind Instandhaltungsprozesse durch einen Mangel an kompetenzbasierter Planung, unzureichende Unterstützung für Techniker bei der Fehlerdiagnose und ineffiziente Dokumentationsmethoden gekennzeichnet. Obwohl bereits einige Aspekte der vorausschauenden Instandhaltung durch bestehende Tools abgedeckt werden, bieten diese keine umfassende kognitive Unterstützung.

Die vorliegende Arbeit präsentiert eine Architektur für ein kognitives Instandhaltungs-Framework (ARCHIE), dass die kompetenzbasierte Instandhaltungsplanung und -ausführung verbessert und die menschlichen Fähigkeiten erweitert. Die Architektur integriert Echtzeit-Sensordaten mit kontextbezogenen Erkenntnissen aus unstrukturierten Textdaten. Die modulare Architektur umfasst fünf Funktionseinheiten, die Fehlererkennung, Aufgabenempfehlungen, halbautomatische Dokumentation und menschenzentrierte Planung unterstützen.

Die Validierung von ARCHIE erfolgte in Halbleiterindustrie, wo signifikante Verbesserungen beobachtet wurden. Zu den wichtigsten Ergebnissen gehören eine Reduzierung der mittleren Reparaturzeit (MTTR) um 20% und eine verbesserte Aufgabenzuweisung. Es lässt sich zusammenfassen, dass ARCHIE einen substanziellen Beitrag zum Bereich der industriellen Instandhaltung leistet und auf kritische personelle und technologische Herausforderungen der industriellen Zukunft eingeht. Seine Skalierbarkeit und Anpassungsfähigkeit macht es zu einem wertvollen Werkzeug für breitere industrielle Anwendungen und ebnet den Weg für intelligentere, datengetriebene Instandhaltungslösungen. Die Ergebnisse zeigen das Potenzial von Künstlicher Intelligenz in der Instandhaltung auf, Anlagen Verfügbarkeit zu erhöhen und nachhaltige Produktionssysteme zu fördern.

Table of Contents

1 Introduction

1.1 Initial Situation and Motivation

The ongoing digital transformation in manufacturing and industrial operations, commonly referred to as Industry 4.0, has prompted significant investments in predictive maintenance (PdM) solutions and cognitive technologies (Nassehi et al., 2022). The global market for PdM is projected to reach $14.3 billion by 2028, with an annual growth rate of 17% (Brügge et al., 2023). This surge in investment reflects the increasing need for advanced maintenance strategies to ensure the reliability and efficiency of capital-intensive assets, where the average unplanned downtime costs exceed $100,000 per hour (Brügge et al., 2021). These economic demands highlight the importance of optimizing industrial operations, particularly in reducing unplanned downtime, improving asset utilization, and lowering maintenance costs (Tolio et al., 2023). Traditional reactive or time-based maintenance approaches are no longer feasible to meet the performance demands of modern production systems (Klees & Wortmann, 2023). Given this, PdM offers an attractive alternative by predicting faults before they occur, thus allowing for timely interventions that minimize production disruptions.

According to a study conducted by McKinsey, enterprises generally utilize only 37% of their data (Shaikh et al., 2024). Moreover, the National Association of Manufacturers (NAM) revealed that while only 25% of manufacturers have high confidence in the collected data, 86% consider using their data essential to their competitiveness (NAM, 2023). Furthermore, McKinsey estimates that artificial intelligence (AI)-enabled PdM can reduce machine downtime by up to 50%, generate 12% savings on repair costs, and lower overall maintenance expenses by 10–40% (Toyoglu et al., 2023). This demonstrates the untapped potential of current data usage strategies, particularly when incorporating AI technologies that can also process unstructured data. Similarly, the Organisation for Economic Co-operation and Development (OECD) reported that AI adoption could increase annual global productivity by 1.2% by 2030, stressing the growing economic imperative to integrate AI-driven maintenance strategies (OECD, 2024).

This shift is significant given the challenges pervading the labor market. For instance, OECD data predicts that over 20% of the global industrial workforce will retire by 2030, exacerbating the skill gap in maintenance and operations (OECD, 2023). This combination of a shrinking workforce and increasingly complex production systems demands automated solutions to maintain operational efficiency. Furthermore, the need for skilled technicians, especially those familiar with data analytics and AI, has become critical in an era where companies invest in cognitive automation to fill this gap and enhance structured as well as unstructured maintenance tasks.

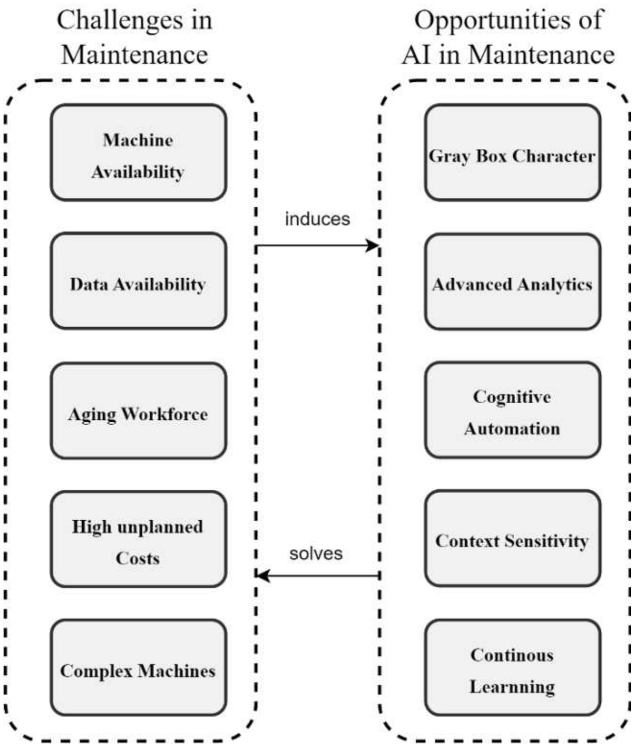
Despite significant progress, several industries have yet to fully capitalize on the potential of AI, especially in the domain of industrial maintenance. For instance, PdM primarily relies on structured machine data, such as sensor readings and equipment logs (McKinsey & Company, 2022). However, this approach neglects unstructured data from textual sources, including technician reports, shift logs, and maintenance manuals. According to Gartner, 80–95% of data in industrial settings is unstructured, with a significant portion remaining untapped (Gartner, 2023).

AI technologies, particularly machine learning (ML), natural language processing (NLP), or, in this context, technical language processing (TLP), and knowledge graphs (KG), are central to the evolution of maintenance management as, together, they can be utilized to design cognitive systems that offer a transformative approach to maintenance planning and operations. McKinsey's estimates on AI's impact—i.e., cutting machine downtime by half and drastically reducing maintenance costs—are driven by its ability to process large volumes of structured sensor data in combination with unstructured textual data representing human knowledge (McKinsey & Company, 2022). The application of ML models to this dataset enables the prediction of equipment faults with greater accuracy and the provision of real-time insights into machine health.

Furthermore, TLP enables the extraction of valuable insights from textual sources, such as technician reports and maintenance logs, which, although often rich in experiential knowledge, remain underutilized. Additionally, KG augments these systems by creating a semantic framework that links different data types, allowing maintenance personnel to contextualize and understand equipment behavior holistically. This enriched data environment supports more effective decision-making and planning.

Cognitive systems for industrial maintenance, combining AI with semantic capabilities to augment human skills and competences, represent the next frontier for improving efficiency, particularly in cognitive tasks that involve complex decision-making processes. By integrating ML, TLP, and KG, these systems can transform vast amounts of both structured and unstructured data into actionable insights. Whereas ML algorithms allow predictive models to continuously learn from historical data and technician input, TLP extracts and processes unstructured textual data from human knowledge, such as maintenance logs or shift notes. Meanwhile, KG provides a structured, semantic representation of these data points, giving cognitive systems reasoning capability that links them to broader maintenance processes and operational contexts.

Figure 1: *Challenges in Maintenance and Opportunities of AI in Maintenance*



This AI-driven approach not only improves machine availability but also reduces mean time to repair (MTTR) by enhancing fault detection accuracy and overall machine availability after predicting and preventing faults more reliably. By automating cognitive tasks, cognitive AI systems compensate for the growing skills shortage in industrial maintenance, enabling maintenance personnel to focus on higher-level decision-making while ensuring that complex systems run smoothly. Thus, AI becomes an essential partner in managing the increasingly sophisticated and data-driven maintenance environments of contemporary industries. According to Verein Deutscher Ingenieure e.V.(VDI) (VDI, 2014), a shift book, more commonly known as a logbook, records machine faults, circumstances, and suggestions for operation and maintenance personnels' activities. The texts entered by maintenance professionals frequently lack the quality necessary to fully leverage the specialist information contained in such books. Consequently, preprocessing is rendered complicated due to the use of various terms, technical language, acronyms, and unfinished texts. Therefore, amplification of current techniques and methodologies is required to address these idiosyncrasies so as to collect all of the information in these texts, transform it into a structured representation through vectorization or quantification of text elements, and then, construct the necessary key performance indicators (KPIs).

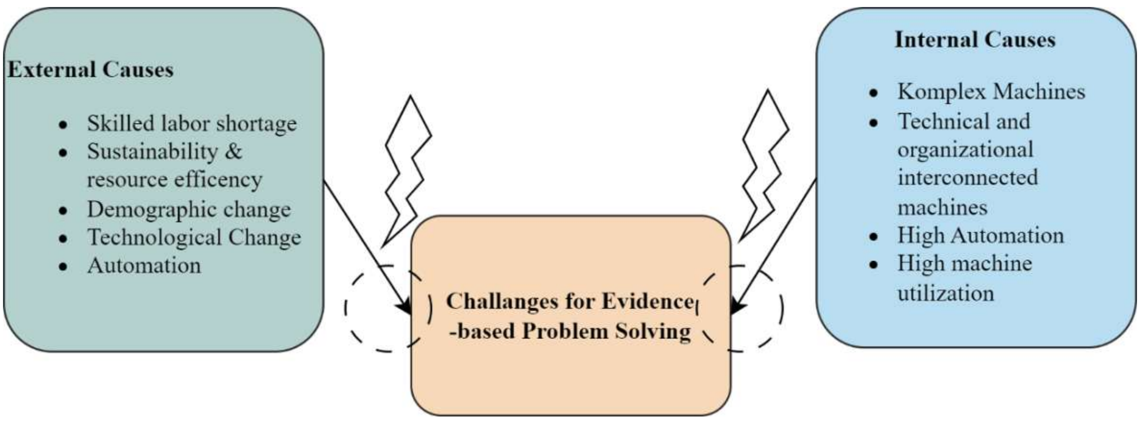
The primary objective behind integrating AI into maintenance management is to improve key challenges associated with industrial maintenance, as illustrated in Figure 1. These challenges are primarily the optimization of KPIs related to reliability, availability, maintainability, and safety (RAMS) (Fink et al., 2021; Passath et al., 2021) Specifically, AI-driven solutions aim to reduce MTTR by enabling more accurate fault detection and diagnostics, facilitating quicker repairs, and minimizing downtime. Moreover, by optimizing maintenance processes, companies can enhance overall equipment effectiveness (OEE), increase uptime, and lower lifecycle costs.

Furthermore, AI systems that incorporate human knowledge through NLP and KG to not only improve prediction accuracy but also offer maintenance personnel enriched information for decision-making. This leads to more efficient planning, faster problem resolution, and a safer operational environment. By combining machine data with human expertise, AI-enhanced maintenance strategies enable companies to maximize asset performance and achieve significant cost savings.

1.2 Problem Definition

Recent market studies reveal considerable attention and increasing investment rates for AI-enhanced maintenance as one of the major use cases of AI in Industry 4.0 (Brügge et al., 2021). Maintenance costs represent between 15 and 70% of the total operating costs of manufacturing companies, depending on asset intensity (Thomas & Weiss, 2021). Approximately 30% of maintenance costs are owing to unnecessary or incorrectly performed maintenance activities (Wireman, 2014).

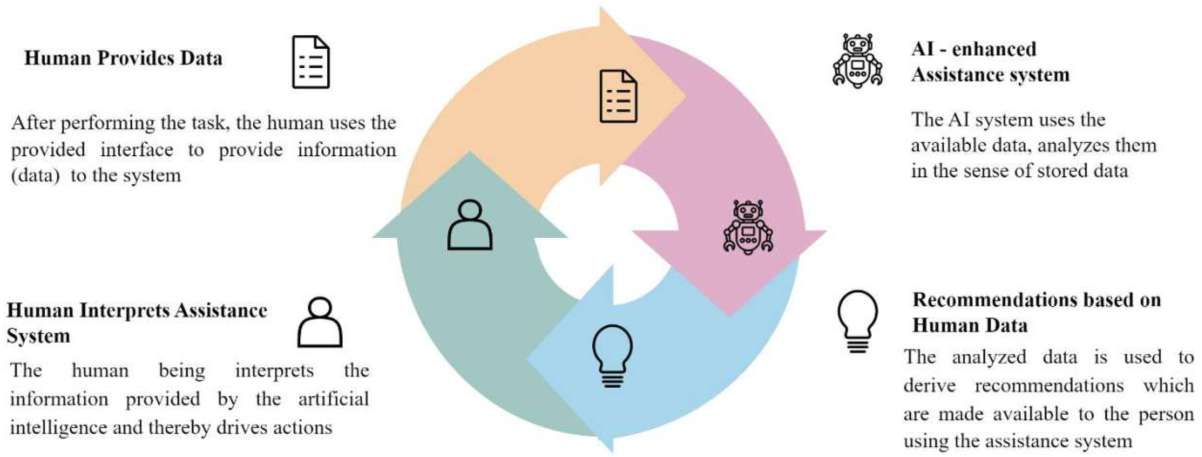
Figure 2: *Challenges for Troubleshooting in Maintenance, adopted from (Matyas, 2022)*



Simultaneously, maintenance technicians and their processes face various challenges, especially when it comes to solving technical problems (Henke, 2019). For instance, utilizing complex systems contributes to machine faults that are challenging to anticipate and render the task of formulating advanced maintenance plans daunting. Concurrently, the process of troubleshooting has become considerably more intricate. This is because modern systems are highly interconnected, operate at high capacity utilization levels, and are characterized by high levels of automation. Consequently, any issues that arise in one system can have a knock-on effect on other connected systems. Furthermore, the global economy is currently facing a shortage of skilled labor, a shift in the demographic profile of the workforce, and intense competition. This is forcing companies to increase productivity to maintain their competitive advantage. As a result, maintenance activities are frequently carried out under significant time and performance pressure, which can lead to the implementation of incorrect maintenance activities.

This lays the foundation for the initial problem statement (P1) of the thesis, which states that the shortage of skilled labor leads to incorrectly planned or performed maintenance activities due to the lack of competence among the available maintenance technicians (Schlund & Ansari, 2021). While these maintenance technicians can be aided by assistance systems, such systems must be capable of assessing their individual competence levels to provide recommendations for problem-solving or planning and scheduling that are tailored to their needs. This is illustrated in the human-in-the-loop model (Bhattacharya et al., 2023) as shown in Figure 2 (Schlund & Ansari, 2021).

Figure 3: *The Human-in-the-Loop Concept in Industrial Maintenance*



The rise in complexity of industrial AI solutions has a significant impact on maintenance staff, as maintenance tasks are becoming increasingly difficult for them to complete as they demand the correct information at the appropriate time. The time spent diagnosing a machine fault and determining a potential solution is generally greater than the time spent resolving it (Sexton et al., 2017). Besides, technicians are often not supported with helpful information during fault diagnosis and subsequent troubleshooting. Often, knowledge management only takes place in the form of lessons learned through forms or records in an information technology (IT) system, such as Maintenance Planning Control and Analysis System (IPSA) (Huber et al., 2021). There are various limits to consider while providing maintenance support in complex manufacturing operations. To begin with, current systems prioritize static data as a central issue. Dynamic information, such as real-time data from sensors or text in machine downtime event descriptions or maintenance reports, cannot be shown since it is unrelated to the machine and maintenance activities, contributing to the second problem statement (P2) of this thesis: the lack of timely and relevant information for maintenance technicians and planners (Flatt et al., 2015). It can be observed that maintenance technicians who use assistance systems for fault identification demonstrate significant improvements in both specificity and time efficiency (24.6% and 25.3%, respectively) over the control group, which does not receive support in their activities (Shin et al., 2021). In these studies, maintenance technicians did not have support for solving existing machine faults and underlying issues.

Owing to the findings of Gandomi and Haider (2015),(2015) it is important to incorporate unstructured data (e.g., text) to go beyond the conventional approach, which states that data in production only come from machines via mounted physical sensors. The diversity of data structures, including multimodal data sources, enables the development of novel virtual sensors, offering an enhanced data foundation for augmenting human capabilities through the analysis of expert knowledge embedded in textual content. A prominent application is in maintenance, where preventive and troubleshooting processes predominantly generate written reports on machine faults, and maintenance activities. These documents, created by maintenance technicians, encapsulate experiential knowledge regarding conditions, faults, causes, and solutions, often recorded in unstructured formats (Kohl et al. 2021). The use of TLP realizes the untapped value of the existing unstructured or semi-structured textual data, where TLP is defined as applying AI algorithms and methods to text to find valuable patterns (Gao et al., 2020; Hotho et al., 2005). Amid all this, there are few applications for retrieving technical assistance reports (Ansari, 2020), since their effectiveness may be restricted by data inconsistencies and inaccuracies caused by the operators' informal language (Biegel et al., 2022), the inability to detect suboptimal solutions by technicians with lower competencies, and the highly unstructured nature of maintenance reports (Ansari, 2020), leading to the third problem statement of the thesis (P3). Additionally, documenting maintenance activities is a time-consuming and error-prone process, which renders AI-based evaluation difficult.

The following section presents the problem statements of study. It should be noted that P1 focuses on the individual provision of information by maintenance planners and technicians, while P2 deals with the depth of this information. On the other hand, P3 addresses the use of this information by maintenance planners and technicians in their daily processes.

The study, therefore, examines the fundamental problem of individual information provision in maintenance under the boundary conditions of demographic change, shortage of skilled workers, heterogeneous data in production, and how human skills can be augmented with current AI technologies.

Research Problems:

- P1: The lack of competence-based scheduling in industrial maintenance leads to non-ideal shift compositions, inefficient task allocation, and sub-optimal problem-solving, resulting in operational inefficiencies, reduced quality, and cognitive overload for maintenance technicians (Ansari et al., 2021; Schlund & Ansari, 2021).
- P2: The absence of adequate support for maintenance technicians and planners in identifying and understanding specific faults and underlying issues in machinery significantly hampers effective maintenance management, leading to a lack of information, prolonged downtime, and inefficient maintenance processes (Flatt et al., 2015; Huber et al., 2021; Shin et al., 2021).
- P3: The current process of documenting maintenance activities is both time-consuming and prone to errors, posing significant challenges for AI-based evaluation and analysis. This inefficiency in documentation hinders the accuracy and reliability of maintenance logs and impedes the effective implementation of AI tools in predictive maintenance and decision-making processes (Biegel et al., 2022; Mark et al., 2021).

1.3 Research Object, Questions, and Objectives

This thesis aimed to lay the foundations for a contemporary maintenance philosophy that focuses primarily on improving asset availability, with secondary goals of reducing MTTR and improving quality. This approach was necessitated by the demographic changes taking place across Europe, the impending shortage of skilled workers, and the simultaneous international competitive pressures, along with the rapid developments in AI. The objective is to provide a framework to support and augment maintenance planners and technicians in their tactical and operational tasks.

Central to this philosophy is facilitating the information delivery tailored to individual skill levels, specific tasks, and contexts. By emphasizing asset availability, the maintenance strategy ensures that equipment remains operational and productive for the maximum duration, directly impacting organizational efficiency and profitability. Reducing MTTR and improving quality further minimize downtime and ensure that maintenance activities meet high standards.

From a human-centric perspective, this approach aims to reduce cognitive load, enabling technicians to focus on problem-solving and rapid repair tasks. Organizationally, it opens up the potential to use personnel with less extensive qualifications in maintenance planning and execution, thereby addressing the skills shortage while maintaining high levels of asset availability. Technologically, it requires creating a system capable of integrating disparate information sources with semantic data to improve accessibility and usability.

To achieve these goals, this thesis designed and validated an architecture for a cognitive maintenance system called ARCHIE, envisioned as a skills-based assistance system that supports operational planning, maintenance planning, shift planning, and task allocation—all aimed at maximizing asset availability. By combining data inputs from physical sensors (structured data) and virtual sensors (unstructured data), ARCHIE constructs a holistic representation of the maintenance environment. This integration allows the system to leverage both explicit and hidden knowledge to make autonomous, proactive, and opportunistic decisions to improve maintenance processes, reduce MTTR, improve quality, and, ultimately, increase asset availability.

Employing a hybrid AI approach, ARCHIE operates as a “gray box” system, which provides greater transparency in decision-making compared to rigid but explainable white box expert systems or flexible but opaque black box systems. The gray box model establishes a transparent core while pursuing flexible, less transparent methods to adapt to different maintenance scenarios, all in the service of improving asset availability and maintenance quality.

The realization of ARCHIE as a human-centered learning support system aligns with the fundamental assumption of providing information at the level of individual competence. It is designed to provide maintenance technicians with the necessary competency-based information to effectively troubleshoot and resolve problems in real time, especially in high-pressure, hands-on environments where asset availability is critical. This supports cognitive relief for human workers and enables the use of less skilled personnel without compromising asset performance or maintenance quality.

ARCHIE's modular structure, consisting of six Functional Units (FUs), enables adaptability and scalability within the maintenance environment. Each FU addresses specific aspects of the maintenance process and is seamlessly integrated to support the overarching goals of contemporary maintenance philosophy. By integrating disparate information sources and leveraging semantic data, ARCHIE embodies the technological imperative of creating an accessible and intelligent maintenance support system focused on maximizing asset availability.

In conclusion, this thesis demonstrates the benefits of competency-based maintenance support systems. By aligning with the proposed contemporary maintenance philosophy and employing a hybrid gray box AI system, ARCHIE represents a significant advancement in supporting and augmenting the maintenance workforce amid current demographic and technological challenges. Its focus on improving asset availability, along with reducing MTTR and improving quality, positions ARCHIE as a central tool in modern maintenance strategies. ARCHIE has the following sub-goals:

- **Generalizability**: The design of the architecture is intended for industrial maintenance but is kept generic enough so that transferability to different industries is easily feasible (e.g., energy industry, semiconductor industry, automotive industry, process industry).
- **Scalability**: The modular structure and the design of the generic components allow for easy adaptation from a one-machine use case to the expansion to a whole manufacturing line or factory.
- **Customizability**: The FUs can be customized individually or even interchanged if the application at hand requires a different approach, which allows for individual adaptation to the use case.
- **Reliability**: The architecture and algorithmic design are aimed at robust implementation. This is particularly important in the industrial application of AI systems since system faults entail high consequential costs.

- **User Acceptance:** When implementing AI systems in an industrial environment, user acceptance of the system is a key factor for success. Therefore, ARCHIE aims for a human-centered design focusing on intuitive and gender-neutral usability.

The aforementioned objectives guided the formulation of objectives for implementing a human-centered cognitive maintenance system in the context of this thesis.

- Objective 1: To derive a morphology on the components of a cognitive human-centered assistance system for competence-based planning and operation of maintenance activities to identify existing research gaps, addressing research problem P1
- Objective 2: To design and implement an AI-enhanced architecture for supporting maintenance technicians and planners with a cognitive assistance system that improves maintenance shift planning, machine fault identification, and documentation, addressing research problem P2
- Objective 3: To design and implement an innovative support system that enhances the capability of maintenance technicians and planners by providing real-time, detailed information about machine faults and underlying issues, addressing research problem P2

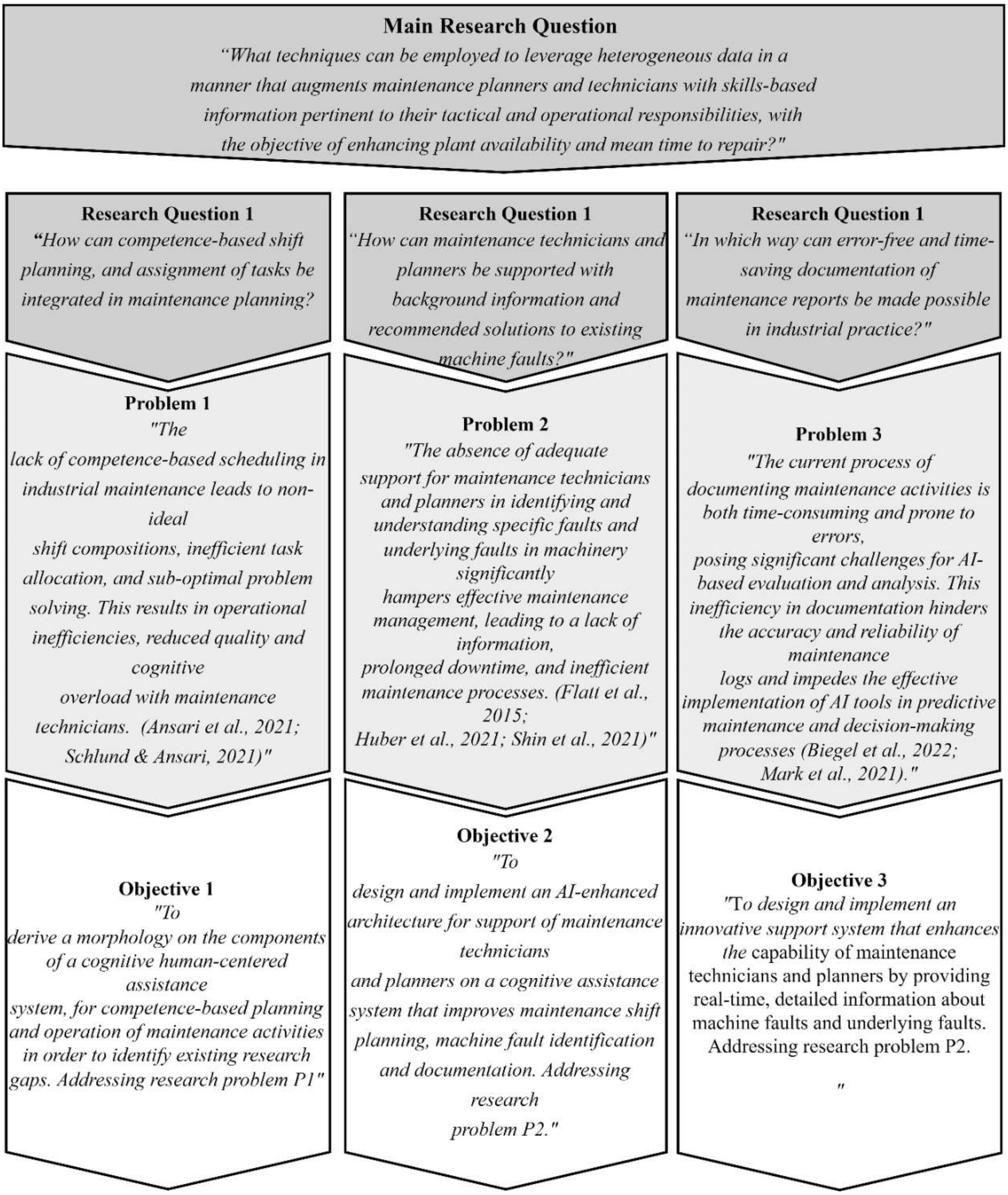
To contribute to the improvement of cognitive maintenance systems, the architecture of a cognitive maintenance system and a procedure model for its implementation were developed with the desire to answer the following research questions:

Research Question: "What techniques can be employed to leverage heterogeneous data in a manner that augments maintenance planners and technicians with skills-based information pertinent to their tactical and operational responsibilities to enhance plant availability and mean time to repair?"

- **Sub-Research Question 1:** "How can competence-based shift planning and assignment of tasks be integrated into maintenance planning?
- **Sub-Research Question 2:** "How can maintenance technicians and planners be supported with background information and recommended solutions to existing machine faults?"
- **Sub-Research Question 3:** "How can error-free and time-saving documentation of maintenance reports be achieved in industrial practice?"

Figure 4 presents the research question, along with its sub-research questions and the associated problems and objectives.

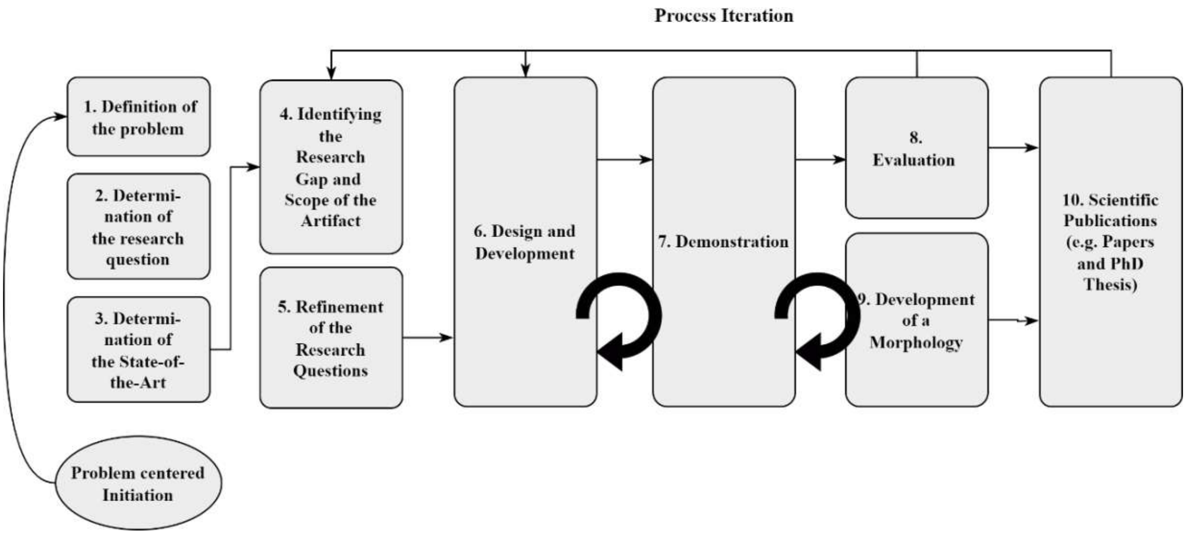
Figure 4: *Context of Research Questions, Problems, and Objectives*



1.4 Research Methodology and Design

The lack of support in shift planning, fault identification, and documentation of maintenance activities renders maintenance labor-intensive, fault-prone, and slow. Loosely based on Archer's (1984) methodology of "the purposeful seeking of a solution", the research problem in this study was broken down into three research questions. For this problem-centered approach, Hevner et al.'s (2004) design science turned out to be a goal-oriented methodology, and the design science research methodology for information system research based on Peffers et al. (2007) was used. The need for a more efficient AI-enhanced assistance system triggered the development of the ARCHIE framework for a competence-based planning assistance system, following the design science in research methodology (DSRM), as illustrated in Figure 5.

Figure 5: *Design Science in Research Methodology a Problem-Centered Approach*



Problem Identification and Research Gap

Industrial maintenance is confronted with three pivotal challenges: (P1) the absence of a competence-based task allocation, giving rise to inefficient shift planning and cognitive overload for technicians (Schlund & Ansari, 2021)(Shin et al., 2021); (P2) inadequate support in fault identification, resulting in protracted downtimes and ineffective troubleshooting (Ansari, 2020); and (P3) the time-consuming and error-prone documentation of maintenance activities (Biegel et al., 2022). While PdM has seen significant advancements, current approaches predominantly rely on structured sensor data, neglecting valuable unstructured information such as technician reports and maintenance logs. This discrepancy limits the accuracy of fault predictions and the ability to provide context-aware decision support. To address these issues, a cognitive assistance system integrating both structured and unstructured data is required to enhance

Objective of the Solution

The objectives of the solution were to (i) construct a morphology for a cognitive assistance system for maintenance and (ii) develop ARCHIE to augment maintenance planners and technicians' capabilities within the example of semiconductor manufacturing. The major challenges included the multimodality of the data sources, the diversity of the maintenance tasks, the various constraints for planning and task allocation, and the user-centric design that supports the maintenance workflow.

Design and Development

The developed artifact enables competence-based shift planning and supports maintenance planners and technicians in fault identification and documentation of maintenance actions. This is implemented through ARCHIE, defined in Chapter 4, which consists of six FUs. Agile development through close user involvement enables user-centered design and value-added functionality.

Demonstrator

After implementing ARCHIE in proof-of-concept demonstrators in two different use cases, these were implemented in an industrial laboratory environment and during industrial use on the shop floor. It was demonstrated that components of the ARCHIE architecture can be used to determine the additive manufacturing capability of spare parts in the railway sector.

Evaluation

ARCHIE's components have been introduced in two maintenance use cases: semiconductor manufacturing and the automotive industry, as well as partially in the area of additive manufacturing of railroad spare parts. In the semiconductor industry, a three-stage evaluation of the system according to the post-study system usability questionnaire (PSSUQ) was carried out on the users, while a survey on AI-based assistance systems was conducted among decision-makers in the industry.

Communication

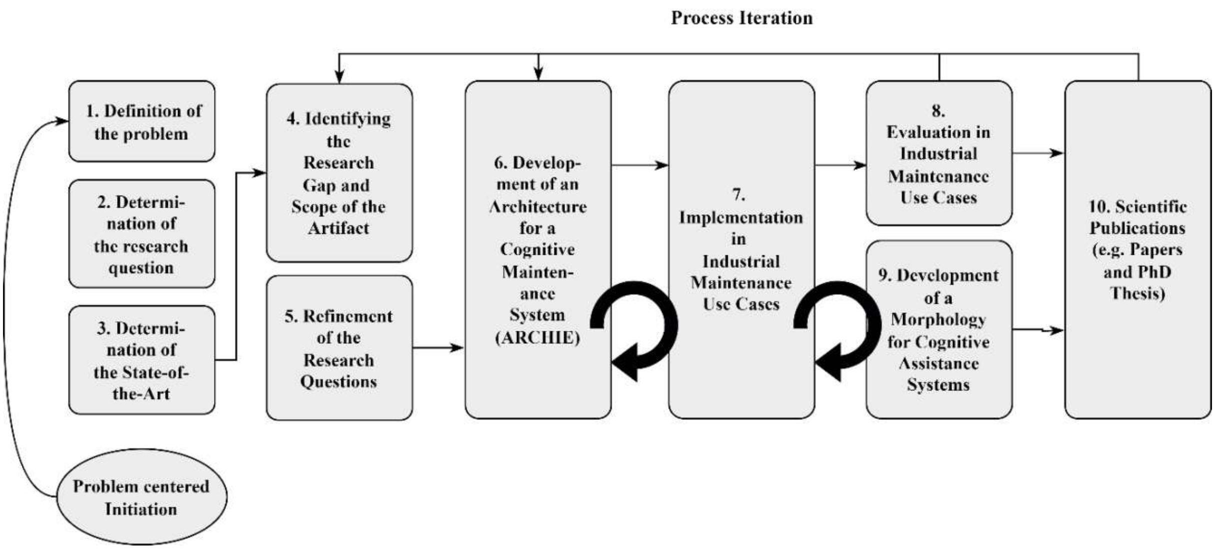
Papers on TM for AI-enhanced fault detection and AI for skills management were published. The developed text mining approach was in described in "*Text Mining for AI-enhanced Fault Detection and Availability Optimization in Production Systems*" (Ansari et al., 2021) while the approach for using AI was provided in "*Artificial Intelligence in Competence Management: A Case Study from the Semiconductor Industry"* (Kohl, Fuchs, Valtiner, et al., 2021). Furthermore, publications on statistical learning algorithms (SLAs) and KG are *"A Knowledge Graph-based Learning Assistance System for Industrial Maintenance"* (Kohl, Anari 2023) and "AI-Enhanced Fault Detection Using Multi-Structured Data in Semiconductor Manufacturing" (Kohl et al., 2024).

Contribution

ARCHIE resulted in a framework. This artifact was applied in both laboratory and industrial environments, and its use was evaluated. The immediate contribution of ARCHIE is the development of a maintenance-specific KG and a SLA that can be used together with Lagrange multipliers for competence-based planning. In a broader scope, the contribution of ARCHIE to the research landscape is the development of a generic architecture for the use of physical and virtual sensors for the competence-based planning of maintenance activities.

Accordingly, the research design of the DSRM starts with a problem-centered approach resulting in contributions to both application and fundamental science. The iterative procedure then enables mutual coordination until the requirements of the application area have finally been met and the knowledge base has been supplemented. Figure 5 illustrates the application of the DSRM in the context of this thesis, based on the outlined methodology in Figure 5.

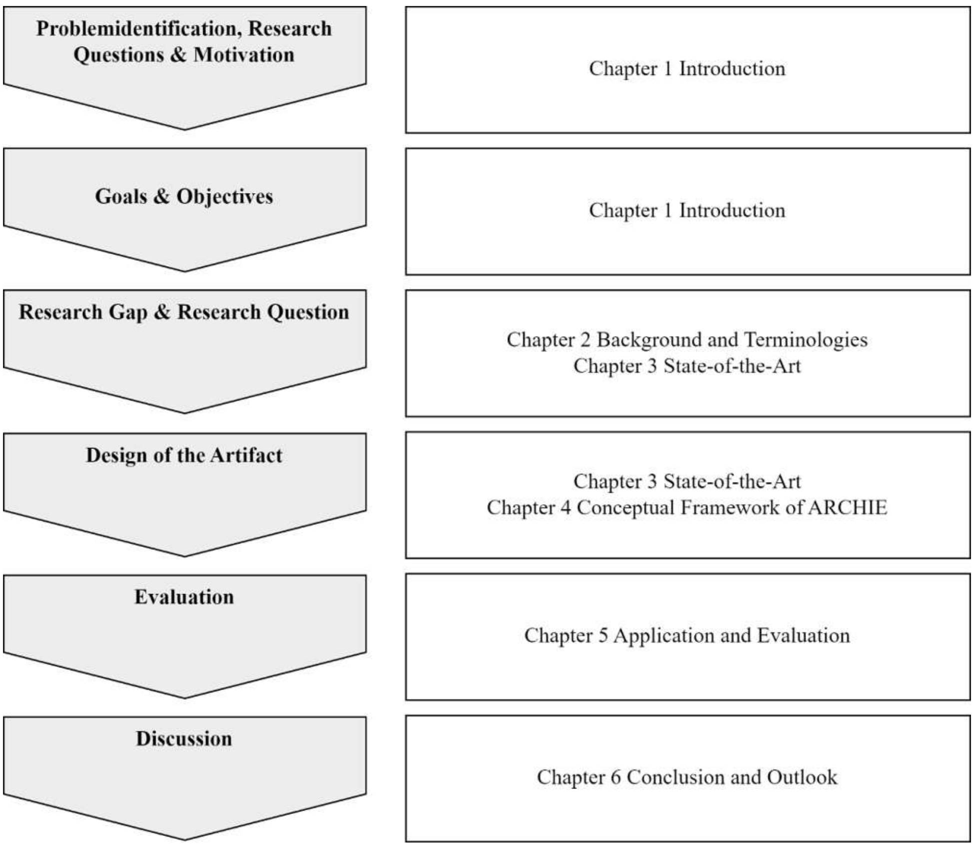
Figure 6: *Application of Design Science Research Methodology in the Context of the Thesis, Based on Peffers et al. (2007)*



1.5 Structure of the Thesis

The thesis is structured as a comprehensive examination of an AI-enhanced framework for competence-based maintenance planning, as illustrated in Figure 7. Chapter 1 introduces the study's context and outlines the research motivation. Chapter 2 defines the objectives and provides a structural overview. Chapter 3 elaborates on the scientific bases, covering necessary topics that are discussed in further detail in the following chapters. Chapter 4 provides an analysis of current technologies and methodologies related to AI in maintenance, establishing the scientific foundation for the framework. Chapter 5 provides a detailed account of the creation and architectural design of the ARCHIE framework, which incorporates specific AI technologies to enhance maintenance planning. In Chapter 6, the framework is applied within the context of semiconductor manufacturing and is subjected to a comprehensive examination. Moreover, this chapter discusses the framework's implementation and operational integration. Chapter 7 presents a discussion of the framework's effectiveness through a use-case scenario, analyzing the results and their implications for maintenance efficiency. Chapter 8 summarizes the research findings, discusses the impacts and limitations, and outlines future research pathways. In particular, the chapter considers potential expansions of the framework's application and further AI integrations.

Figure 7: *Structure of the Thesis*



1.6 Published and Supervised Works

This section details the author's published works and supervised student theses during his PhD studies. These publications influenced and informed parts of this thesis. The list focuses solely on academic articles directly related to the thesis research, where the author made significant contributions.

I. **Kohl, Linus**; Eschenbacher, Sarah; Besinger, Philipp; & Ansari, Fazel. (2024). Large Language Model-Based Chatbot for improving human-centricity in Maintenance Planning and Operations. In *Proceedings of the European Conference of the PHM Society 2024* (pp. 12–12).

II. **Kohl, Linus,** Madreiter, Theresa, & Ansari, Fazel. (2024). AI-Enhanced Fault Detection Using Multi-Structured Data in Semiconductor Manufacturing. In *Multimodal and Tensor Data Analytics for Industrial Systems Improvement* (pp. 297-312). Cham: Springer International Publishing.

III. **Kohl, Linus**; Stricker, Philipp; Reisinger, Julia; Ansari, Fazel (2024): digiTeachVR: A Digitally Enhanced Teaching Platform for Improving Data Science Skills and Virtual Reality Competencies in Cross-Disciplinary Engineering Education. In: *Proceedings of the 14th Conference on Learning Factories (CLF 2024)*

IV. Ansari, Fazel; **Kohl, Linus**; Sihn, Wilfried (2023): A Competence-Based Planning Methodology for Optimizing Human Resource Allocation in Industrial Maintenance. In: *CIRP Annals* 72 (1), pp. 389–392. DOI: 10.1016/j.cirp.2023.04.050.

V. **Kohl, Linus**; Ansari, Fazel (2023): A Knowledge Graph-based Learning Assistance System for Industrial Maintenance. In: *Procedia CIRP, 126, 87-92.*

VI. Biegel, Tobias; Jourdan, Nicolas; Madreiter, Theresa; **Kohl, Linus**; Fahle, Simon; Ansari, Fazel et al. (2022): Combining Process Monitoring with Text Mining for Anomaly Detection in Discrete Manufacturing. In: *Proceedings of the 12th Conference on Learning Factories (CLF 2022)*

VII. Ansari, Fazel; **Kohl, Linus** (2022): AI-Enhanced Maintenance for Building Resilience and Viability in Supply Chains. In: Alexandre Dolgui, Dmitry Ivanov, and Boris Sokolov (Eds.): Supply Network Dynamics and Control, Vol. 20. Cham: Springer International Publishing (Springer Series in Supply Chain Management), pp. 163–185.

VIII. Tuli, Tadele Belay; **Kohl, Linus**; Chala, Sisay Adugna; Manns, Martin; Ansari, Fazel (2021): Knowledge-Based Digital Twin for Predicting Interactions in Human-Robot Collaboration. In: 2021 26th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA). 2021 IEEE 26th International Conference on

Emerging Technologies and Factory Automation (ETFA). Vasteras, Sweden, 07.09.2021 - 10.09.2021: IEEE, pp. 1–8.

IX. Madreiter, Theresa; **Kohl, Linus**; Ansari, Fazel (2021): A Text Understandability Approach for improving reliability-centered maintenance in Manufacturing Enterprises. In: Alexandre Dolgui, Alain Bernard, David Lemoine, Gregor von Cieminski, and David Romero (Eds.): Advances in Production Management Systems. Artificial Intelligence for Sustainable and Resilient Production Systems, Vol. 630. Cham: Springer International Publishing (IFIP Advances in Information and Communication Technology), pp. 161–170.

X. Passath, Theresa; Huber, Cornelia; **Kohl, Linus**; Biedermann, Hubert; Ansari, Fazel (2021): A Knowledge-Based Digital Lifecycle-Oriented Asset Optimization. In: *Teh. glas. (Online)* 15 (2), pp. 226–334. DOI: 10.31803/tg-20210504111400.

XI. Ansari, Fazel; **Kohl, Linus**; Giner, Jakob; Meier, Horst (2021): Text Mining for AI enhanced fault detection and Availability Optimization in Production Systems. In: *CIRP Annals*. DOI: 10.1016/j.cirp.2021.04.045.

XII. **Kohl, Linus**; Ansari, Fazel; Sihn, Wilfried (2021): A Modular Federated Learning Architecture for Integration of AI-Enhanced Assistance in Industrial Maintenance - A Novel Architecture for Enhancing Industrial Maintenance Management Systems in the Automotive and semiconductor industry. In: Wilfried Sihn und Sebastian Schlund (Hg.): Competence Development and learning assistance systems for the Data-Driven Future: Goto Verlag, S. 229–242.

XIII. **Kohl, Linus**; Fuchs, Benedikt; Berndt, Rene; Valtiner, Daniel; Ansari, Fazel; Schlund, Sebastian (2021): Künstliche Intelligenz im Kompetenzmanagement. In: *Zeitschrift für wirtschaftlichen Fabrikbetrieb* 116 (7-8), S. 534–537. DOI: 10.1515/zwf-2021-0100.

Additionally, the following student theses were co-supervised, which have impacted the present work:

Master's Thesis

I. Künig, L. (2024). Transformer Event Extraction Explainer: A Tool for Improving the Explainability of Transformer Models for Industrial Maintenance Applications, supervised by **Kohl, L**; Ansari, F, TU Wien

II. Neubauer, P. (2024). Evolution of Prescriptive Maintenance in the Pharmaceutical Industry: Design and Evaluation of a Novel Data-Driven Maintenance Strategy for the Example of a Chromatography Machine, supervised by **Kohl, L**; Ansari, F, TU Wien

III. Kölbl, J. (2024). Large Language Model-Based Chatbots in Industrial Maintenance Applications, supervised by **Kohl, L**; Ansari, F, TU Wien

Bachelor's Thesis

I. Pfanner, L., (2021). Text-Mining in Maintenance: A Comparison Between Different Vectorization Models for Maintenance Documents Enabling Better Information Retrieval in Predictive Maintenance, supervised by **Kohl, L**; Ansari, F, TU Wien

II. Özylmaz, B. (2023). The Use of Digital Twin Morphology Towards More Sustainable Maintenance, supervised by **Kohl, L**; Ansari, F, TU Wien

III. Koch, A. (2023). An Overview of the Application of Assistant Systems in Maintenance Strategies, supervised by **Kohl, L**; Ansari, F, TU Wien

2 Background and Terminologies

This chapter examines the fundamental concepts of PdM, industrial AI (IAI), cognitive systems, and assistance systems in the industrial environment, along with the underlying principles most relevant to this thesis, as illustrated in Figure 8.

To refine the research objectives, an understanding of the research topic is established for the following discussions. To lay a scientifically rigorous foundation for the thesis, the fundamental principles and specifics concerning the background of the developed framework are provided, based on which subsequent chapters are build.

Defining the research topic begins with establishing the scope of analysis. Before determining the process of extracting competencies from text, an introduction to the fundamentals of maintenance theory, IAI, and assistance systems is presented. These domains intersect in significant ways that are crucial for understanding the context and objectives of this research.

The overlapping area between assistance systems and IAI lies in their collective aim to augment human capabilities within industrial environments through intelligent technologies. Assistance systems leverage IAI to provide planners and technicians with adaptive support, real-time feedback, and insights. By incorporating ML algorithms and data analytics, these systems can offer personalized guidance, automate routine tasks, and facilitate decision-making processes, thereby improving efficiency and reducing the likelihood of human error.

The intersection of maintenance and IAI is evident in the application of AI techniques to optimize maintenance activities. IAI enables predictive analytics by analyzing large volumes of operational data to identify patterns, anomalies, and potential faults before they occur, allowing for condition- or knowledge-based maintenance strategies and improving asset availability. AI-driven tools can also assist in scheduling maintenance tasks, allocating resources efficiently, and providing diagnostics that enhance the effectiveness of maintenance operations.

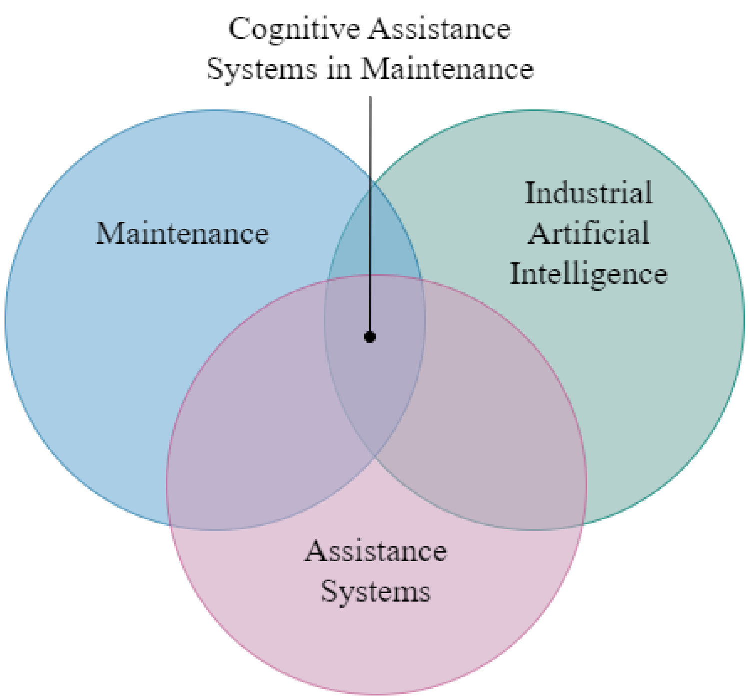
The convergence of maintenance and assistance systems focuses on supporting maintenance personnel through technology that augments their performance and efficiency. Assistance systems in maintenance may include augmented reality interfaces that overlay instructions onto equipment, interactive manuals, and automated troubleshooting guides.

All three domains—namely, maintenance, assistance systems, and IAI—meet during the development of cognitive assistance systems in maintenance that synergistically combine human expertise with AI technologies. This intersection is the focal point of the research, aiming to create a competence-based assistance system powered by IAI to comprehensively support maintenance activities. By providing humans with timely and necessary information based on heterogeneous data sources in a usable way, maintenance planners and technicians can improve their efficiency, leading to increased availability and reduced MTTR.

In this overlapping area, IAI provides the analytical and predictive capabilities necessary for proactive maintenance, while assistance systems deliver this information in an accessible and context-sensitive manner to maintenance personnel. The integration ensures that technicians receive the right information at the right time, aligned with their skill levels and the specific tasks at hand. This holistic approach not only addresses the technical challenges of maintenance but also considers human factors, such as cognitive load and skill diversity among personnel.

Furthermore, the content orientation is designed to enable a targeted assessment of the state-of-the-art (see Chapter 4). Understanding the interplay between these domains allows for a comprehensive evaluation of existing solutions and the identification of gaps that the proposed research aims to fill. The chapter concludes with the approach to task allocation, the principle of which will be crucial for the subsequent derivation of the solution approach. Task allocation strategies will leverage insights from maintenance, assistance systems, and IAI to assign tasks effectively based on individual competencies, further enhancing operational efficiency and asset performance.

Figure 8: *Cognitive Assistance Systems in Maintenance at the Intersection of Maintenance, Industrial Artificial Intelligence, and Assistance Systems*



2.1 Maintenance

This section provides a comprehensive examination of the conceptual and operational foundations of maintenance within industrial contexts, situated within the definitions and guidelines set forth in DIN 31051. The text defines maintenance as a structured interplay of technical, administrative, and managerial interventions designed to sustain or restore asset functionality over the equipment lifecycle. Moreover, this section presents a systematic categorization of core maintenance activities, including inspection, repair, servicing, and improvement. Each of these activities is elucidated in terms of its contribution to enhancing system reliability, availability, and operational sustainability.

The strategic aspect of maintenance is emphasized to investigate the preventive and corrective paradigms and their respective roles in mitigating unplanned downtime while optimizing resource utilization. The discussion highlights the transformation of maintenance from a historically costly necessity to a strategic facilitator of operational efficiency and cost optimization. Analytical models, including maintenance cost curves and maturity frameworks, are employed to illustrate the complex trade-offs inherent in balancing maintenance expenditure with performance outcomes.

By addressing the critical importance of aligning maintenance strategies with organizational objectives such as minimizing downtime costs, enhancing asset reliability, and supporting sustainability goals, this section provides a robust foundation for subsequent analyses of advanced maintenance methodologies. It sets the stage for exploring predictive, cognitive, and knowledge-based maintenance approaches that leverage emerging technologies to meet the demands of modern industrial systems.

2.1.1 Maintenance Strategies

According to DIN 31051:2019-06, maintenance can be defined as follows:

"*Combination of all technical and administrative measures as well as management measures during the lifecycle of a subject matter to preserve or restore its functional condition to secure the functional liability required.*"

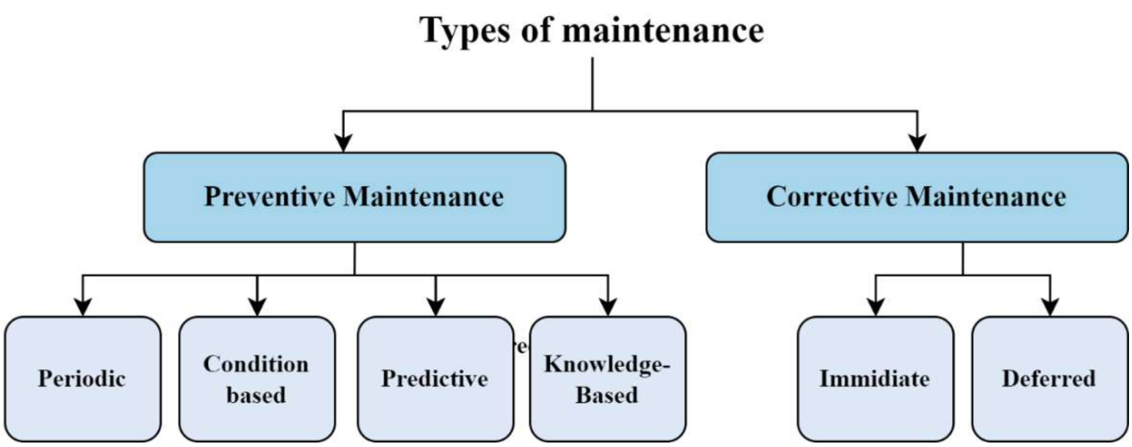
Maintenance is, therefore, a part of plant management, responsible for production safety, maintaining value, and ensuring the effective use of equipment (Pawellek, 2013). Its main goal is to maximize equipment reliability and reduce total costs (Matyas, 2022). Important areas of maintenance include the planning, organization, execution, and monitoring of all processes. According to DIN 3051, maintenance measures encompass the inspection, repair, and improvement of machines and equipment.

The terms maintenance, repair, inspection, servicing, and improvement are defined as follows in DIN 31051:

- **Maintenance:** Defined as the collection of technical, administrative, and managerial activities throughout the life cycle of an asset to preserve or restore its operational condition to perform its intended function
- **Repair:** Refers to the tangible actions taken to restore the functionality of a malfunctioning unit
- **Inspection:** Includes activities focused on evaluating the current condition of a unit, identifying causes of wear, and formulating implications for its future use
- Service: Includes activities designed to slow the deterioration of the existing wear margin
- **Improvement:** Includes a mixture of technical, administrative, and managerial actions designed to increase the reliability, maintainability, and safety of a unit while maintaining its original function

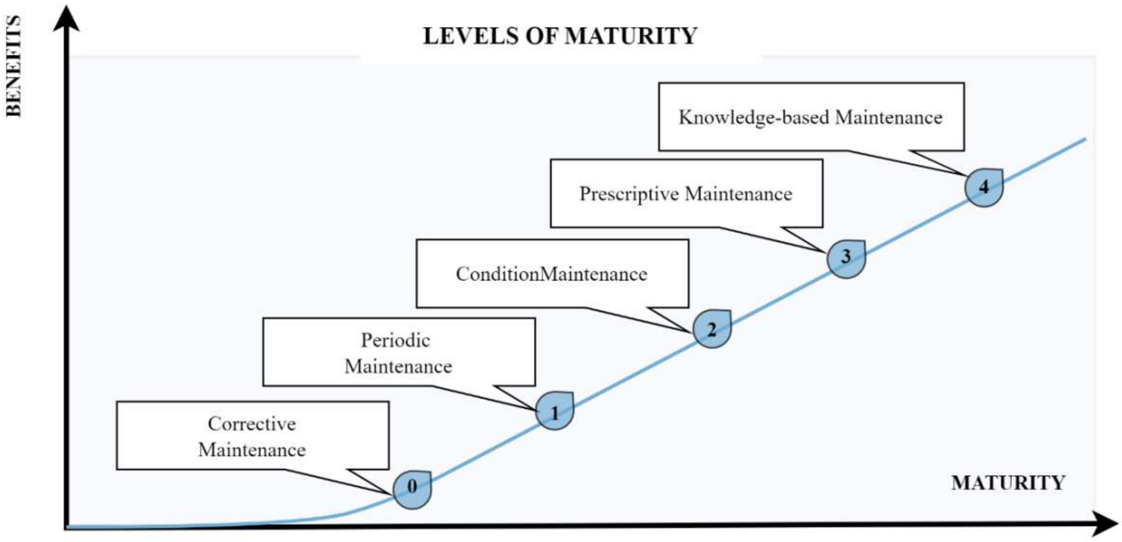
The maintenance strategy plays a significant role in maintenance because it determines the timing, content, and scope of the maintenance measures to be applied (Hölbfer, 2014).

Figure 9: *Types of Maintenance Strategies Inspired Extended from DIN 31051*



Maintenance strategies, see Figure 9, are crucial for ensuring the longevity and reliability of machines and equipment in industrial settings. The types of maintenance strategies can be broken down into two main components: preventive maintenance (PM) and corrective maintenance. These two can subsequently be subdivided into different types of strategies, each with varying levels of maturity, which can be selected depending on the application areas and necessary requirements. It can be demonstrated that each phase in the evolution of maintenance strategies, driven by a clear rationale, must be selected, yielding corresponding benefits, see Figure 10.

Figure 10: *Maturity Levels of Maintenance Based on Ansari et al. (2019)*



Rewards	Short term savings	Fixed schedule	No spuprises	Competitive Advanage	Best in class
Motivator	Meeting budget	Quick respnse times	Avoiding unplanned maintenance	Maximize uptime	Organizational growth
Behaviour	Decaying	responding	Discipline	Learning	Innovating

In this context, the systems and the maintenance organization must exhibit specific behaviors. Consequently, the highest level KBM is not always the optimal choice; instead, a maintenance strategy with a lower level of maturity may be more suitable.

The primary objective of maintenance within industrial operations is to optimize the balance between minimizing total production downtime costs and the expenditure on maintenance resources, thereby contributing to favorable business outcomes. Operational goals of maintenance, such as enhancing plant availability (see equation 1), reliability, reducing downtime, and minimizing both direct and indirect maintenance expenses in the form of maintenance intensity (see equation 2) stem from this overarching objective. However, it is important to recognize that these goals may sometimes conflict with each other (Biedermann, 2016a; Henke, 2019). Additionally, maintenance objectives may extend to include other goals aligned with company policies, such as reducing energy and resource consumption (Pawellek, 2016).

$$Availability = \frac{average\ operating\ time}{Average\ operating\ time + \ average\ downtime} \quad (1)$$

$$Maintenance\ intesity = \frac{Yearly\ maintenance\ costs}{Replacement\ value\ of\ the\ machine} * 100 \quad (2)$$

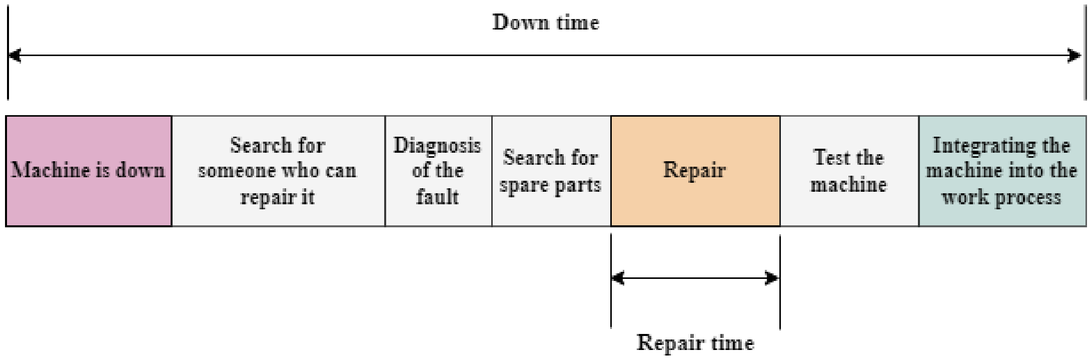
Matyas’ (2022) examines the parameter of availability and the potential for enhancement by implementing a cognitive assistance system leveraging AI. Primarily, the proposed augmentation of the competencies of maintenance planners and technicians is reflected in a reduced MTTR, which directly impacts availability by reducing average downtime. Furthermore, the cognitive assistance system proposed in this dissertation and its underlying maintenance philosophy align with the knowledge-based maintenance strategy (Matyas, 2022).

2.1.2 Preventive Maintenance

Preventive maintenance (PM) aims to minimize unplanned downtime and reduce maintenance costs by executing scheduled maintenance activities. DIN EN 13306 characterizes PM as a process aimed at evaluating and mitigating deterioration, thereby reducing the likelihood of an object’s fault (Deutsches Institut für Normung, 2018).

Periodic maintenance, which is a prevalent strategy in this context, involves the proactive replacement or refurbishment of components after a specified usage period, irrespective of their current state. This period is determined based on time or operational events, such as operating hours, calendar duration, lifting operations, or production output. For this approach, it is critical to understand the fault behavior of the components involved, typically quantified by the mean time between faults (MTBF). This knowledge facilitates the scheduling of component replacements during non-production periods, optimizing maintenance duration as personnel and spare parts are pre-arranged (Schenk, 2010). However, this strategy often leads to premature component replacement to avoid unscheduled downtimes, resulting in underutilization of the total wear capacity. Depending on the implementation, this can lead to over- or under-utilization of the components. In both cases, the availability of the system is reduced accordingly. In the case of underutilization, increased planned maintenance activities lead to more downtime. Conversely, in the case of overutilization, unplanned maintenance activities lead to more downtime. Both have a corresponding negative effect on the total wear inventory of the asset and its underlying equipment and thus on the remaining useful life (RUL). Furthermore, the MTBF for different components within a system can vary significantly, necessitating varied intervals for periodic maintenance activities. For new assets, the lack of extensive statistical data hampers the ability to make reliable predictions about average downtime. Due to these uncertainties, components are often replaced well before reaching their functional limits, leading to inefficiencies in maintenance practices.

Figure 11: *Distribution of Times in Maintenance Matyas (2022)*



2.1.3 Predictive Maintenance

The term PdM is defined as a condition-based approach where sensor data and predictive data analytics are used to forecast asset or equipment faults (DIN 31051:2019–2006). It is essential to perform these predictions on parameters that characterize the deterioration of the equipment (Lee et al., 2013) and (Lee et al., 2006) PdM relies on historical data, models, and domain knowledge to anticipate potential faults in advance. By utilizing statistical or ML models, PdM can predict trends, behavioral patterns, and correlations, thereby enhancing the decision-making process for maintenance activities and primarily preventing downtime (Lee et al., 2006; Sezer et al., 2018). The implementation of PdM, along with methodologies aimed at enhancing manufacturing capabilities, has led to the emergence of terms such as intelligent industry or intelligent manufacturing (Kiangala & Wang, 2018). However, the implementation and use of PdM require a wide range of competencies, especially for identifying critical systems and developing analysis models. Implementation and operation also place high cognitive demands on maintenance planners, who must now use the appropriate AI systems, interpret their results, and translate them into actionable tasks for maintenance technicians for planning purposes.

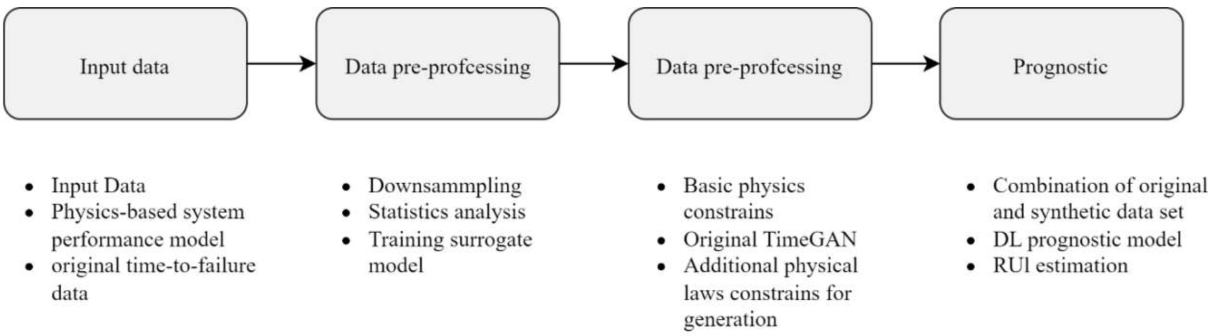
The maintenance impact constitutes a significant portion (15–60%) of the overall operational costs in manufacturing (Haarman et al., 2017; Mobley, 2002). However, companies often lack accurate measurements of their maintenance-related expenditures, highlighting the need for research on the utilization of new technologies that can address this issue. This is particularly significant regarding the impact of PdM and its potential as a differentiating factor in implementing Industry 4.0.

In particular, the use of large amounts of sensor data for the RUL calculation enables a far-reaching implementation of an Industry 4.0 approach in the company (Yan et al., 2017). In addition, there are new AI-related possibilities that enable subsequent model calculations for the RUL, one of which is to use RUL in a PdM approach (Yan et al., 2017). The notion that

PdM can generate actionable scheduling based on equipment performance or condition over time enables cost-efficient (Wu et al., 2016) and opportunistic maintenance strategies (Colledani et al., 2018). A key requirement for effective PdM implementation is the availability of sufficient data from all stages of the manufacturing process (Kiangala & Wang, 2018); consequently, this can lead to reduced maintenance costs, minimized downtime, and improved productivity and quality.

The difficulty of forecasting the RUL of assets is typically encountered during their operational lifespan, which is when sensors must be retrofitted, and models may have to be developed without the original equipment manufacturer (OEM). Therefore, a comprehensive understanding of the asset or equipment structure is crucial.

Figure 12: *Framework for Physics-Informed RUL Prediction Adopted From (Xiong et al., 2023)*



This concept is an integral component of prognostics and systems health management (PHM), representing a comprehensive framework for RUL prediction, see Figure 12. Overall, PHM encompasses three main dimensions: observation, analysis, and action (Atamuradov et al., 2017; Kwon et al., 2016; Terrissa et al., 2016). In this context, research related to PdM is directly associated with the observation dimension of PHM, employing intelligent methods to forecast faults (T. P. Carvalho et al., 2019). Some authors consider PdM to be one of the components of PHM, alongside equipment health monitoring and MTTR (Iskandar et al.., 2015).

As for PdM, another crucial aspect concerns the three classifications of prediction approaches, as outlined by Deutsch et al. (2017), B. Li et al. (2019), and (Wu et al., 2016):

- **Physical model-based:** This approach relies on mathematical modeling, considering the component's condition and fault measurements. It necessitates precision in condition assessment and employs statistical methods to constrain these variables (Ansari & Kohl, 2022).

- **Knowledge-based:** These approaches aim to simplify the complexity associated with physical models and often serve as hybrid strategies. Examples include expert systems or fuzzy logic, which combine domain knowledge with computational techniques (Ansari et al., 2023; Ayad et al., 2018)
- **Data-driven:** This category encompasses models commonly found in the current advancement of PdM solutions. They are predominantly based on statistics and pattern recognition (Iftikhar et al., 2023; Wu et al., 2018).

In examining the maintenance philosophy, it is evident that the presented PdM approaches demonstrate that, while pure white-box models such as physical PdM models offer a high degree of explainability, developers require a comprehensive understanding of the asset, equipment, or even the components installed. Moreover, purely data-based models are black box models and neglect certain contexts when data are missing or unbalanced. Therefore, the knowledge-based approach, with its gray box design, represents the optimal methodology for a cognitive assistance system, as it allows for the incorporation of expert knowledge in addition to data-based models.

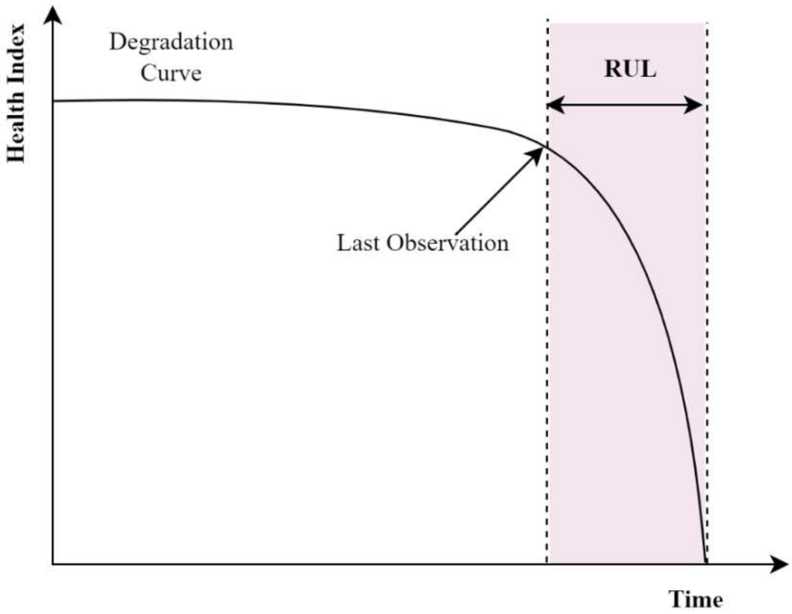
The predominant methodology in several PdM models is the pure black box approach, which presents a significant challenge in combining data from diverse sources—including condition monitoring, historical error records, and machine and quality data—to accurately predict fault behavior (Nemeth et al., 2019). While the scientific community typically classifies PdM approaches based on the methodologies (Ansari et al., 2019; Nunes et al., 2023; van Dinter et al., 2022; Zonta et al., 2020) employed, this study endeavored to categorize PdM approaches according to the means by which the defects are identified.

- **Forecasting Remaining Useful Life**: This approach utilizes traditional reliability management methods, such as determining Weibull parameters (Jeon & Sohn, 2015) or calculating survival functions using Kaplan–Meier estimators (Ragab et al., 2016) for predicting Rul, see Figure 13. While effective for planning multiple similar components, such as in spare parts forecasting or service personnel scheduling, its applicability to specific components is limited.
- **Forecasting Wear Stock:** This method aims to predict the future condition of a component, typically through condition monitoring measurements or expert assessments. By incorporating historical fault data and additional sensor and process

data, this approach yields more meaningful forecasts applicable to specific components (Zhai et al., 2021).

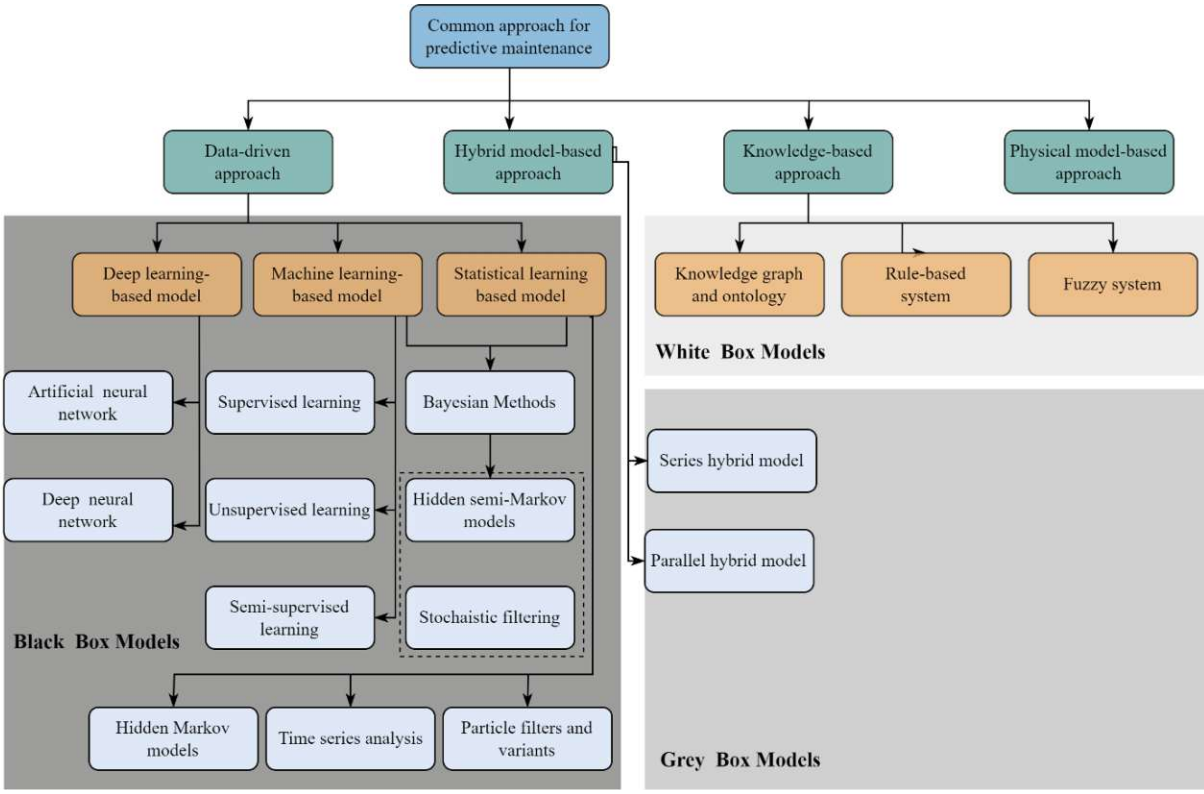
- **Prediction of Time to Fault (TTF):** Unlike wear stock prediction, this approach estimates the remaining time until a component's fault. It is particularly relevant when direct condition assessment is infeasible, relying primarily on historical fault data and various process, machine, and quality data. However, the availability and quality of historical fault documentation often pose challenges in industrial (Jalali et al., 2019).

Figure 13: *Maintenance Fault Curve*



In industrial applications, a gray-box PdM model is advantageous, combining different approaches to cater to the specific maintenance requirements of various components (Luo et al., 2020), see Figure 11. While the development of a bespoke hybrid PdM approach for a particular application is more labor-intensive than traditional PdM approaches (Shcherbakov & Sai, 2022), the latter are increasingly becoming insufficient for ensuring the required reliability (Wellsandt et al., 2022). The growing complexity of modern production facilities and the variable load spectra of flexible manufacturing systems render constant MTBF values impractical, complicating the definition of intervention limits for wear stock based on uniform load, thereby making the need for hybrid PdM more prevalent (Montero Jimenez et al., 2020). It is particularly noteworthy that the physical model-based approaches must be developed individually, component-specifically, and, in this case, possibly even for the corresponding application. Consequently, these models cannot be further subdivided.

Figure 14: *PdM Approaches Based on Montero Jimenez et al. (2020)*



2.1.4 Knowledge-Based Maintenance

Knowledge-based maintenance (KBM) has evolved with the integration of AI, transforming maintenance into a more intelligent and prescriptive practice. This transformation is primarily attributable to ongoing advancements in semantic technologies and their integration with generative AI methodologies. Presently, KBM involves leveraging AI to automate and enhance the processes of knowledge creation, acquisition, storage, retrieval, and application within the maintenance domain (Ansari et al., 2019). Notably, AI advancements facilitate predictive maintenance, improve diagnostic accuracy, and optimize maintenance schedules, revolutionizing how knowledge is utilized in maintaining industrial systems. This approach not only streamlines maintenance tasks but also significantly boosts efficiency and accuracy, ensuring that maintenance organizations remain at the forefront of technological innovation (Kovacs et al., 2019; Nemeth et al., 2019).

Moreover, KBM is an advanced approach that integrates expert knowledge, data analytics, and intelligent systems to optimize maintenance activities in industrial environments. By ensuring that personnel possess an expert-level understanding of production and manufacturing processes through education, training, and critical thinking, KBM enhances decision-making, improves efficiency, and reduces downtime.

In the era of intelligent manufacturing, KBM plays a crucial role by facilitating semantic interoperability and efficient knowledge exchange among various stakeholders. This integration supports the development of smart manufacturing systems capable of adapting to new challenges and incorporating innovative technologies to improve production processes (Adamczyk et al., 2020).

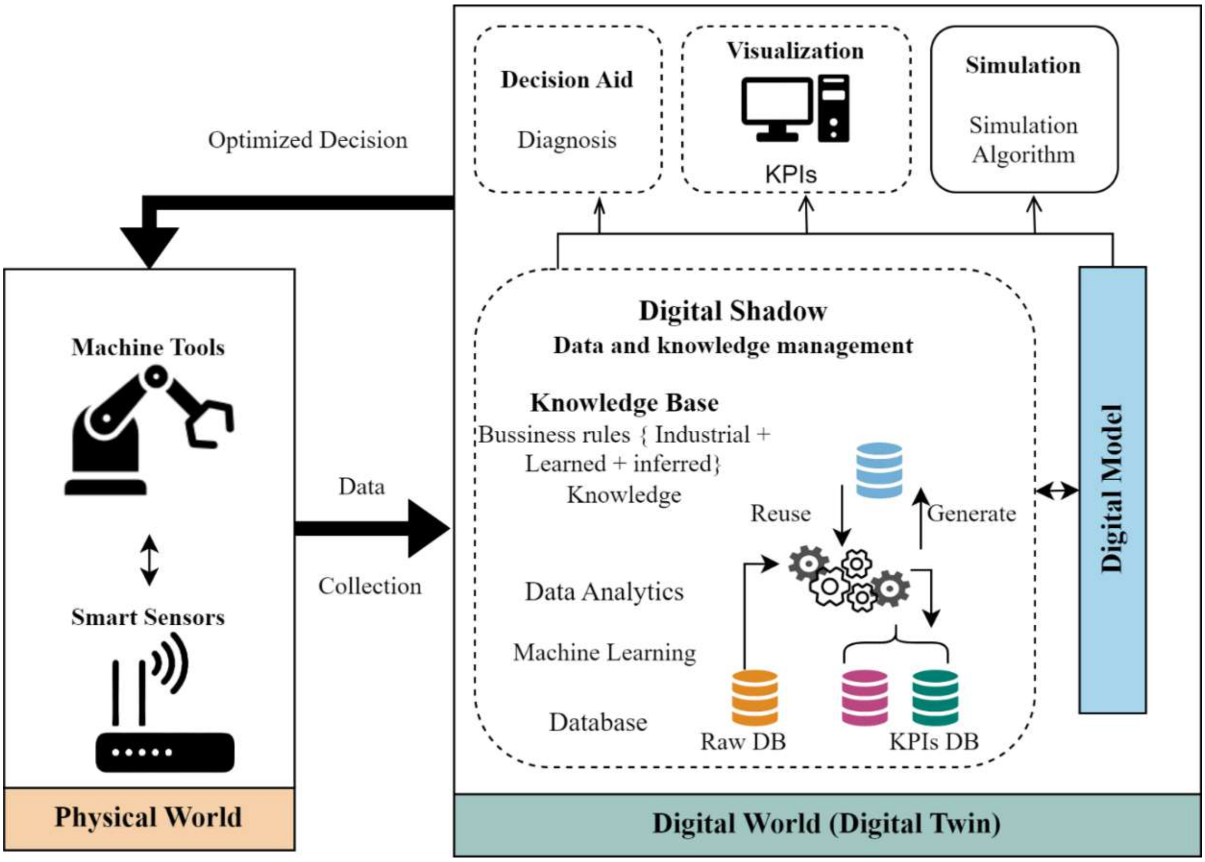
Effective knowledge representation is essential for KBM, with ontology-based models being a common method to structure and formalize knowledge within a domain. Ontologies enable the sharing and reasoning of knowledge by defining concepts and relationships, which is vital for promoting information exchange among different parties involved in maintenance activities (Wan et al., 2019). However, challenges exist in representing unstructured or tacit knowledge that cannot be easily depicted in predefined formats.

KBM often incorporates decision support systems that utilize data analytics to process substantial volumes of diverse and complex information (Ansari, 2019). These systems generate actionable insights and recommendations that enhance and optimize future maintenance plans, integrating seamlessly with production planning and control systems (Ansari et al., 2019). The development of maintenance control centers and mobile interfaces facilitates the presentation of crucial key figures, technical information, and the spatiotemporal development of assets, supporting prescriptive maintenance measures.

Combining data analytics with knowledge management systems enhances the ability to derive insights from both explicit and inferred knowledge. Moreover, organizing information using ontology-based knowledge models allows for the incorporation of various types of insights, including industrial data, expert principles, and findings from data analytics (Ladj et al., 2021). Further, ML techniques such as unsupervised learning are utilized to categorize contexts and identify detrimental incidents, demonstrating the practical applicability of these methods in maintenance.

While ML and AI offer significant potential for enhancing KBM, understanding their full impact on maintenance methodologies remains a challenge, as seen in Figure 15. Hence, there is a need for further research to explore how ML-based approaches and AI developments are transforming maintenance practices within industrial environments (Merkt, 2019). Addressing issues related to knowledge representation, scalability, and the processing of diverse data is essential for advancing KBM.

Figure 15: *The Overall Flow of the Digital Shadow Based on Merkt (2019)*



As proposed in the maintenance philosophy, a gray-box model that focuses on the integration of KBM with assistance systems and IAI forms the foundation of modern maintenance strategies:

- KBM provides the expert knowledge and procedural guidelines essential for effective maintenance.
- Assistance systems support maintenance personnel by delivering real-time guidance, interactive tools, and user-friendly interfaces tailored to individual competencies and tasks.
- IAI enhances these systems by offering advanced analytics, predictive capabilities, and intelligent decision-making support.

This synergy enables organizations to develop intelligent maintenance support systems that improve asset availability, reduce MTTR, and maintain high-quality standards. By leveraging AI-driven insights within user-friendly assistance tools, maintenance technicians receive the right information at the right time, enhancing their ability to troubleshoot issues effectively, especially in high-pressure, hands-on settings.

2.2 Assistance Systems

Assistance systems are designed to aid workers in manufacturing, assembly, and logistics sectors by enhancing the efficiency of their tasks, tailored to the situation and context at hand. These systems facilitate a transition toward paperless management across manufacturing, assembly, and logistics operations. Additionally, they can minimize the time spent on orientation, training, communication, and information retrieval, thereby streamlining various (Hold et al., 2017).

Accordingly, assistance systems enable a human-centered approach to production and, in particular, maintenance, as highlighted in this study. These assistance systems can be extended with hybrid intelligence that provides maintenance planners and technicians with information based on a gray-box approach. The approach here involves not automating activities but enabling the augmentation of human skills and competencies in line with the maintenance philosophy.

This results in an expansion of the assistance systems toward definition; digital assistance systems (DAS) integrate advanced technologies such as AI, ML, and the Internet of Things (IoT) to enhance operational efficiency and decision-making (Kohl & Ansari, 2023). They provide real-time data analysis, predictive maintenance insights, and personalized guidance to workers, thereby streamlining processes and reducing the likelihood of errors. These systems can also adapt dynamically to changing conditions and requirements in the workplace, offering contextual and situation-specific assistance. Further, by augmenting human capabilities with digital AI, these systems not only optimize current operations but also pave the way for continuous improvement and innovation in the fields of manufacturing, assembly, and logistics (Mark et al., 2021). The implementation of DAS marks a significant step toward the realization of KBM, where intelligent automation and human expertise converge to create more agile, efficient, and competitive industrial environments. This approach, therefore, aims to augment human performance rather than automate or rationalize it away (Schuh et al., 2015). While this may suggest increased physical, mental, or emotional stimulation, it is important to note that most tasks necessitate a combination of these resources and require a minimum level of energy. This definition now needs to be supplemented with hybrid gray-box intelligence so that a cognitive system can be defined as “an autonomous system that can perceive its environment, learn from experience, anticipate the outcome of events, act to pursue goals, and adapt to changing circumstances” (Vernon, 2021). Cognitive assistance systems (CASs) aim to augment human skills in intricate tasks without replacing human labor in production, assembly, and logistics. By providing tailored support and learning to individual competence levels from a knowledge base, CASs offer more fitting, accurate, and timely data in applications (Kohl, Ansari, & Sihn, 2021) Moreover, CASs support sophisticated tasks such as lifelong learning, machine operation, and task management through improved human-machine interaction, employing tools such as NLP, ergonomic risk assessment through pose estimation, and augmented reality for an intuitive user experience.

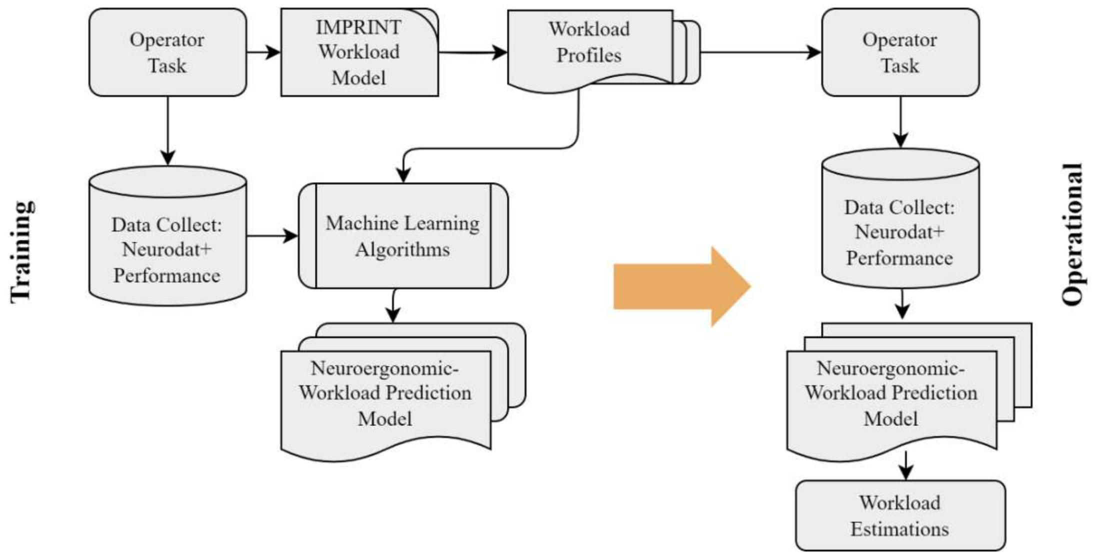
Conversational agents, which use NLP for understanding, generating, and managing dialogues, are at the forefront of interaction within CAS (Kang et al., 2020). Capable of meaningful interactions, these agents support labor-intensive activities in various domains such as customer support, healthcare, education, and manufacturing, enhancing communication efficiency and reliability (Eloundou et al., 2023). Within the industrial sphere, the deployment of CAS is an emerging area of research offering notable advantages (Mark et al., 2021), including centralized access to varied information systems, support in decision-making processes (Rožanec et al., 2022), task delegation (Burggräf et al., 2021), and facilitating interactions that do not require hand or eye coordination (Romero & Stahre, 2021). This not only improves operational efficacy and safety but also provides on-the-job training tools (Wang et al., 2022) and enables real-time adjustments to machine settings, thus enhancing the versatility and agility of manufacturing operations (Zheng et al., 2022). These implementations underscore the pivotal role of cognitive assistants in enriching human labor, refining task management, and promoting ongoing learning and adaptability in sophisticated industrial settings.

CASs are designed to alleviate mental strain by reorganizing information, aiding in decision-making, and offering support during periods of distraction due to non-standard tasks or demands. Moreover, they also facilitate the assessment of different options. The concept of mental workload is tied to the balance between an individual's cognitive capacity and the cognitive demands of a specific task (Ansari, Hold, et al., 2018). Work-related mental strain is understood as a fluctuating dynamic, alternating between overburdening and under-stimulation (Zäh et al., 2007). Within this continuum, there exists an optimal workload range for peak productivity, which varies among individuals depending on various factors. Transgressing these boundaries temporarily may not significantly impact performance (Azari et al., 2023).

A CAS demands an in-depth comprehension of the working station's processes. Thus, a process modulation such as work domain analysis needs to be implemented. Rusnock and Borghetti (2018) presented a method for continuously estimating work without interrupting the operator. Initially, the continuous workload evaluation is represented as a time series, which transforms into a workload profile that can be utilized before, during, and after the execution of a task Figure 16. Therefore, process analysis serves as more than just an initial reference. Moreover, it also functions as a foundational model that depicts the distribution of mental stress states characterized by high and low levels during the assembly or maintenance process.

On the other hand, an approaching advancement is the generation of assembly systems that operate on a cyber–physical systems (CPS) framework. For example, in their study, Hold et al. (2017) outlined a recently manufactured curriculum intended for industrial engineering graduate students. Methods for planning and assessing complex cyber–physical assembly systems and their corresponding DAS were covered in this course. An approach that can even be extended to DAS, focusing on virtual learning factories and scenarios, was described in (Kohl, Stricker, et al., 2024). In another study, Gerdenitsch et al. (2021) presented an analysis for optimizing user acceptance of head-mounted displays concerning their perceived usability and simplicity of use. First, the study recommended maintaining its design in its current form due to the notably better feedback received regarding hands-free operation, adaptability, and control of the technology. Second, the study focused on accurate visualization on the screen. Third, the study recommended using the technology adaptation, as well as assessing appropriate use cases and work duties for each technology that is to be implemented in the workplace.

Figure 16: *The Overall Flow of the System (Rusnock & Borghetti, 2018)*



In conclusion, it can be said that assistance systems have become prevalent in manufacturing, assisting workers in various tasks, including maintenance (see Figure 17). Maintenance assistance systems can support maintenance planners and technicians in problem-solving (Ansari, Khobreh, et al., 2018), identifying root causes more quickly, optimizing maintenance operations, and streamlining non-value-adding activities, such as documentation. One crucial development area in maintenance assistance systems is CAS, designed to enhance human cognition and decision-making abilities. By combining human expertise with AI techniques, CAS can improve maintenance effectiveness and efficiency.

Notably, CASs are particularly effective in maintenance operations, where complex problems require expertise and experience to diagnose and resolve, thereby supporting maintenance technicians in various areas, see Figure 17. These systems can provide maintenance technicians with real-time access to information such as historical maintenance data, sensor data, and equipment manuals. By integrating structured data from sensors and other sources with unstructured data, such as text from equipment manuals or maintenance reports, these systems can provide technicians with a comprehensive view of equipment health and maintenance needs.

Figure 17: *Morphology of Assistance Systems in Maintenance (Samtleben & Rose, 2019)*

Dimension	Feature							
Knowledge retrieval	Employee obligation	System obligation						
Application possibility	Permanently installed	Flexible use possible	Flexible					
Implementation of the assistance system	Open	Partly open	Closed					
Type of support	Cognitive	Mechanical	Visual	Auditive				
Planned use of information	Process	View	Analyze	Forecast				
Through assistance system provided information	Building plan	Checklists	KPIs	Machine values				
Technical support means	Screen	AR-googles	Cameras	Robots	Projection	Headset	Identifi-Cation Feature	Tablet
Process monitoring	Automatic	Hybrid	Manuel					
Automation degree	Fully automatic support	Support with fixed target	Intension detection with solution suggestions					
Activity restriction	Without restriction	Restricted	No paralell working possible					
Input	Touch-Screen	Mouse/ Keyboard/ joystick	Feature Scanning	Gesture control	Voice control			

PM and PdM strategies can significantly benefit from CAS. Whereas PM involves performing regular maintenance activities on equipment to prevent breakdowns or faults, PdM utilizes real-time data from sensors to identify potential problems and take action before equipment faults occur. With the assistance of cognitive assistance systems, maintenance technicians can identify patterns in sensor data, predict equipment faults, and take proactive measures to prevent downtime.

The impact of CAS on maintenance costs, availability, and human knowledge is significant. These systems can reduce maintenance costs and improve equipment availability by automating routine tasks and augmenting technicians with real-time information. They can also help to increase human knowledge by providing access to information that might otherwise be difficult to obtain, such as the root cause of a particular problem or the best maintenance strategy for a specific piece of equipment. Furthermore, by reducing the incidence of human error, cognitive assistance systems can help to reduce the risk of machine and human faults, resulting in safer and more reliable operations (Ansari, Khobreh, et al., 2018). It is, therefore, possible to assign the role of the assistance system to the respective maintenance levels described in Figure 9 and their roles (see Table 1).

Table 1: *Role of Assistance Systems in Maintenance Maturity Levels*

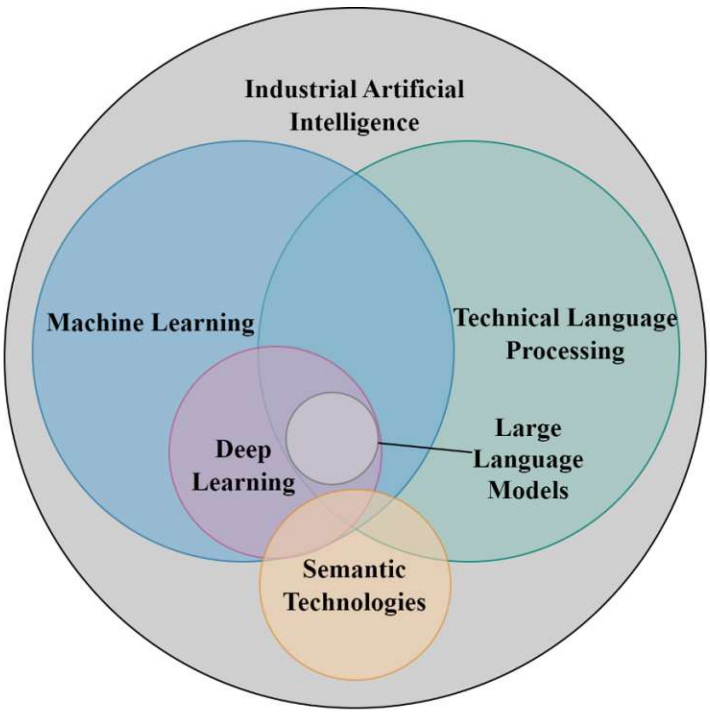
Maintenance Maturity Level	Role of Assistance Systems
Reactive	Problem-solving, root cause identification with historical maintenance reports, documentation
Planned	Shift planning and task allocation
Predictive	RUL estimation for improved planned maintenance
Prescriptive	Improved shift planning and task allocation based on advanced RUL forecasts and identified root causes
Knowledge-based	Reasoning over semantically interlinked data enables continuous improvement of maintenance activities

2.3 Industrial Artificial Intelligence

Section 2.3 introduces the foundational concepts of IAI and its integration into maintenance systems, exploring how AI techniques, ranging from ML to KG frameworks, are employed to enhance maintenance efficiency, precision, and decision-making. This section delves into the role of IAI in optimizing processes such as fault diagnosis, predictive analytics, and resource allocation, leveraging structured and unstructured data from diverse sources. Further, by addressing key methodologies, including anomaly detection, NLP, and cognitive systems, this section establishes the groundwork for understanding how AI-driven technologies transform maintenance into a proactive, knowledge-intensive discipline.

While the field of AI lacks a universally accepted definition, it can be broadly understood as a branch of computer science focused on creating systems that mimic human cognitive abilities such as reasoning, learning, and self-improvement (Russell et al., 2010). The OECD defines an AI system as a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing physical or virtual environments (Yeung, 2020). Often, ML is considered a core subfield of AI due to its emphasis on learning from data, closely related to statistical learning (Hastie et al., 2001). However, AI encompasses a broader range of techniques. From an industrial perspective, AI technologies empower systems to perceive their surroundings, process acquired data, solve complex problems, and continuously improve their performance through experience. This capability to learn and adapt makes AI valuable for various industrial applications (Peres et al., 2020).

Figure 18: *Relation Between AI, ML, DL, and NLP*



AI serves as the overarching domain encompassing various subfields, including ML, NLP, computer vision, and speech recognition. Generative AI refers to a subset of AI models designed to generate new data instances that resemble a given dataset. These foundation models learn the underlying patterns, structures, and features of the input data and can produce original content such as text, images, audio, or designs. Within this framework, ML is a subset of AI dedicated to enhancing algorithms through experiential learning and data analysis, which involves techniques like supervised, unsupervised, and reinforcement learning. Deep learning (DL), a branch of ML, employs DNNs to interpret complex data structures and patterns.

IAI can be defined as AI within industrial contexts to enhance operational efficiency, productivity, and safety, involving the integration of AI algorithms with industrial systems and processes, leveraging data from equipment, sensors, and operations to make informed decisions (Peres et al., 2020; Zonta et al., 2020). IAI plays a pivotal role in modernizing manufacturing processes, significantly enhancing efficiency and precision. Moreover, it improves critical steps in modern manufacturing, such as design, engineering, diagnosis, prediction, optimization, and decision-making.

The intersection of generative AI and IAI presents significant opportunities for advancing maintenance practices. Some of these opportunities are as follows:

- **Enhanced Predictive Capabilities:** Generative models can simulate a wide range of operating conditions and fault modes, enriching the predictive analytics of IAI systems.
- **Personalized Assistance:** Generative AI can create customized support materials, such as step-by-step repair instructions or training modules tailored to individual technician competencies.
- **Adaptive Scheduling:** Combining generative simulations with real-time data allows for dynamic adjustments of maintenance schedules to optimize asset availability and workforce utilization.
- **Innovative Problem-Solving:** Generative AI can propose novel solutions to maintenance challenges, which can then be evaluated and implemented within the IAI framework.

These advancements, crucial for the evolution of the manufacturing sector, have been extensively analyzed in recent literature (Çınar et al., 2020; Gualtieri et al., 2021), which also emphasizes key application areas, including product quality detection and production process monitoring. Furthermore, research suggests that using IAI in PdM significantly reduces manufacturing costs and enhances asset performance.

The convergence of AI with digital twins (DTs) and the industrial metaverse presents new opportunities for augmenting knowledge, expertise, and skill acquisition in industrial maintenance. DTs are virtual replicas of physical systems that allow for real-time monitoring, simulation, and analysis. The industrial metaverse extends this concept by creating immersive virtual environments where industrial operations can be visualized and interacted with in innovative ways (Bordegoni & Ferrise, 2023). This integration enhances training and maintenance applications by offering novel approaches to simulate maintenance tasks, troubleshoot issues, and optimize system performance within a virtual setting (Neuhüttler et al., 2022). By leveraging AI within these environments, organizations can improve the efficiency and effectiveness of maintenance activities, leading to reduced downtime and enhanced operational reliability.

Advancements in IAI have led to the development of cognitive maintenance systems that utilize multi-sensor data fusion for comprehensive environmental sensing, which focus on generalizability, extensibility, scalability, consistency, and user acceptance, addressing common challenges in implementing IAI within manufacturing environments (Kohl, Ansari, & Sihn, 2021). Such systems aim to overcome basic drawbacks in IAI implementation by ensuring that maintenance solutions are transferable and scalable across different contexts. However, limitations exist, including the need for multiple data repositories for data acquisition and processing. Therefore, integrating FUs such as data analysis, storage, decision-making, and human-centered support requires domain knowledge and appropriate IAI algorithms, highlighting the importance of interdisciplinary expertise in deploying these systems effectively.

ML plays a critical role in PdM, particularly during tasks such as anomaly detection and prediction. Generally, anomaly detection involves classification or clustering to identify deviations from normal operational patterns, while prediction tasks often utilize regression analysis to forecast future states of equipment (Biegel et al., 2022).

Frameworks that incorporate ML and cognitive approaches in Industry 4.0 settings enhance the effectiveness of PdM strategies by enabling advanced reasoning and decision-making capabilities. However, implementing ontologies and ML methods in PdM poses challenges, especially in facilitating reasoning for RUL prediction (Dalzochio et al., 2020; Traini et al., 2019).

Effective data integration is crucial for IAI applications due to its significant impact on organizational performance (Ansari et al., 2020). Thus, gathering data from devices, providing contextual information, semantically enhancing data, and integrating information from various sources are vital for accurate maintenance assessments (Ansari et al., 2019). By implementing comprehensive data integration strategies, organizations can more accurately assess maintenance efficacy by quantifying costs saved and time devoted to maintenance activities.

Big data introduces additional challenges related to scalability, latency, and data security. IAI and, therefore, PdM heavily depend on substantial volumes of data (Arinez et al., 2020) necessitating collaborative data platforms and the use of heterogeneous data sources to address these concerns (Gao et al., 2020). Ensuring data security and managing the computational demands of processing large datasets are essential for successful deployment. For ML and DL approaches in particular, high volumes of data of sufficient quality prove crucial for achieving good practical results as well as high model quality (W. Zhang et al., 2019).

2.3.1 Semantic Technologies

Knowledge bases are a fundamental component of knowledge engineering, an AI discipline that originated in the 1950s. They store structured information about specific domains, thereby enabling intelligent systems to reason and make informed decisions (Abdelillah et al., 2024). Semantic technologies further enhance this by encoding meanings and knowledge separately from data, content files, and application code, enabling machines to interpret and process information with greater efficacy. Central to these technologies are ontologies, which are formal representations of concepts and relationships within a domain. Ontologies can be categorized as weak (i.e., lightweight) or strong (i.e., heavyweight). Weak ontologies provide basic classifications and simple relationships that are suitable for straightforward interoperability, whereas strong ontologies offer detailed concepts and logical constraints that are necessary for complex reasoning (Zheng et al., 2024). KGs structure knowledge in graphs, connecting entities and their relationships to facilitate semantic searches and data integration (Fensel et al., 2020). One way to store and organize the information extracted by NLP models is through the use of KGs (Paulheim, 2016). Notably, KGs have become increasingly popular for representing and managing complex knowledge and information. The Journal of Web Semantics defines a KG as "a large network of entities, their semantics types, properties, and as between entities" (Kroetzsch & Weikum, 2016). Färber et al. (2017) refers to a KG as a resource description framework (RDF), stating, "An RDF graph consists of a set of RDF triples where each RDF triple (s, p, o) is an ordered set of the following RDF terms: a subject $s \in U \cup B$, a predicate $p \in U$, and an object $U \cup B \cup L$. An RDF term is either a URI $u \in U$, a blank node $b \in B$, or a literal $l \in L$." In IAI applications, KGs enhance information retrieval by providing structured, semantically linked domain information to improve response accuracy and contextual relevance (Balas et al., 2022; Paulheim, 2016), which is especially valuable in domain-specific applications like manufacturing (Ladj et al., 2021). Therefore, KGs enhance the capabilities of cognitive assistance systems (CAS) by providing them with structured context information on specific user requests (J. Li et al., 2022). Leveraging the rich semantic relationships within KGs, chatbots can understand and process user queries more effectively, navigating through complex information networks to retrieve or infer accurate answers (Ansari et al., 2020).

The advent of Google©'s KG in 2012 marked a significant evolution in the application of KGs, leading to their widespread adoption across various domains for organizing KBs (Chicaiza & Valdiviezo-Diaz, 2021). This graphical representation not only illustrates direct relationships but also facilitates the exploration of complex, arbitrary interrelations among these entities, allowing for complex data analysis (Buchgeher et al., 2021).

In the realm of the manufacturing industry, particularly in maintenance, the role of KGs has become increasingly prominent as a means to enhance efficiency and optimize resource allocation. Implementing KGs in this sector is a transformative process, enabling a more systematic and interlinked approach to managing and utilizing heterogeneous information. By leveraging KGs, the industry can effectively organize and analyze data on machine performance, maintenance histories, and operational metrics (Buchgeher et al., 2021).

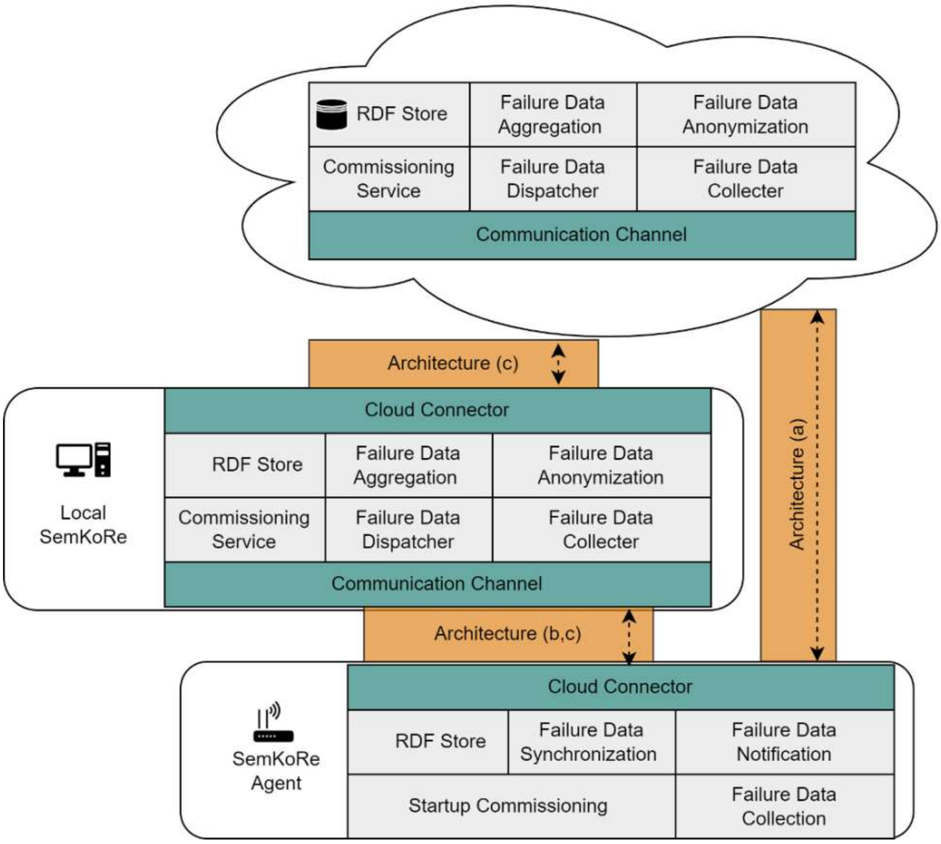
Cognitive intelligence refers to the simulation of human thought processes in a computerized model, which enables systems to learn from data, recognize patterns, and make decisions. While cognitive intelligence has been achieved through various established methodologies and applied across diverse fields (Lyu et al., 2022; D. Zhang et al., 2020), its application in PdM remains relatively uncommon (Mukherjee et al., 2024).

Graph-based approaches are emerging as highly efficient means to attain cognitive intelligence within IAI. By utilizing and aggregating relationships between assets, equipment, components, events, attributes, and semantic relations, these approaches can model complex interactions in industrial systems. Transforming data into graph structures allows for advanced analytics and reasoning, which is particularly beneficial in fault diagnosis and maintenance, where understanding the interplay between system components is crucial.

- **Fault Diagnosis with Graph Convolutional Networks:** DL techniques such as graph convolutional networks have been applied to fault diagnosis in industrial settings. By converting acquired signals from the general data domain to the graph domain, these networks effectively identify patterns indicative of faults. For instance, intelligent acoustic fault diagnosis methods using deep graph convolution networks have demonstrated improved performance in identifying roller-bearing faults (Zhang et al., 2020).

- **Development of Digital Assistants and KGs:** The advancement of digital assistants in industrial maintenance has been facilitated by creating KGs that model the knowledge required for maintenance tasks. Structured methodologies have been developed to address design challenges in creating these graphs, involving collaborative efforts between industry and academia (Fensel et al., 2020).
- **Actionable KGs for Maintenance Recommendations:** Solutions based on fuzzy frameworks have been proposed to develop actionable KGs that aggregate heterogeneous historical data. These models enable the retrieval and recommendation of practical suggestions to assist in daily maintenance tasks, thereby enhancing decision-making processes (Teern et al., 2022).

Figure 19: *Architecture of SemkoRe Enhanced (Buchgeher et al., 2021)*



KG-based approaches offer significant potential for improving maintenance processes by capturing and leveraging machine fault data from diverse sources. For instance, systems have been developed to collect data generated by various machines owned by multiple enterprises across different locations and industry sectors. By facilitating the exchange of maintenance experiences among OEM clients, these systems can reduce maintenance expenditures and improve OEE (Hossayni et al., 2020).

Frameworks have been introduced to convert maintenance report data, typically in natural language text, into formal KGs. These graphs encapsulate semantic relationships among maintenance entities derived from historical data in work orders. The generated KGs serve multiple decision-making applications, including maintenance performance monitoring and root-cause analysis (Martinez-Gil et al., 2022). They contribute to the development of an open-source, expandable industrial maintenance knowledge base that can be progressively enhanced (Buchgeher et al., 2021).

2.3.2 Technical Language Processing

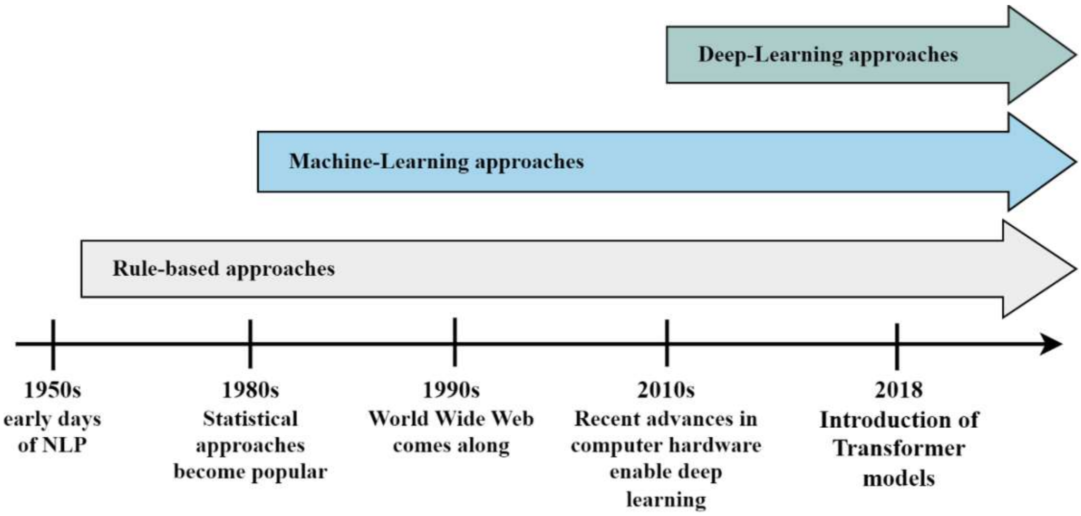
NLP, a pivotal area within AI and computer science, is defined as "the analysis of linguistic data, most commonly in the form of textual data such as documents or publications, using computational methods" (Verspoor & Cohen, 2013). Thus, NLP aims to bridge the gap between human language and computer understandability, acheiving this by leveraging linguistic principles to create structured representations of unstructured text. These representations can be syntactic, focusing on the grammatical relationships between words and phrases within the text, or semantic, capturing the underlying meaning conveyed by the language. In the area of industrial application and its grammatical specialties (Brundage et al., 2021), such as semantically incomplete short texts, the term TLP has been introduced. This structured format allows computers to analyze and manipulate natural language in a way that facilitates tasks such as machine translation, sentiment analysis, and automated text summarization (K. B. Cohen & Verspoor, 2013). As an interdisciplinary field, TLP integrates aspects of computer science, linguistics, manufacturing domain knowledge, and mathematics. The relationship between TLP and related domains such as AI, ML, and DL is depicted in Figure 18. Its fundamental objective is to convert human language into executable computer commands. TLP is bifurcated into two primary research streams: natural language understanding and natural language generation (Russell et al., 2010). Natural language understanding focuses on interpreting human language, analyzing texts, and extracting pertinent information for subsequent applications, as highlighted by Schank (1972). Conversely, natural language generation involves generating human-readable text from structured inputs such as data, text, graphics, audio, and video, as described by (Indurkhya & Damerau, 2010). Natural language generation is further categorized into text-to-text translation and summarization, as well as text-to-other forms such as image generation from text and other-to-text, exemplified by text generation from video content (Kang et al., 2020).

Contemporary TLP relies heavily on ML and DL techniques, integrating these computational methods to process and understand human language. Traditional NLP methods, such as word counting and text similarity measurements are fundamental to ML-based models in NLP, although not always classified under ML (Vajjala et al., 2020).

Historically, NLP began with rule-based and template-based systems, where language experts manually encoded patterns and rules for tasks such as translation and information extraction , see

Figure **20**. These methods, while leveraging expert knowledge, were limited in addressing the complexity and diversity of human language. The 1980s saw a significant shift with the advent of statistical methods and ML algorithms in NLP, using language data and statistics to improve performance and address language complexities. Unlike rule-based systems, statistical methods in NLP learn from data, allowing for more flexible predictions. However, they require large, high-quality datasets to perform effectively.

Figure 20: *Development of NLP*



In the 2010s, advancements in computer hardware enabled the adoption of powerful DL techniques in NLP, facilitating efficient learning from vast datasets and significantly advancing NLP capabilities. This development exemplifies the transition from rule-based systems to contemporary, data-driven methodologies that leverage the most recent advancements in AI, ML, and DL.

In the realm of NLP, various methodologies are employed depending on the task at hand, encompassing rule-based, statistical, and DL techniques (Verspoor & Cohen, 2013). Vaswani et al. (2017) categorized NLP approaches into heuristic-based, ML, and DL.

Initially, heuristic-based NLP utilized digitized resources such as dictionaries for tasks such as lexicon-based sentiment analysis. Advanced knowledge bases, such as WordNet, provided deeper semantic relationships between words (Miller et al., 1990; Vajjala et al., 2020). Regular expressions (regex) are instrumental in text analysis, enabling the integration of domain knowledge into NLP systems and facilitating deterministic frameworks like TokensRegex (Stanford NLP, 2013).

In ML for NLP, supervised techniques such as classification and regression are applied to tasks such as news categorization and stock price prediction. Moreover, algorithms like Naive Bayes and support vector machines are commonly used, with support vector machines transforming data to higher-dimensional spaces for classification (Hastie et al., 2001). Further, Hidden Markov models effectively capture the sequential nature of language, modeling the hidden parts of speech in text (Vajjala et al., 2020)

DL for NLP has seen the rise of neural network architectures such as recurrent neural networks and long short-term memory networks, with the latter addressing the memory limitations of recurrent neural networks. However, transformer networks have recently dominated the field, offering state-of-the-art performance by employing self-attention mechanisms for contextual word representation (Vajjala et al., 2020). The subsequent chapters will delve into transformer architectures and their applications in domain-specific NLP tasks.

Transformers, a groundbreaking development in DL for TLP, have rapidly ascended to the forefront of the field, with their unique approach, which diverges from traditional sequential models, leveraging self-attention mechanisms. This enables simultaneous processing of all input words during both encoding and decoding phases, significantly enhancing their efficacy in a range of TLP tasks (Hotho et al., 2005).

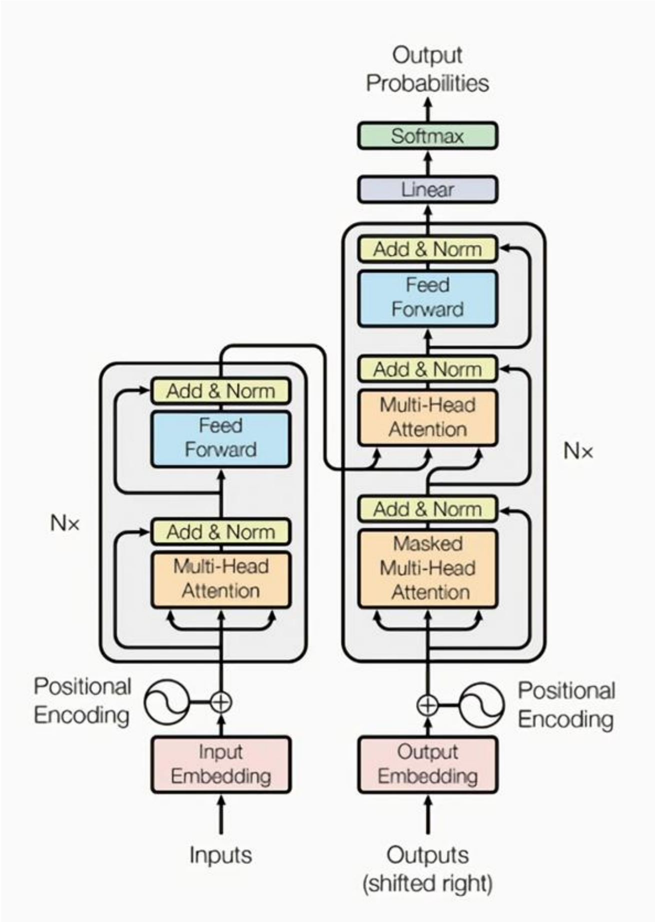
Transformer Architecture

At the heart of the transformer model, as described by Tunstall et al. (2022), lies the encoder-decoder architecture, see Figure 21. This design is characterized by the following:

- An encoder, which converts input sequences into a series of embedding vectors, representing the contextualized state of each input token
- A decoder, which iteratively generates an output sequence from the encoder's contextual embeddings

Each component comprises six layers, with the encoder and decoder stacks featuring attention and feedforward network sub-layers. The architecture's design facilitates the sequential processing of layers, where each layer's output becomes the input for the next (Kang et al., 2020).

Figure 21: *Transformer Architecture, Comprising an Encoder and Decoder Stack (Vaswani et al., 2017)*



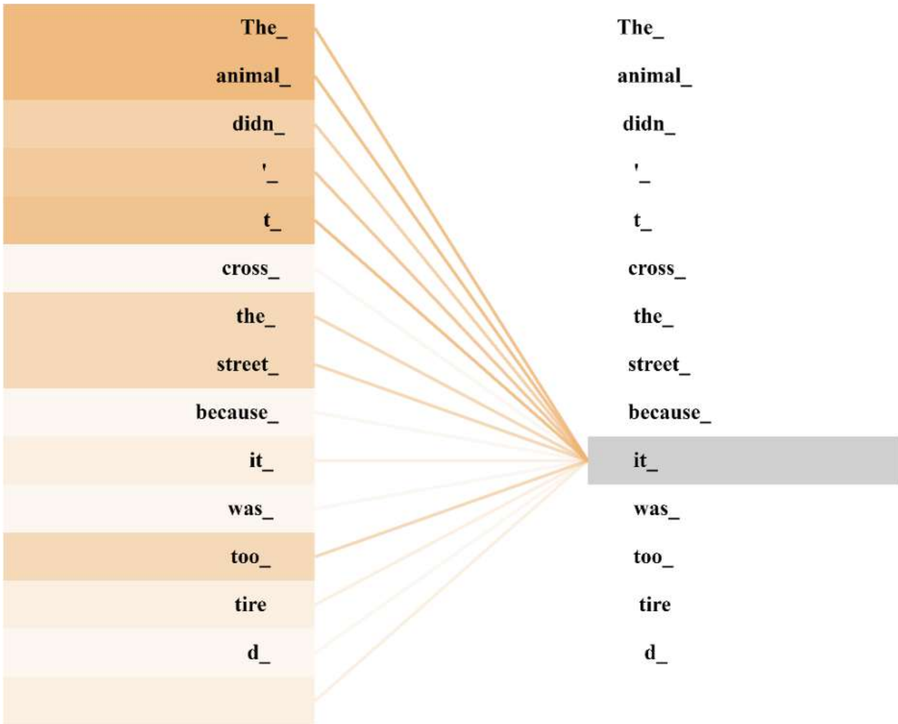
Self-Attention Mechanism

The self-attention mechanism, a pivotal innovation in transformers, has supplanted traditional recurrence-based models, operating on a word-to-word basis and assessing the interrelations of each word with every other word in a sequence. This mechanism calculates dot products between word vectors to discern the most significant relationships, enhancing the model's understanding of word context and sequence (Alammar & Grootendorst, 2024).

Encoder Stack

Within the encoder, each of the six layers follows a uniform structure, consisting of a multi-headed attention mechanism and a fully connected feedforward network. A key feature of these layers is the inclusion of residual connections and layer normalization, ensuring the preservation and stabilization of information throughout the processing stages. The input embeddings are augmented with positional encodings, which is a crucial addition given the absence of recurrence in the transformer model, allowing the network to maintain awareness of word order in sequences (Tunstall et al., 2022; Vaswani et al., 2017)

Figure 22: *Self-Attention Mechanism Identifies the Relationship Between Words of the Input Sequence During the Encoding Process (Jay Alammar, 2019)*



Decoder Stack

The decoder stack, mirroring the encoder's structure, introduces an additional masked multi-head attention mechanism. This mechanism is instrumental in enabling the model to infer based on the preceding context without prematurely accessing future sequence parts. The decoder's layers consist of three sub-layers: masked attention, an attention mechanism linked to the encoder, and a feedforward network. The output of the decoder is transformed into specific words through a linear layer followed by a softmax layer, converting the output vectors into probabilities and selecting the most probable word at each decoding step (Tunstall et al., 2022). Although transformers were initially conceptualized for sequence-to-sequence tasks such as machine translation, their versatility has led to their application in various NLP tasks, using configurations such as auto-encoding, auto-regressive, and full sequence-to-sequence models. A critical aspect of their deployment involves transfer learning, where models are initially pre-trained on extensive text corpora to learn general-purpose language representations, followed by fine-tuning on specific downstream tasks to adapt to particular NLP challenges (Tunstall et al., 2022). The following sections will provide a more detailed exploration of these model types and the intricacies of the transfer learning process in the context of NLP.

The transformer architecture is the base of most generative models. These generative models learn data distributions to create new, similar instances, which is essential in IAI for TLP tasks such as root cause identification in maintenance (Goodfellow et al., 2016).

- **Generative Adversarial Networks:** They consist of generator and discriminator networks trained adversarially to produce realistic data samples (Goodfellow et al., 2016).
- **Variational Autoencoders:** They encode data into a latent space and reconstruct it, useful for generating new data (Kingma & Welling, 2013).
- **Autoregressive Models:** They predict sequence elements based on previous ones, aiding in time-series forecasting.

The most influential transformer-based models with self-attention mechanisms are (Vaswani et al., 2017):

- **Bidirectional Encoder Representations from Transformers (BERT):** It understands context bidirectionally and excels in tasks like text classification (Devlin et al., 2018). In IAI, it extracts information from maintenance logs.
- **Generative Pre-trained Transformer (GPT):** It is designed for text generation and is effective in creating maintenance reports and documentation (Alec Radford et al., 2019)

2.3.3 Automated Planning

Optimizing factories, a cornerstone of production engineering for the past half-century, has traditionally relied on addressing challenges within limited degrees of freedom; however, recent technological advances propose a daydreaming framework for factories that utilize their cognitive capacity for foresighting—that is, anticipating and evaluating future possibilities (Nassehi et al., 2022). By assessing and learning from potential eventualities, these daydreaming factories become antifragile, thriving amid uncertainty without the need for revolutionary reactions to external or internal stimuli. In this context, automated planning emerges as a critical component, enabling factories to perform “what-if” analyses that explore various operational scenarios. By integrating automated planning with IAI, factories can simulate decisions, optimize processes dynamically, and facilitate the co-evolution of people, processes, and products.

Automated planning and learning, a subfield of AI, is concerned with developing intelligent systems that excel in both planning and learning (Russell et al., 2010). These systems can not only generate plans of action (i.e., courses of action) for solving problems but also learn and reason about past experiences, anticipate future challenges, and develop strategies to address them effectively (Kuter, 2009). Typical drawbacks in manufacturing include logistical and social constraints, as well as limitations in infrastructure. A feasible alternative approach to achieving this objective is a phased progression, initiating the process of changing different aspects of operations management (Y. Cohen et al., 2019). From this standpoint, scheduling is a crucial procedure that demands investigation. Parente et al. (2020) provided insight into the scheduling concerns that require attention within the novel I4.0 framework. On the other hand, the objective of production planning and control activities is to assist the organization in aligning manufacturing performance with consumer demand by determining what, how much, and when to produce, purchase, and deliver (Bonney, 2000). Thus, production planning and control enhances the overall value of the manufacturing process. Moreover, production planning and controls must consistently respond to shifting strategic and operational environments, which evolve the opportunities in the supply chain and complex customer demands (Wiendahl et al., 2005). In a nutshell, a dynamic industrial environment requires an evolutionary and integrative strategy regarding operations management, planning, and production. Furthermore, advancements in information and communication technology have greatly improved the performance of the function (Shamsuzzoha et al., 2016). Moreover, the substantial increase in capabilities has paved the way for the development of CPS and the IoT in industrial environments, enabling enhanced production and maintenance processes (Pivoto et al., 2021). Likewise, production planning involves allocating resources to achieve production goals. Production is primarily regarded as an activity that adds the most value to raw materials. A well-planned production timeline allows the manufacturer to meet the quantity and distribution needs of the consumer on time. Moreover, a detailed production timeline involves information including the amount to be produced and task assignment to electronic machines. Additionally, it is designed to achieve high performance on one or more evaluation parameters, including operational expenditures and completion time. Conversely, research has made progress by integrating methodologies such as IAI, mathematical modeling, and optimization algorithms to enhance decision-making systems' effectiveness.

Additionally, maintenance planning plays a role in efficiency as it focuses on activities that ensure the availability and reliability of production assets. Similarly, maintenance planning is viewed as an activity that keeps machines in good working order and helps to increase equipment availability for output (Ansari et al., 2023). In the absence of maintenance, machines deteriorate rapidly, resulting in unexpected malfunctions and excessive downtime (Moghaddam, 2020).

In recent years, the nurse scheduling problem, a persistent challenge in shift planning, has been the subject of extensive research. These studies have focused on finding solutions to the challenges that arise in managing the healthcare workforce. The research aims to create schedules for each employee that guarantee the completion of tasks as per the predetermined plan. This is especially difficult in the healthcare industry due to the presence of a variety of distinct staffing needs on various days and shifts. Jafari (2021) investigated the nurse scheduling challenge through the optimization of nurses' shift preferences while minimizing the overall number of required nurses; consequently, four game-theoretic models were proposed based on the participants' diverse cooperative and competing interactions. Furthermore, a model for mathematical programming was constructed to derive equilibrium strategies.

Dummer et al. (2023) proposed a metaheuristic optimization method that addresses the practical limitations and scope of real-world planning scenarios to plan medical residency training in Austria, presenting the autonomously optimized planning approach for residency training in real-world planning scenarios. The approach was developed and assessed by utilizing a use case and dataset acquired from a medical facility in Austria.

Güler and Geçici (2020) handled the challenge of physician shift scheduling in a healthcare facility during the COVID-19 pandemic by proposing a mixed-integer programming model and constructed a decision support system. In this model, every physician, apart from their regular divisions, was required to provide their services in the COVID-19 departments. Additional time was allocated for leisure activities complying with prolonged work shifts, thus creating a schedule that encouraged greater mental productivity. Furthermore, by implementing suitable workloads, the resultant schedules successfully reduced the susceptibility of the physicians to the virus while simultaneously ensuring uninterrupted access to healthcare services for all departments. This approach was also transferred to metaheuristics using fair scheduling for medical training (Gaal et al., 2023).

F. Liu et al. (2019) proposed a novel decision model that combines predictive maintenance decisions based on forecasting data to minimize total projected costs. The health indicator and dummy age subjected to machine deterioration were taken into consideration in the integrated model. Moreover, a parametric decision-making framework for PdM based on system health was developed by Huynh et al. (2019) to minimize expected long-run maintenance costs. The performance of the integrated solution was compared to the solutions achieved by independently tackling the predictive maintenance planning and production scheduling challenges, with the outcomes demonstrating its effectiveness. In another study, Moghaddam (2020) introduced a new optimization model to identify Pareto-optimal maintenance and substitution schedules for a repairable manufacturing machine, along with discussing a simulation-based optimization approach. The proposed model and solution method have been found to be an effective strategy for resolving maintenance and replacement planning issues in automated manufacturing equipment such as computer numerical control machines.

In another study, K. T. Nguyen and Medjaher (2019) introduced an optimized schedule for security personnel via integer linear programming. The primary objective of this study was to optimize the security personnel's satisfaction by granting them the autonomy to select their desired work schedule and day off while duly considering the limitations imposed by organizational regulations.

In another study, Stock-Williams and Swamy (2019) encompassed three significant aspects of offshore and wind farms. The initial step involved constructing a robust and adaptable metaheuristic optimization model in which modifications to the objective and simulation algorithms required no adjustments to the optimizer. Next, an approachable valuation methodology was established, which utilized historical wind farm data to assess the merits and demerits of maintenance planning methods and to approximate the financial gain on investment resulting from their implementation. The methodology was finally put into practice and evaluated by applying the valuation approach to data.

2.3.4 Automated Task Allocation

Task allocation involves the transfer of responsibilities to resources or agents, which is critical for maximizing operational effectiveness and attaining intended results in diverse fields. As the demand for customized manufacturing solutions increases, modern manufacturing companies are recognizing the significance of small batch production. Nevertheless, this form of variable production, which caters to various product types, incurs additional setup and transportation time. Q. Feng et al. (2023) presented a solution to the problem of flexible task scheduling, which takes setup time, transportation time, and processing into consideration.

Furthermore, it is suggested that the improved GA could serve as a viable approach to reduce the overall duration of makespan, preparation, and transportation (K. T. Nguyen & Medjaher, 2019). The evaluated algorithm uses three methods to generate solutions, which helps improve both the quality and diversity of the initial population. Kohl and Ansari (2023) employed a competence-based maintenance planning approach that combines a genetic algorithm, knowledge graph, and linear programming. Additionally, the research also integrated competence factors into shift scheduling and assignment allocation processes. Furthermore, the study integrated competence factors into the processes of shift scheduling and assignment allocation. In contrast, competence-based maintenance planning strives to optimize human resources (HR) planning while aligning with the objectives of production planning. This integrated approach strategically coordinates maintenance operations with production goals, highlighting the importance of a unified strategy to ensure that maintenance efforts positively contribute to the production process.

In another study, B. M. Nguyen et al. (2019) implemented a novel approach to optimize the task scheduling problem for Bag-of-Tasks applications in a Cloud-Fog environment. This innovative method takes both costs and execution time into account while also allowing users to prioritize either cost efficiency or processing speed. Hosseinioun et al. (2020) proposed a solution for saving energy that entails using dynamic voltage and frequency scaling. They developed an algorithm called IWO CA, which generates valid task sequences. Their experiments demonstrated that this algorithm performs better in terms of energy consumption than existing algorithms. Z. Zhou et al. (2020) introduced an optimization algorithm for scheduling tasks, which combines the genetic algorithm with a strategy. The experimental results clearly indicate that their proposed system outperforms task scheduling algorithms.

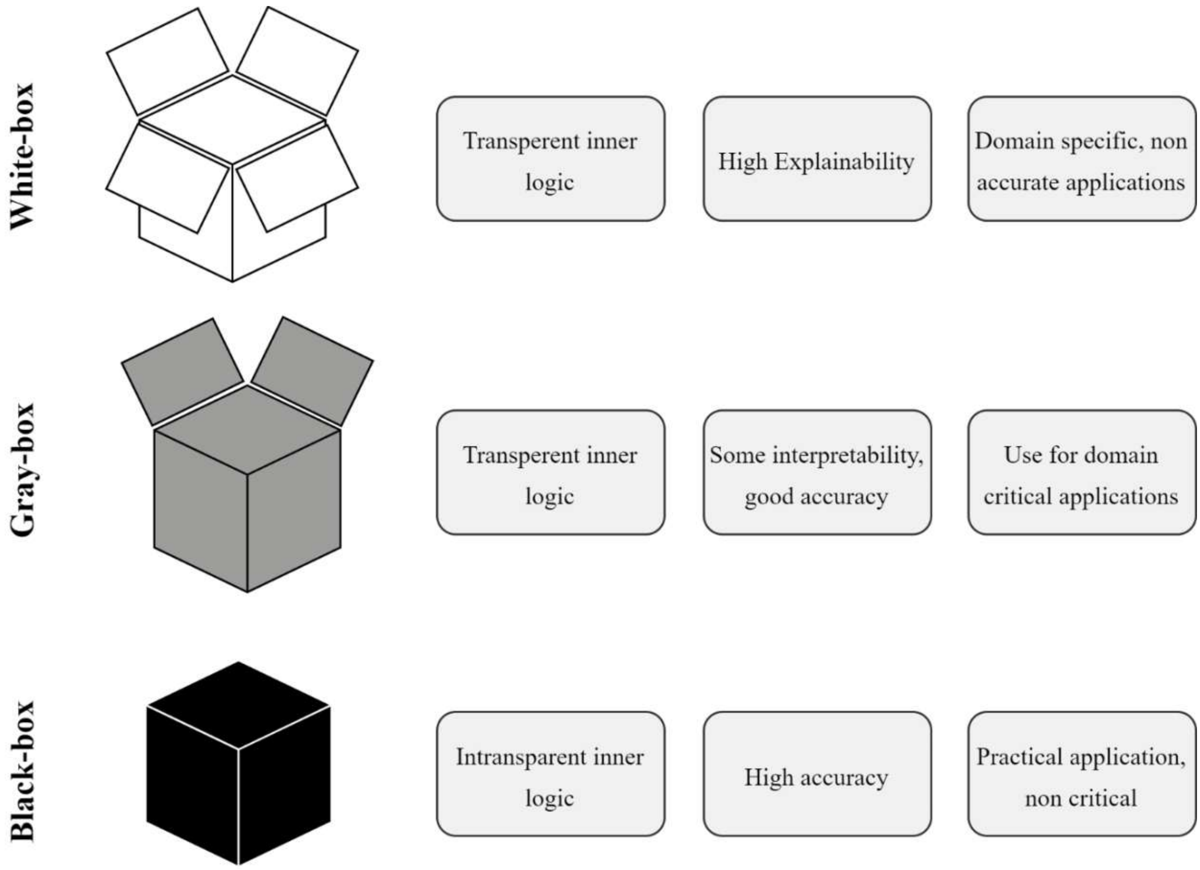
Q. Liu et al. (2019) introduced a method for optimizing a scheduling system that considers both resources and engineering teams by developing a model to help allocate service resources for products. To enhance the system, they combined the nondominated sorting genetic algorithm with enhanced genetic algorithms to incorporate an optimization mechanism. An alternative study conducted by Ehtesham Rasi and Sohanian (2021) aimed to maximize profitability while minimizing pollution emissions and optimizing the utilization of premium raw materials. The study utilized two algorithms—multi-objective particle swarm and multi-objective genetic—to resolve the issue. The major aim of this study was to optimize sustainability performance indicators and incorporate sustainable supplier selection into the design of supply chain networks.

2.4 Interim Conclusion: Opening Up the Research Areas of Competence-Based Maintenance Through Artificial Intelligence

IAI is crucial in the modern maintenance landscape as it enables the generation of plans and actionable recommendations from heterogeneous data sources. By processing and analyzing data from sensors, maintenance reports, manufacturing execution system (MES), and other inputs, IAI facilitates predictive maintenance, optimizes resource allocation, and enhances decision-making processes.

In alignment with the cognitive maintenance philosophy presented, the core of the system must be a white box that can solve specific tasks using flexible black box models, see Figure 1. The white box, comprising a KG, serves as the transparent and interpretable component of the system, leveraging ontologies and Knowledge bases to represent domain knowledge explicitly. Moreover, it provides explainable reasoning and decision-making capabilities, ensuring that maintenance actions are grounded in a well-defined understanding of the domain.

Figure 23: *Differentiation Between White, Gray, and Black-Box AI Models*



The flexible black box models, such as DL algorithms and other advanced ML techniques, contribute adaptability and learning capabilities to the system. While these models may lack inherent interpretability, they excel at processing complex, high-dimensional data and identifying patterns that may not be apparent through traditional analysis. By integrating these black box models with the semantic white box core, the system can utilize their strengths while mitigating their limitations.

Overall, this integration allows for the use of a flexible and scalable gray box model. The gray box approach combines the transparency and explainability of white box models with the flexibility and performance of black box models, enabling the system to adapt to various maintenance scenarios, handle diverse data types, and provide reliable, explainable recommendations. Further, it ensures that the decision-making process is both data-driven and aligned with domain expertise, thereby enhancing trust and acceptance among users.

Furthermore, the gray box model supports scalability by accommodating new data sources, models, and algorithms as they become available. It can also evolve with technological advancements and changing organizational needs, making it a sustainable solution for long-term maintenance planning and operations.

By adopting this gray box approach, organizations can harness the full potential of IAI in maintenance, improving asset availability, reducing downtime, and optimizing resource utilization. It aligns with human-centered design principles by providing maintenance personnel with understandable and actionable insights, tailored to their competencies and the specific context of their tasks. This ensures that even personnel with less extensive qualifications can effectively engage with the maintenance system, addressing skill shortages while maintaining high standards of asset performance and maintenance quality.

2.4.1 Maintenance

This thesis examines the evolution and optimization of maintenance strategies in industrial contexts. Historically, the primary focus was on minimizing direct maintenance costs. However, it has become apparent that the greater cost implications arise from production downtime, shifting the perspective of maintenance from a mere cost center to a critical cost optimization function. This reorientation underscores the importance of strategic maintenance planning to minimize total costs by effectively managing production downtimes.

As previously discussed, PM aims to reduce unplanned downtime through scheduled maintenance activities. This approach hinges on understanding the fault behaviors of components, typically quantified by metrics such as MTBF or MTTR. However, the inherent challenge in PM lies in the potential for premature replacement of components, leading to increased spare part and labor costs and reducing overall efficiency and sustainability.

Conversely, PdM represents a more sophisticated strategy that leverages historical data and advanced modeling techniques to predict equipment faults before they occur. By analyzing trends and patterns, PdM aims to optimize maintenance schedules, thus reducing downtime and extending the life of components. The integration of PdM with Industry 4.0 technologies further enhances its potential, enabling real-time monitoring and data-driven decision-making.

Furthermore, the concept of KBM is examined as a pivotal element of contemporary maintenance practices, with KBM employing domain-specific knowledge, historical data, and expert systems to inform maintenance decisions. The integration of KBM facilitates more precise diagnostics, enhanced decision-making, and the development of more effective maintenance strategies. Additionally, the formalization and systematization of knowledge within an organization facilitate the standardization of maintenance procedures, the improvement of training programs, and the assurance of consistent maintenance quality.

In the broader context of CAS, the maintenance strategies discussed illustrate the progression toward more intelligent and responsive systems. KBM, in particular, aligns with the principles of cognitive assistance by utilizing data analytics and ML to provide anticipatory guidance and support to maintenance technicians and planners. This not only improves operational efficiency but also enhances the adaptability and resilience of industrial systems.

2.4.2 Assistance Systems

Assistance systems, particularly DAS and CAS, are transformative tools in the manufacturing, maintenance, assembly, and logistics sectors. These systems utilize advanced technologies such as AI, ML, and the IoT to provide real-time data analysis, predictive maintenance insights, and personalized operational guidance. The dynamic adaptability of assistance systems to changing workplace conditions represents a significant advancement over static, traditional systems.

In the broader context of cognitive assistance systems, the role of assistance systems involves augmenting human capabilities rather than replacing human labor. By providing real-time feedback, error detection, and corrective suggestions, CASs enhance the efficiency and accuracy of complex tasks. For example, in maintenance operations, CASs can offer real-time diagnostics and troubleshooting support, reducing the time and effort required to resolve issues and preventing potential faults.

Implementing CAS involves several layers of technology, including sensors for data collection, cloud computing for data processing, and AI algorithms for data analysis and decision support. The integration of these technologies creates a robust framework that supports maintenance personnel in their tasks, improves task allocation through competence-based scheduling, and ensures that maintenance activities are performed efficiently and effectively. Furthermore, the human-centered design of CAS ensures that these systems are user-friendly and intuitive, promoting higher acceptance and usage among maintenance staff. This design philosophy is crucial for successfully deploying CAS in industrial environments, where user acceptance can significantly influence the overall effectiveness of the system. By focusing on enhancing human capabilities and providing contextualized support, CASs contribute to more efficient and reliable maintenance operations.

2.4.3 Industrial Artificial Intelligence

IAI represents a convergence of AI and industrial engineering aimed at enhancing the cognitive capabilities of industrial systems. The application of IAI in maintenance is particularly promising, as it enables the development of intelligent systems that can perceive, process, and autonomously solve complex problems. The present study explored various AI and ML techniques that can mimic human cognitive functions, thereby improving maintenance processes.

The practical applications of IAI in maintenance entail the ability to analyze vast volumes of data to identify patterns, predict faults, optimize maintenance schedules, and provide maintenance technicians with real-time insights and guidance. These capabilities not only reduce maintenance costs and downtime but also enhance the overall reliability and safety of industrial operations. Moreover, the study emphasized the importance of a robust implementation framework for IAI systems, which includes ensuring data quality, integrating AI with existing industrial systems, and providing adequate training for maintenance personnel. The successful deployment of IAI systems in industrial environments necessitates a multidisciplinary approach that encompasses data science, industrial engineering, and human factors.

Within the framework of CAS, IAI plays a pivotal role in transforming traditional maintenance approaches into more intelligent and adaptive processes. As outlined in Section 1.3, ARCHIE aims to provide an AI-enhanced framework designed for maintenance, combining data from physical sensors and unstructured data sources to provide a comprehensive view of the maintenance environment. This holistic approach enables the system to make autonomous, proactive decisions that improve maintenance processes, such as shift planning and fault identification in line with KBM principles.

In conclusion, integrating IAI within CAS for KBM represents a significant advancement in industrial maintenance. By enhancing the cognitive capabilities of maintenance systems, IAI enables more efficient, effective, and reliable maintenance operations. Ongoing research and development in this field are crucial for realizing the full potential of AI in transforming industrial maintenance practices.

3 State of the Art

This chapter presents a systematic literature review inspired by the literature research strategy developed by O'Donovan et al. (2015) and, subsequently, Zonta et al. (2020) and Peres et al. (2020). The aim of the selected methodology was to develop a reproducible search string. This string, which encompassed scientific literature on CAS and its applications in industry, manufacturing, and particularly maintenance, was then queried in the most commonly utilized literature databases. The results from the databases were filtered based on the defined criteria, after which the duplicates were exchanged and then merged. These merged literature sources were subjected to bibliometric analysis, broken down into their topics in a morphology to identify a research gap. Therefore, a search string and databases were defined. As a starting database, the constructed search string, seen in Figure 21, was adapted to Google Scholar, IEEE Xplore, ScienceDirect, and Web of Science on March 31, 2024, with a filter considering the time frame of 2015–2024, with 2015 marking the beginning of DL-based TLP, 2018 marking the rise of transformers, and 2022 marking the release of GPT.

The search string used for this review was constructed to capture relevant scholarly works that explore this intersection, ensuring comprehensive inclusion of studies relevant to the research questions, as seen in Figure 21.

Figure 24: *Structured query representation for produced fo the literature review*

(("industry 4.0" OR "smart manufacturing" OR "cognitive system") AND ("predictive maintenance" OR "PdM" OR "prescriptive maintenance" OR "knowledge-based maintenance") AND ("Artificial Intelligence" OR "AI" OR "machine learning" OR "machine learning technique") AND ("model" OR "method" OR "architecture") AND (predict*))

The main research question ("What information can be extracted from textual data to improve the efficiency and effectiveness of maintenance planning?") was addressed through terms such as "knowledge-based maintenance" and "artificial intelligence," which focus on methods for data extraction and utilization. Predictive models and ML techniques are essential for enhancing maintenance planning through data analysis. Sub-research question 1 ("How can competence-based shift planning and task assignment be integrated in maintenance planning?") was supported by including terms such as "cognitive system" and "smart manufacturing," ensuring the inclusion of literature on AI-driven task allocation and competence-based planning.

To address sub-research question 2 ("How can maintenance technicians and planners be supported with background information and recommended solutions to existing machine faults?"), the search string was expanded to include "predictive maintenance" and "prescriptive maintenance." This approach enabled the identification of studies that addressed the issue of providing proactive solutions and background information, which can assist maintenance technicians with AI-generated recommendations. Sub-research question 3 ("How can error-free and time-saving documentation of maintenance reports be achieved in industrial practice?") was addressed by terms such as "knowledge-based maintenance" and "machine learning technique," which are key to automating and improving the accuracy of maintenance documentation, thereby reducing errors and increasing efficiency. Finally, sub-research question 4 ("How can an AI-enhanced cognitive assistance system in industrial maintenance be developed and implemented in an industrial setting to effectively ensure its seamless integration and maximization of its potential benefits for maintenance technicians and planners?") was supported by the inclusion of the terms "model," "method," and "architecture," ensuring that relevant studies on developing and implementing cognitive systems were captured. This comprehensive approach ensured that the review encompassed technological contexts, advanced maintenance strategies, AI applications, and practical implementation aspects, thereby directly addressing the research questions and objectives of the thesis.

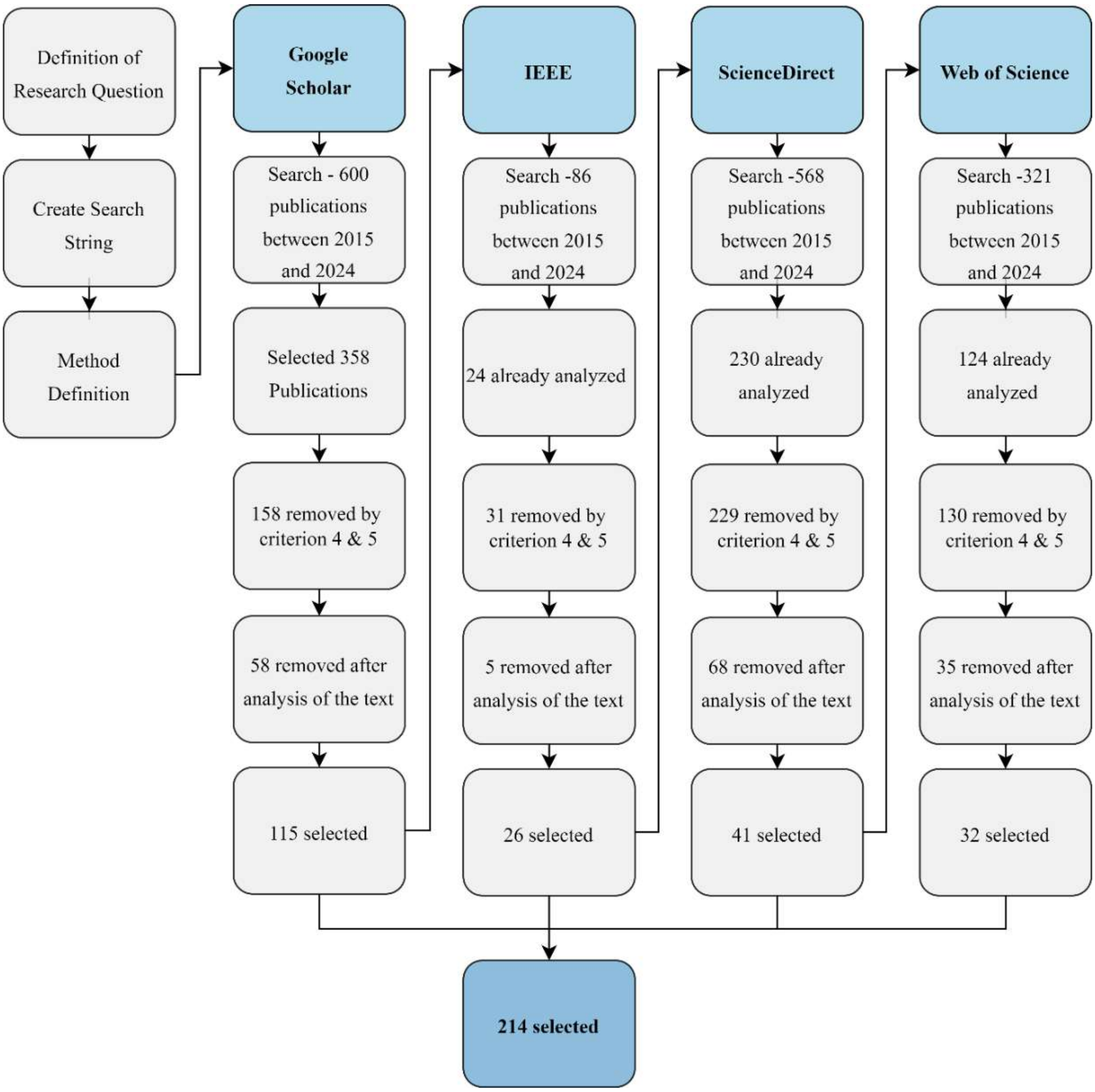
The screening of the selected documents was based on five exclusion criteria, as outlined in Table 2. It should be noted that criterion 1 was selected as the language of scientific communication to prevent any geographical bias. Additionally, criterion 4 was chosen to ensure that no roadmap publications were included, as these predominantly pertain to subjects that have yet to be implemented.

Table 2: *Screening Criteria*

Criteria	Description
Criteria 1	Filter from 2015 to 2024
Criteria 2	Publication must be in English
Criteria 3	Publication must include the search terms Industry 4.0, Smart Manufacturing, Predictive Maintenance, Artificial Intelligence, or Machine Learning
Criteria 4	Publication must not be solely a review or roadmap
Criteria 5	Publications must address CAS in maintenance as a model, method, architecture, approach, or methodology

The methodology for literature selection, as delineated in Figure 25, encompassed a multi-stage process across various academic databases. Utilizing a modified algorithm proposed by Wittmann (2017), an initial corpus of scholarly articles was identified on Google Scholar, which served as the primary database for this research. A preliminary threshold was established, limiting the scope to the top 600 most relevant search results. Following this initial selection, a rigorous screening process was implemented, which involved applying a predetermined fourth criterion, resulting in the exclusion of 58 articles. An additional four articles were eliminated following a detailed text analysis, ensuring the relevance and quality of the remaining literature. The search methodology was then replicated on IEEE Xplore, ScienceDirect, and Web of Science databases, employing the same selection criteria and procedures. During this phase, any duplicates previously identified and analyzed in the Google Scholar dataset were meticulously removed to avoid redundancy.

Figure 25: *Screening Process of Literature Research*



The final phase of the selection process entailed a comprehensive analysis of the abstracts, keywords, conclusions, and content sections of the remaining articles. This thorough examination facilitated the distillation of the literature to a final set of 214 articles, which were deemed most relevant and informative for the research objectives at hand.

These papers were analyzed to build a morphology describing CAS in maintenance, based on their content, as seen in Appendix I. Table 3 lists a selection of the used articles sorted by year, along with types of publications, publishers, and conference or journal names.

Table 3: *Sample of Selected Articles*

Article	Type	Periodical
Bagheri et al. (2015)	Journal Article	Advances in Service-Oriented and Cloud Computing
Kuo et al. (2017)	Journal Article	Journal of Systems Architecture
Li and Wang (2017)	Journal Article	Advances in Manufacturing
Butte et al. (2018)	Journal Article	Computers & Industrial Engineering
Candanedo et al. (2018)	Book Section	Knowledge Management in Organizations
Cho et al. (2018)	Book Section	Advances in Production Management Systems: Smart Manufacturing for Industry 4.0
Mashhadi et al. (2018)	Journal Article	Procedia Manufacturing
Schmidt and Wang (2018)	Journal Article	Procedia Manufacturing
Sezer et al. (2018)	Conference Paper	IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC)
Strauß et al. (2018)	Conference Paper	IEEE International Conference on Big Data
Bergmann and Klein (2019)	Conference Paper	Proceedings of the 16th International Conference on Informatics in Control, Automation and Robotics
Brueno and Vita (2019)	Conference Paper	IEEE International Conference on Smart Computing
Calabrese et al. (2019)	Journal Article	Procedia Manufacturing
Cao et al. (2019)	Journal Article	Procedia Computer Science
Lee et al. (2019)	Journal Article	Procedia CIRP
Lindemann et al. (2019)	Journal Article	Procedia CIRP
Aydemir and Acar (2020)	Journal Article	Journal of Manufacturing Systems
Calabrese et al. (2020)	Journal Article	Information

Article	Type	Periodical
Caroff et al. (2020)	Journal Article	Procedia Manufacturing
Chen et al. (2020)	Journal Article	IEEE Access
Fila et al. (2020)	Journal Article	Procedia Computer Science
Kiangala and Wang (2020)	Journal Article	IEEE Access
Kozłowski et al. (2020)	Journal Article	Expert Systems with Applications
Luo et al. (2020)	Journal Article	Robotics and Computer-Integrated Manufacturing
Ansari et al. (2021)	Journal Article	CIRP Annals
Kohl, Ansari, and Sihn (2021)	Conference Paper	WGAB Conference
Bourezza and Mousrij (2021)	Journal Article	Advances in Intelligent Systems and Computing
ElMaraghy et al. (2021)	Journal Article	CIRP Annals
Kohl, Ansari, and Sihn (2021)	Conference Paper	WGAB Conference on Competence Development and learning assistance
Curry et al. (2022)	Book Chapter	Technologies and Applications for Big Data Value
Quandt et al. (2022)	Conference Paper	Procedia CIRP
D'Amico et al. (2022)	Conference Paper	CIRP Journal of Manufacturing Science and Technology
Liu et al. (2022)	Conference Paper	Robotics and Computer-Integrated Manufacturing
Liu et al. (2022)	Conference Paper	IEEE Transactions on Industrial Informatics
Haase et al. (2022)	Conference Paper	Industrie 4.0 Management
Balas et al. (2022)	Conference Paper	Communications in Computer and Information Science
Ansari et al. (2022)	Book Chapter	Supply Network Dynamics and Control
Li et al. (2022)	Book Chapter	Procedia CIRP
Gaal et al. (2023)	Conference Paper	WGAB Conference
Li et al. (2023)	Journal Article	Robotics and Computer-Integrated Manufacturing
Ansari et al. (2023)	Journal Article	CIRP Annals
Kohl et al. (2023)	Book Chapter	Procedia CIRP
Jan et al. (2023)	Conference Paper	Expert Systems with Applications
Zhang et al. (2023)	Conference Paper	Advanced Engineering Informatics
Teoh et al. (2023)	Journal Article	IEEE Internet of Things Journal
Xie et al. (2023)	Book Chapter	Big Data and Cognitive Computing

Article	Type	Periodical
Mallioris et al. (2024)	Journal Article	CIRP Journal of Manufacturing Science and Technology
Jaenal et al. (2024)	Journal Article	Engineering Applications of Artificial Intelligence
Chakroun et al. (2024)	Journal Article	Journal of Intelligent Manufacturing
Justus et al. (2024)	Journal Article	International Journal of System Assurance Engineering and Management
Mukherjee et al. (2024)	Journal Article	Computers & Industrial Engineering
Zheng et al. (2024)	Journal Article	Journal of Intelligent Manufacturing

Figure 26 presents the distribution of article types. The majority, with over two-thirds, were published as journal articles, whereas only 46 documents out of 220 were conference papers. Additionally, only 21 were selected from book sections.

Figure 26: *Distribution of Articles by Type*

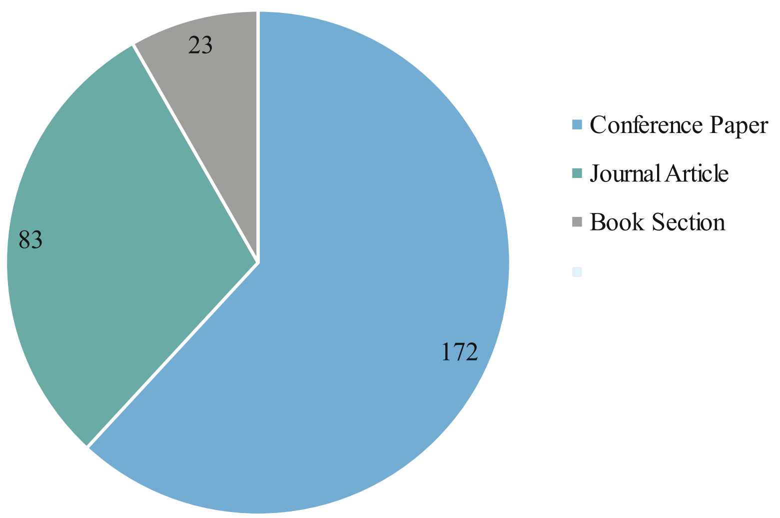
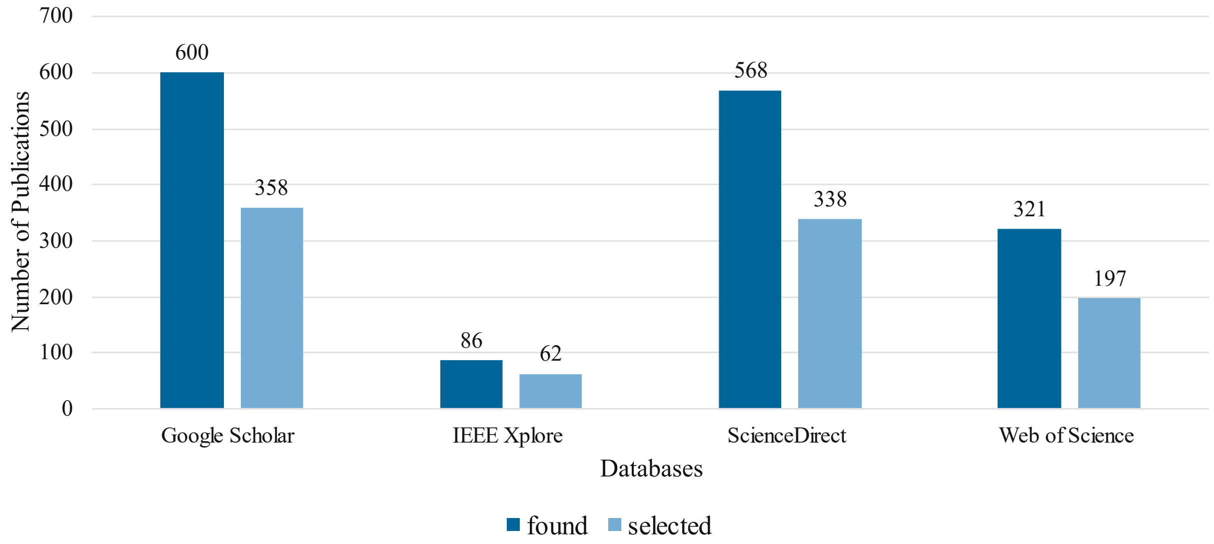


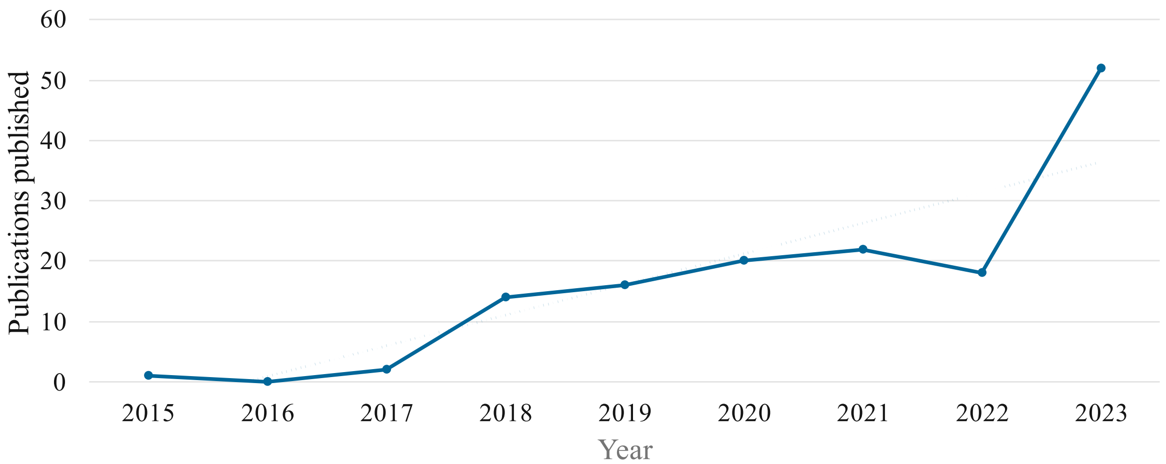
Figure 27 illustrates the number of publications in the databases after applying the search string. On Google Scholar, there was a cut-off at 600 publications, from which 115 publications were selected after applying the extraction criteria. In IEEE, only 86 publications were found, and 26 publications were used. On the other hand, ScienceDirect had a remarkably high number with 568 publications. As for the other two databases, only 41—a relatively low amount—were usable for this thesis. Similarly, on Web of Science, 321 publications were found but only 32 were selected after applying the criteria. In conclusion, the search string performed the best on the IEEE database and the worst on ScienceDirect. In addition to analyzing highly relevant publications, a reverse search was also conducted to identify potentially related publications that had not been previously considered.

Figure 27: *Number of Publications in the Databases Using the Extraction Criteria*



The annual evolution of the CAS topic from 2015 to 2024, including redundant papers, can be clearly observed in Figure 28. Even though it is only a short period, the graph and the trend line demonstrate that CAS, as a relevant maintenance approach in the 21st century, is gaining traction, indicating a tendency toward even greater attention in the future. Moreover, it also signifies that substantial research is being conducted on this specific topic.

Figure 28: *Distribution and Tendency of Publications Per Year (2024 Excluded)*



Existing taxonomies, frameworks, and literature reviews have addressed the implementation of emerging technologies in the maintenance process and described the extent of PdM; however, they generally do not exclusively focus on the CAS approach in manufacturing.

Peres et al. (2020) conducted a systematic review of current IAI and its application areas in manufacturing, such as process optimization, quality control, human-robot collaboration, PdM, and relevant cognitive approaches in Industry 4.0. This can be seen as a steppingstone for Cheng et al. (2022) and Runji et al. (2023), whose works discussed visual analytics and its sub AI methods for maintenance and manufacturing. While these reviews offer a relevant overview of emerging technologies and AI application possibilities, they only mentioned cognitive approaches in maintenance in passing.

In contrast, Jan et al. (2023), Dogan and Birant (2021) and Iftikhar et al. (2023) carried out comprehensive literature reviews on the ML and DL techniques and tools used in a manufacturing context. They discussed solutions in manufacturing based on their tasks, algorithms, and learning types, laying the groundwork for cognitive systems. Furthermore, they explained the steps of the knowledge discovery in databases. These studies provided a good overview of the prevalent ML techniques and how data can be obtained in manufacturing. However, they did not explicitly deal with CAS.

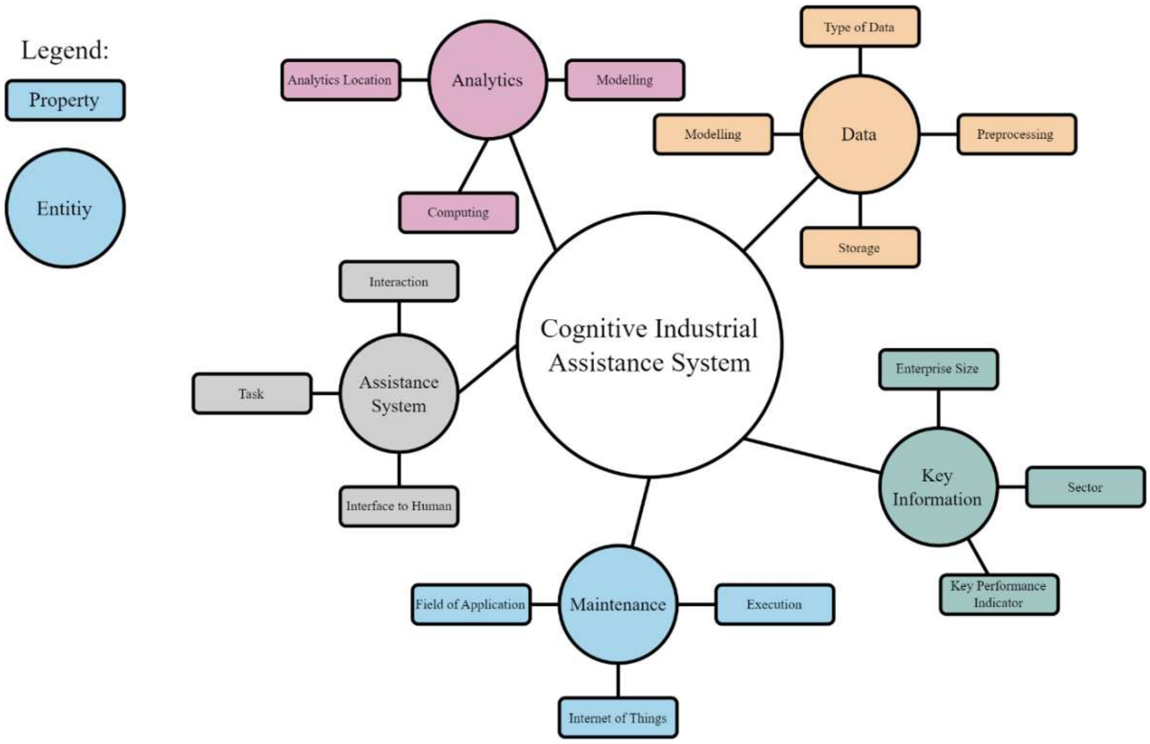
Moreover, T. P. Carvalho et al. (2019), Siraskar et al. (2023), Divya et al. (2023) and Scaife (2024) focused on the implemented ML methods in PdM, including cognitive approaches, describing which equipment is subjected to PdM techniques and what data are used to apply PdM. While these papers focused on specific steps, they did not consider the CAIS approach as a complete process.

Zonta et al. (2020) developed a PdM taxonomy that deals with architectures for PdM approaches in Industry 4.0, which involved an in-depth categorization of these approaches. Taking KBM approaches into account (Cao et al., 2022; Xiaoqiao Wang et al., 2023) and the fundamental work of ElMaraghy et al. (2021) on the categorization of cognitive approaches in manufacturing, a detailed picture of the CAS landscape can be drawn with the help of Zonta (Zonta et al., 2020), the relevant work mentioned above, and the results of the literature analysis, as seen in Table 3. A review of the literature reveals that IAI, particularly ML, has attained a high level of relevance and maturity. Moreover, the challenging role of combining data sources has also been illustrated by several published works. It is noteworthy that the literature search has demonstrated the increasing importance of combining heterogeneous data sources, especially with semantic technologies as a white box approach in the field of IAI. However, these are still treated marginally in several literature reviews. Only a few publications use semantic technologies as a central element, whereas black box approaches, often used in ML, usually form the core of the architectures or methods used.

3.1 Dimensions of the Morphology

In this systematic state-of-the-art review, a wide array of literature was analyzed to construct a comprehensive morphology for CAIS, specifically tailored for maintenance applications, as seen in Figure 26. The synthesis of these findings provided the foundations for a cognitive maintenance framework that integrates the latest advancements and theoretical underpinnings in the field. This morphology encapsulated key aspects of cognitive systems, leveraging the potential of AI and digital technologies to advance maintenance planning and operations in industrial settings. The resulting structure serves as a blueprint for developing and implementing cognitive assistance systems that can significantly enhance maintenance efficiency, accuracy, and reliability (Stricker, 2021).

Figure 29: *Concept of CAS Morphology*



Key Performance Indicators

KPIs are an indispensable component of the maintenance control process; consequently, they must be considered when utilizing IAI technologies. They are an intrinsic element of the business understanding phase within the CRISP-DM framework and serve as quantifiable metrics for maintenance performance (Wirth & Hipp, 2000). The development of KPIs emphasizes ease of interpretation and the ability to trace the derivation of these metrics (Matyas, 2022). KPIs not only facilitate the identification of cost drivers but also elucidate cause-and-effect relationships. Their timely availability is crucial for both internal and external benchmarking, forming the foundation for setting planned values and implementing corrective actions.

Task Definition

Dogan and Birant (2021) categorized tasks within ML and DM in a manufacturing and PdM context as scheduling, monitoring, forecasting, and prediction. A significant proportion of industrial applications focus on fault forecasting (33%) and RUL prediction (35%). Other approaches include standard fault triggers such as anomaly detection (7%), drift detection (Zenisek et al., 2019), and condition monitoring (Candanedo et al., 2018).

Sectoral Application of CAS

T. P. Carvalho et al. (2019) and Zonta et al. (2020) identified diverse sectors for PdM implementation, notably manufacturing (70%), transportation (15%), and energy (10%), along with healthcare and the food industry.

Enterprise Size

The scale of data flow and enterprise size are pivotal in tailoring CAIS systems. Sezer et al. (2018) and Paranitharan et al. (2024) advocated for a low-cost system specifically designed for SMEs, suggesting a categorization of companies into small, medium, and large enterprises for CAIS application.

Implementation Requirements

CAIS implementation necessitates sensor data collection, which should be both cost-effective and straightforward, especially in the planning phase. This includes digitizing machine centers or integrating them into cyber-physical production systems (CPPS). For older machinery, retrofitting approaches are recommended to facilitate Industry 4.0 compatibility without incurring extensive reengineering costs (Hesser & Markert, 2019; Schulte et al., 2023)

Interactions in Manufacturing

Zonta et al. (2020), Xiaoqiao Wang et al. (2023) and ElMaraghy et al. (2021) discussed interactions in manufacturing involving products and industrial services, particularly in the context of CPS and service innovation. The data flow in CAIS approaches stems from interactions with workers, products, processes, tools, and machinery.

Data Storage Solutions

The shop floor generates substantial data that requires storage in various formats, including unstructured, structured, or relational databases. Data warehouse solutions of CAIS (Chen et al., 2020; Z. Li et al., 2017), DL for unstructured data (Sezer et al., 2018; Strauss et al., 2018; Yu et al., 2020), and KGs (Ansari et al., 2023; Cao et al., 2022; Kohl & Ansari, 2023; Rosati et al., 2023) are employed to ensure timely, accurate, and appropriately located data availability for CAIS.

Data Preprocessing

Data preprocessing is vital in industrial maintenance, particularly for predictive and preventive strategies. Large volumes of sensor and machinery data in industrial contexts often contain noise, incompleteness, and inconsistencies, potentially leading to erroneous predictions and suboptimal maintenance actions (Al-amri et al., 2021). Therefore, implementing data preprocessing techniques, including data cleaning, normalization, outlier detection, and feature extraction, is imperative to enhance data quality and reliability (Dogan & Birant, 2021). Data preprocessing is a critical step in converting raw data into actionable insights, which is essential for successfully applying advanced maintenance strategies in industrial settings.

Analytic Location

Fog computing or edge computing is an emerging technology that provides cloud-like services to the edge of the IoT infrastructure (Iftikhar et al., 2023). Factory floors produce a large volume of data, and, with fog computing, these data can be analyzed quickly, accelerating the decision-making and reaction process for CAS (Foukalas, 2020; Islam et al., 2024). With the emergence of Big Data and the rise of CPPS (Partovi et al., 2021), cloud computing paradigm can be used as a hosting platform for DL and ML algorithms. All data processing and analysis algorithms can be located in the cloud server for CAS (Fila et al., 2020).

Data Modeling

As Zonta et al. (2020) stated, data modeling can differ among knowledge-based, data-driven, and physical model–based approaches. Knowledge-based models build upon rules, facts, or cases gathered over years of operation and maintenance of the technical system (Molęda et al., 2023; Montero Jimenez et al., 2020). Data-driven approaches utilize large amounts of data generated daily from technical systems. Physical model–based approaches "use the laws of physics to assess the degradation of components" (Montero Jimenez et al., 2020). Knowledge-based approaches are the basis for CAS (A. Carvalho et al., 2020; Ghobakhloo et al., 2023; Xia et al., 2022). It can be concluded that most papers focused on a black or gray box machine learning approach without leveraging the possibilities of semantic modeling (e.g., Dynamic Bayesian Networks, KG) for interlinked data and resulting IAI applications.

Error Metric

Error metrics can measure the quality of an algorithm, i.e., the predictive performance of a regression model in terms of the mean deviation of its estimation from the actual values. The lower the error values, the higher the accuracy in making predictions (Traini et al., 2019). The purpose is to compare different ML algorithms and their performances. Various error metrics exist, such as regression-related metrics (25% of the use cases), classification-related metrics (53% of the use cases), ranking-related metrics, and statistical metrics. The performance results can be visualized in the confusion matrix, illustrating prediction accuracy and error (Lee et al., 2018; K. T. Nguyen & Medjaher, 2019; Zenisek et al., 2019).

Data Computing

As the data computing process is cost-intensive and energy-consuming, the availability of model results should be determined (Teoh et al., 2023). Offline analysis focuses more on accuracy and considers the entire dataset when building the model, while online approaches prioritize real-time analysis (Calabrese et al., 2020)

Data Origin

Cognitive systems are based on building prediction models constructed from historical data using algorithms and then applying this approach in real-time data streams to identify anomalies and possible faults and generally estimate future behavior (Apiletti et al., 2018; T. P. Carvalho et al., 2019; Paolanti et al., 2018; Rivas et al., 2020)

Data Acquisition

In maintenance, the acquisition of data, particularly sensor and text data, is crucial for predictive maintenance strategies. Sensor data, sourced from various devices within machinery, offer real-time insights into equipment performance and facilitates early fault detection (Al-amri et al., 2021). These sensors, which include temperature and vibration detectors, produce structured data for continuous monitoring and anomaly analysis (Müller et al., 2020). Additionally, text data from maintenance logs, operator reports, and manuals provide a wealth of unstructured information. Advanced text mining techniques applied to these data can uncover patterns and insights beyond what sensor data alone reveals (Biegel et al., 2022). Most studies have focused on singular data sources, missing the diverse data landscape that modern production environments offer.

Type of Data

Time series data is an outcome of repeated measurements over time, providing a database for prognostics and forecasts. Overall, 88% of the analyzed use cases base their approaches on time series data and 8% combine time series and numerical data, representing values expressed in numbers. As manufacturing data can be classified into structured, semi-structured, and unstructured, manufacturers have valued structured data storage in classical relational databases. With advancements in data management techniques, unstructured data management has become increasingly attractive (Gao et al., 2020). As a result, data are not stored in a structured database format; rather, they are most likely to be in visual, textual, or audio form in manufacturing use cases. For unstructured data such as images with many dimensions (e.g., pixels), there are no semantically meaningful data attributes, and dimension reduction methods such as t-distributed stochastic neighbor embedding are used (Gao et al., 2020). Furthermore, DL algorithms such as convolutional neural networks utilize visual formats like encoded images to perform classification tasks (Kiangala & Wang, 2018). Ansari et al. (2021) used text mining in the CAIS context to extract further information from maintenance reports and use it as an additional tool for downtime detection. Moreover, text mining is a valuable tool for identifying patterns, categorizing and clustering unclassified data, and recognize similarities between texts. The knowledge gained can expedite maintenance services by supporting the planning and decision-making processes. It can be observed that current research has primarily focused on singular types of data, with multi-modal data being scarcely discussed in the context of maintenance, especially maintenance planning and execution.

Field of Application

The field of application can either be domain-specific or adopt a generic approach. It is often necessary to build a framework that can be widely implemented across various systems and not solely application-specific (K. T. Nguyen & Medjaher, 2019). Overall, 77% of the analyzed use cases can be implemented into multiple physical systems with slight or no additional adaptation.

Maintenance Execution

The growing importance of maintenance management has generated increasing interest in developing and implementing efficient maintenance strategies and decision processes, which humans alone cannot deliver (Ansari et al., 2019; Biedermann, 2016b). Cachada et al. (2018) introduced the architecture of an intelligent PdM system that supports technicians during maintenance interventions by providing guided intelligent decision support. Furthermore, the decision support system is extended with augmented reality technologies to enhance the interaction between humans and machines (Cachada et al., 2018) or augmented by cognitive aspects (Kohl, Ansari, & Sihn, 2021). In addition, physical assistance systems aim to relieve employees physically in terms of an improvement in force diversion, force application, force amplification, stabilization, support, advances in ergonomics, or the precision of work execution (Schlund et al., 2018). Current support systems for maintenance execution have addressed singular tasks but have not aimed to combine different information sources for holistic support (e.g., planning, execution, and spare parts).

Human Interface

As mentioned above, collaboration between humans and machines is an essential advancement for maintenance actions (Bhattacharya et al., 2023). The information and feedback for the human can be mechanical, acoustical, or optical, with optical communication being the most common for human operators (C. Li et al., 2023; Ranz et al., 2018). With digital networking in the context of assistance systems, employees can utilize emerging technologies to monitor work activities, manage time, and distribute activities as cognitive relief. As an organizational relief, better monitoring of system states, work sequences, processes, and dissemination of knowledge can be achieved (Schlund et al., 2018).

Maintenance Architecture

With the ever-increasing production speed and the merging of several production areas, the maintenance sector can provide vital information to a supply chain, both intra-company or even in cross-company cases (Ansari et al., 2019). Monostori et al. (2016) developed such an architecture to enable cross-company information exchange and a high level of transformation transparency in the supply chain, which can be used to improve maintenance resilience (Ansari & Kohl, 2022). It can be concluded that a generalizable cognitive maintenance architecture is currently lacking.

Maintenance Organization

Most CAS approaches are designed for use in a clear industrial use case with in-house infrastructure and computing capacity, often defined as in-house PdM (Canciglieri Junior et al., 2022). Another approach is a cloud-based PHM, using which, the company no longer owns the infrastructure as its supplier makes it available. The cloud can serve as a hosting platform for autonomous data mining and cognitive learning algorithms, providing appropriate and efficient

RUL prognosis as a service. Accessibility to the cloud can be realized with a server or smart device connected via a user interface as a CAIS application (Fila et al., 2020; Foukalas, 2020).

3.1.1 Morphological Box

The developed morphological box can be seen in Figure 30. This includes all the dimensions listed in Section 3.1, their characteristics, and further detailed breakdowns, such as in the area of data modeling, as seen in Figure 32, to allow for more precise distinctions. The morphological box was used to develop a cognitive maintenance framework and clearly categorize the special features of cognitive assistance systems and their necessary characteristics in maintenance.

Figure 30: *Main Morphology for CAIS*

Knowledge-based Maintenance							
KPIs	Maintenance Cost		Planning Quality		Production Efficiency		
	Maintenance Intensity	Maintenance Cost Rate	Failure Time Share	Maintenance Quote	Availability	Performance	Quality
Task	Scheduling	Monitoring	Fault Forecast	RUL Prediction	Downtime Forecast		
Sector	Energy	Manufactiuring	Health Care	Food Industry	Transport		
Company Size	Small Enterprises	Medium Enterprises	Large Enterprises				
Implementation	Retrofitted PS	Digitized	CPPS				
Interactions	Products	Processes	Tools	Machines			
Data Storage	Data Warehouse	Data Lake	Knowledge Graph				
Data Preprocessing	Data Preparation	Feature Engineering					
Analytic Location	Edge/Fog Computing	Cloud Computing	Digital Twin	On-Premises			
Data Modeling	Data Driven	Physical Model based	Knowledge Based	Hybrid			
Data Computing	Online	Offline					
Data Origin	Historical	Real-Time					
Data Acquisition	Physical Sensor	Virtual Sensor	Simulated Sensor				
Type of Data	Structured		Unstructured				
	Numerical	Time Series	Visual	Textual	Audio		
Maintenance Utilization	Application-Specific	Generic Approach					
Maintenance Task Execution	Human	Machine	Human-Machine Interaction				
Interface to Human	Mechanical	Acoustical	Optical				
Maintenance Communication	Cross-Company	Intra-Company					
Maintenance Organisation	in-house PdM	PdM as a Service					

Figure 31 provides specific input on the individual data preprocessing steps, including 10-fold cross-validation to help benchmark the ML techniques.

Figure 31: *Sub Morphology Data Preprocessing*

Data Preprocessing					
Data Preparation	Data Cleaning	Data Parsing	Data Smoothing	Data Imputation	Data Sampling
Feature Engineering	Feature Extraction				
	Supervised	Unsupervised			
	Linear Discriminant Analysis (LDA)	Autoencoder	Principal Component Analysis (PCA)		
	Feature Selection				
	Filter Methods	Wrapper Methods	Embedded Methods		
	Feature Normalization				
	Feature Standardization	Feature Scaling normalization	Discretization		
	Feature Encoding				
	One-Hot Encoding	Label Encoding			
	Data Partitioning				
	No Splitting	Train-Test	Train-Validation-Test	K-flod cross Validation	

Figure 31 illustrates the data modeling techniques used, such as multi-layer perceptron, case-based reasoning, KG, and their error metrics (e.g., mean square error and root mean square error).

Figure 32: *Sub Morphology (Modified) Data Modeling*

Data Modelling									
Data Driven	Supervised Machine Learning								
	Neural Network			Classification					
	ANN	RNN	CNN	Gradient Boosting	SVM	Bayesian Belief Network	Logestic Regression	Decision tree	
	MLP	Gatewd Recurrent Unit	LSTM						
	Unsupervised Machine Learing								
	Anomaly Detection/Denoising	Clustering		Dimensionality Reduction Algorithm					
	Autoencoder	Gaussian Mixture	K-means Clustering	PCA	t-SNE				
Physical Model Based	HMM	Bayesian Network	Filters	Statistical Analysis					
			Hidden Gamma Process partical- Filter	ARIMA	Hotelling's T2 and SPE statistics	Weibull			
Knowledge base	Fuzzy Logic	CBR	Knowledge Graphic						
Error Metric	Regression Related Matrics				Classification Related Metrics				
	MSC	RMSE	MAE	RSE	Classification Accuracy	Precision	Recall	F1-Score	ROC-Curve

Figure 33 presents the data acquisition step and the importance of distinguishing between physical and virtual sensors. For physical sensors, data from machines are collected. In the digital shift book, the information consists of machine fault reports and regular preventive maintenance reports.

Figure 33: *Sub Morphology Data Acquisition*

Data Acquisition							
Physical Sensor	Parameter						
	Vibration	Acceleration	Temperature	Force	Current	Pressure	Other Types
	Data Collection						
	Automatic	Semi-Automatic	Manual				
	Approaches for data Generation						
	Synthetical	Empirical					
	Communication						
	Wireless	Non-Wireless					
	Output Signal						
	Analog	Digital					
	Type of Sensor						
	Regular Sensor	Smart Sensor					
Virtual Sensor	Origin						
	ERP	Digital Twin	Shift Book	MES	Expert Knowledge	Others	
	Data Collection						
	Automatic	Semi-Automatic	Manual				
Simulated Data	Approaches for Data Generation						
	Fully Synthetical	Synthetical Based on Pervious Data	Synthetical based on a virtual simulation model	Based on a simplified Physical Model			

3.2 Research Gap

Based on the state-of-the-art review and accompanying methodology, a research gap emerged from the analysis of the existing body of knowledge. The discussion in this chapter highlights several areas where current CAS practices in maintenance planning and execution are insufficient to meet the nuanced requirements of modern manufacturing environments, particularly in the semiconductor industry.

- **Limited Real-World Application:** Current AI applications in industrial maintenance are predominantly theoretical or limited to controlled environments such as laboratories. This indicates a gap in the practical, real-world application and validation of these AI systems in complex industrial environments.
- **Data Utilization Challenges:** It is evident that while structured data from machine sensors are used extensively, unstructured data (e.g., maintenance logs and operator

notes) are significantly underutilized. Existing systems have not fully realized the potential of integrating these data types to improve predictive maintenance capabilities.

- **Integration and Interoperability Issues:** Integrating AI tools into existing industrial maintenance systems is a challenge that has yet to be fully addressed. While current systems offer suggestions for problem-solving, they often lack the necessary background information to be truly effective.
- **Lack of Expertise:** There is a noticeable dearth of robust research that can accurately assess and categorize the skills and competencies of maintenance technicians. Current systems do not adequately capture the dynamic and diverse skills required in high-tech industries such as semiconductor manufacturing.
- **Dynamic Skill Matching:** Existing methodologies fall short in dynamically matching the skills of personnel to the evolving needs of maintenance tasks. This is critical to ensuring that the most qualified technicians are deployed.

4 Conceptual Framework of ARCHIE

This thesis introduces an architecture for a cognitive maintenance system dubbed ARCHIE. ARCHIE signifies a substantial advancement in the domain of industrial maintenance, particularly pertinent to the automotive and semiconductor industries, due to its generalizable and scalable gray box approach. ARCHIE was developed based on the gaps in the research landscape identified in Section 3.2, which, in turn, are based on the building blocks identified in Section 3.1.1. The present chapter lays the groundwork for a comprehensive exploration of ARCHIE's modular framework, its alignment with the principles of AI-enhanced assistance, and its potential to improve industrial maintenance KPIs (e.g., availability and MTTR).

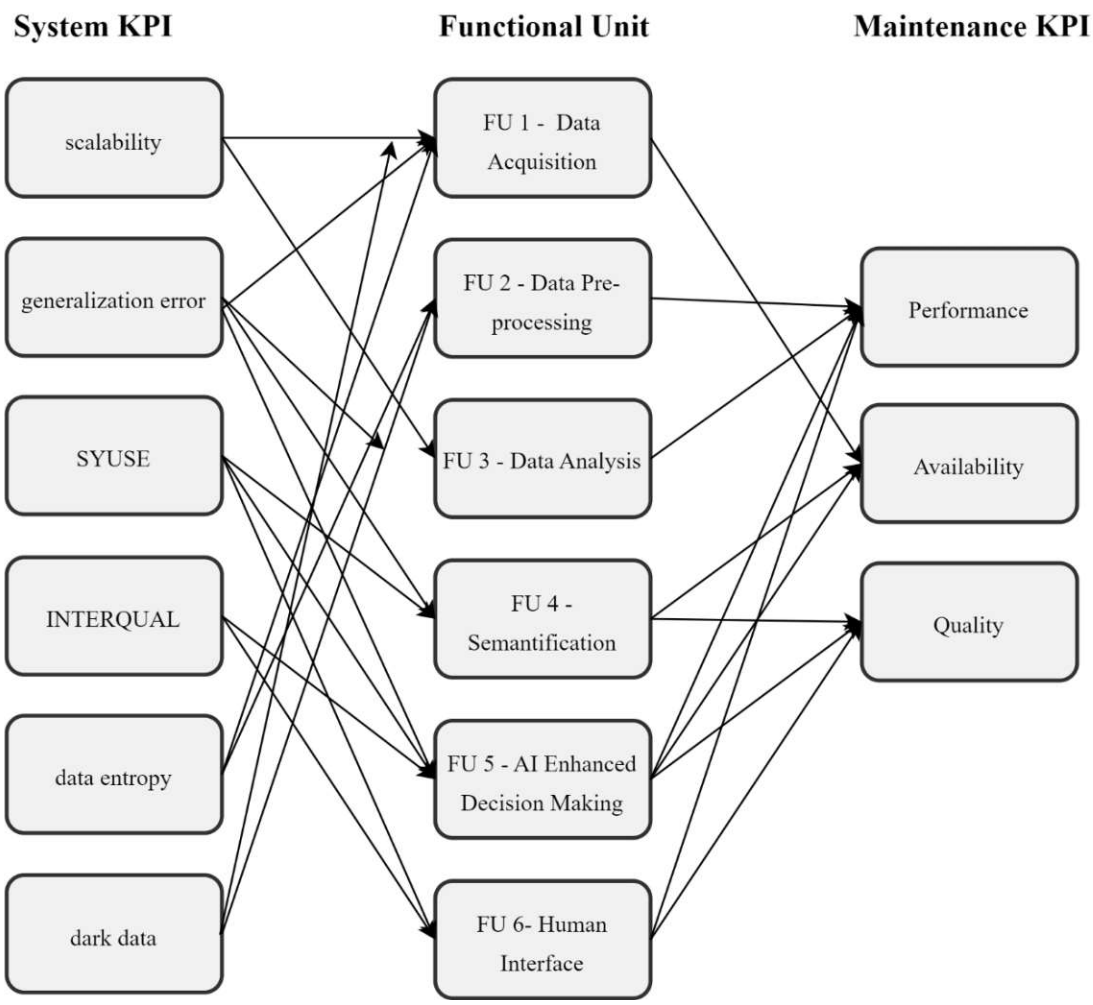
The primary aim of human-centered cognitive systems within the domain of industrial maintenance is to augment the cognitive capabilities of planners and technicians, as delineated by Fischbach et al. (2020). The objective is to achieve this improvement through the synergistic integration of unstructured and structured data. A significant challenge present here is the extraction of explicit and implicit knowledge, which can be linked to individual historical actions and the competencies utilized to perform them. Such a linked knowledge base can serve as a foundation for recommendation systems for planning and action through reasoning (Borgianni et al., 2023). The proposed ARCHIE is designed to actualize this AI-driven, human-centric assistance paradigm. ARCHIE integrates an array of both physical and virtual sensors to monitor machine states and parameters, as well as to augment human expertise and competencies. These optimizations encompass reductions in planning time, documentation time, root cause identification, problem-solving, and human error rates, thereby improving availability and MTTR.

The foundational design principle of ARCHIE is predicated on the assimilation of multi-channel data. Data acquisition is executed through physical sensors (e.g., machine and image data) and virtual sensors (e.g., audio, text), offering a holistic environmental overview essential for the cognitive maintenance system's optimal functionality. Virtual sensors are a crucial element in this process as they facilitate the utilization of previously untapped, predominantly unstructured data sources, which offer a supplementary dimension to the asset's condition assessment. This dataset incorporates both objective machine and process information from interconnected systems and subjective evaluations from machine operators and maintenance technicians based on their individual competence levels.

This comprehensive data landscape facilitates the strategic implementation of maintenance interventions, precisely timed and directed at the appropriate machinery, executed by optimally suited maintenance technicians. This approach directly enhances KPIs such as machine availability and performance. Indirectly, the human-centric augmentation of maintenance activities serves to elevate the performance and quality of these operations, thereby positively influencing the quality of KPIs. Collectively, these enhancements result in a general increase in OEE.

The following design principles underpin the ARCHIE framework. These design principles, along with ARCHIE's system KPIs, are foundational to understanding how ARCHIE, with its FUs, effectively addresses OEE and its underlying KPIs.

Figure 34: *Influence of the System KPIs on the FUs and Impact of the FUs on the Maintenance KPIs*



- **Generalizability and Scalability:** ARCHIE's architectural design is not constrained by the limitations of specific industries, exemplifying a generic design that can be adapted for use in various domains. The generalizability of this type is ensured by the openness of the data sources and the framework conditions in the planning and recommendation logic. This adaptability is crucial for ensuring that the architecture remains relevant and effective in diverse industrial contexts. Furthermore, scalability is a key characteristic of ARCHIE, enabling it to be deployed effectively in scenarios ranging from single-component applications to entire factory-wide systems. This scalability ensures that the system can grow and evolve in tandem with the expanding needs and complexities of industrial environments. In the context of ARCHIE, the term "scalability" is defined as the ratio of load to the demand of system recommendation $s = \frac{\sum x}{\sum y}$, where s is the scalability factor, x is the number of computational loads of recommendations, and y is the number of computational loads of demanded recommendations. Then, the concept of generalizability is calculated using the generalization error $I_n[f] = \frac{1}{n}\sum_{i=1}^{n} V(f(x_i), y_i)$ (Hastie et al., 2001), where $I_n[f]$ is the generalization error of the data-depended function f, with its predicted out values y and inputs x, where n is the number of data points.

- **Customizability:** The ARCHIE system can be customized to align with specific employee competencies. While skills specifically refer to the ability to use methods or instruments for defined tasks in particular settings, competence encompasses a broader scope that enables individuals to face new situations and challenges independently. Competency differs from competence as it focuses on specific capabilities rather than the holistic ability to perform (Hernandez-de-Menendez et al., 2020). Unlike qualifications, which primarily certify formal learning, competence emphasizes practical application and proven ability in real-world situations (European Commission et al., 2017). This level of customization enhances the practical utility of the system, making it more effective and user-friendly. The customizability can be measured with the interface quality (INTERQUAL) in the PSSUQ (Jordan et al., 1996).
- **Human-Centered Design:** ARCHIE places significant emphasis on user experience, incorporating a human-centered design that prioritizes intuitive interaction. This approach ensures that the system integrates seamlessly into the existing workflow, enhancing user engagement and reducing the learning curve. The design of the user interface, developed in close collaboration with end-users and decision-makers, ensures that ARCHIE meets the real-world needs of its operators, thereby enhancing its practical utility and acceptance. Human centricity can be measured by system usefulness (SYUSE) in the PSSUQ (Jordan et al., 1996).
- **Advanced Data Processing Techniques:** At the core of ARCHIE's functionality is its sophisticated data processing capability. This encompasses multi-channel data acquisition, preprocessing, and analysis using ML or TLP, including large language

models, based on the data utilized. The system can normalize sensor data, transform audio data into textual format, and extract textual data from corporate resources. This comprehensive data processing allows ARCHIE to utilize a vast array of data types, making it exceptionally versatile in handling diverse industrial maintenance scenarios. ARCHIE's strength lies in its flexible analytical options for analyzing dark data. Dark data, as defined in this context, refers to information that is collected but not used by companies $dd = \frac{\sum data_{used}}{\sum data_{available}}$ (Corallo et al., 2023; Gartner, 2024) where dd is the ratio of dark data, $data_{used}$ is the number of data sources used, and $data_{available}$ is the total number of data sources availale in the enterprise. The analytical capabilities of ARCHIE are, therefore, measured by the degree of dark data in relation to the total data.

- **Knowledge-based Data Storage Solutions:** ARCHIE employs a KG for data storage, which maps real-world entities and their relationships in a graph structure. This storage method is crucial for AI-enhanced decision-making, as it allows for the identification of patterns and relationships within the data. The KG's ability to represent complex interrelations between various entities, such as maintenance technicians, machine operators, and sensor values, adds a layer of depth and sophistication to ARCHIE's analytical capabilities. The ability of ARCHIE to represent various data sources and complex relations can be calculated through its data entropy $H(X) := -\sum_{x \in \mathcal{X}} p(x) \log p(x)$, a measure of its nonuniformity (Hastie et al., 2001), where $H(X)$ is the entropy and X is a descrete random variable of the dataset $\mathcal{X}$.
- **I-Enhanced Decision Making:** Utilizing the information stored in the KG, ARCHIE applies statistical and similarity-based learning algorithms to suggest the most suitable actions. This AI-enhanced decision-making process is critical for optimizing maintenance strategies, recommending appropriate maintenance measures, and evaluating required competencies for specific tasks. This aspect of ARCHIE not only increases operational efficiency but also significantly enhances the strategic decision-making process in maintenance management. Human-centricity can be measured by information quality (INFOQUAL) in the PSSUQ (Jordan et al., 1996).

The metrics collected and the methodology employed in their calculation are presented in Table 4. These benchmarking values are utilized to facilitate continuous improvement and expansion of ARCHIE.

Table 4: *Metrics of ARCHIE Design Principles*

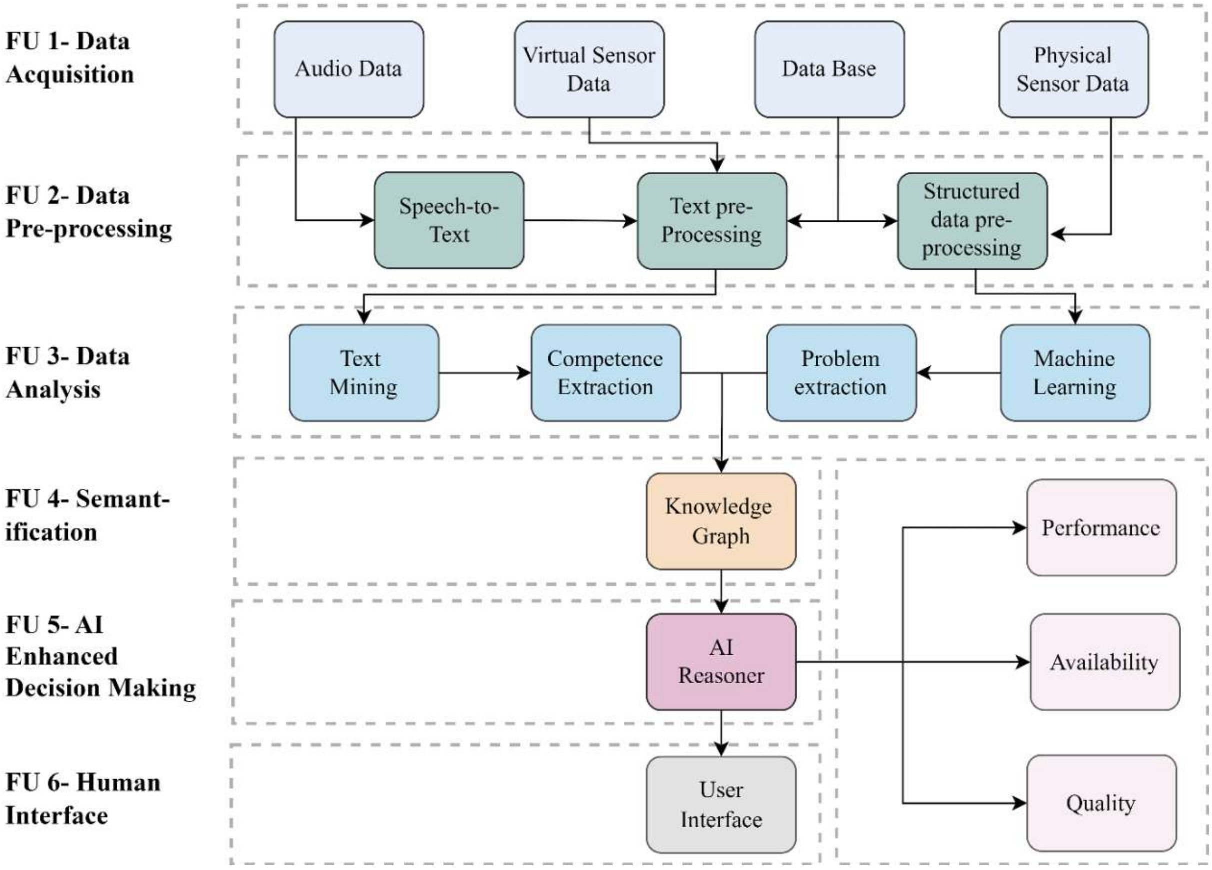
Metric	Calculation
scalability	$scalability = \frac{load\ of\ recommendations}{demand\ for\ recommendations}$
generalization error	$I_n[f] = \frac{1}{n}\sum_{i=1}^{n} V(f(x_i), y_i)$
SYUSE	$\mathrm{SYUSE} = \sum_{i=1}^{n} sysuse$
INFOQUAL	$\mathrm{INFOQUAL} = \sum_{i=1}^{n} infoqual$
INTERQUAL	$\mathrm{INTERQUAL} = \sum_{i=1}^{n} interqual$
data entropy	$H(X) := -\sum_{x \in \mathcal{X}} p(x) \log p(x)$
dark data	$dd = \frac{\sum data_{used}}{\sum data_{available}}$

As Figure 35 shows, ARCHIE is a modular system comprising six fundamental FUs, each integral to its operational framework.

- **FU1 - Data Acquisition:** ARCHIE's architecture is engineered with adaptable interfaces to facilitate multi-channel data acquisition, a critical component in comprehensive data analysis. This unit captures data in two primary forms: structured data from sensors and unstructured data, including text, images, or audio, which is subsequently converted into a textual format. This approach to data acquisition is pivotal for ensuring a robust dataset for nuanced analysis by providing a better semantic database interlinked with physical and virtual sensors.
- **FU2 - Data Pre-processing:** This unit in ARCHIE is responsible for the initial processing of data, tailored to its source. Structured data, such as sensor outputs, undergo normalization or one-hot encoding, depending on their categorical nature, to prepare them for downstream ML applications. Conversely, unstructured data, like audio recordings, are converted into text via speech-to-text algorithms. Textual data, whether sourced directly or extracted from corporate repositories via web scraping, are subjected to tokenization, stopword removal, lemmatization, and part-of-speech

tagging. These processes are crucial for identifying semantic relationships within the data, thereby preparing it for in-depth analysis. In specific scenarios, the creation of specialized dictionaries is necessary to address technical jargon or enterprise-specific terminologies.

Figure 35: *Design Principle and functionality of an Architecture for a Cognitive Maintenance System*



- **FU3 - Data Analysis**: The data analysis unit in ARCHIE is selected based on the data type and source, employing various algorithms tailored to the data's nature. For structured data, ML techniques, such as random forest or neural networks, are utilized for tasks such as downtime prediction and critical machine fault detection. In contrast, TLP models are applied to unstructured data to extract information such as maintenance activities and competencies. Image data are also analyzed using DL models, such as convolutional neural networks, to identify potential defects or quality issues.
- **FU4 - Data Storage:** ARCHIE employs a KG for data storage, a sophisticated structure that maps real-world entities and their interrelations. This KG encompasses a wide array of entities, from maintenance technicians and machine operators to sensor data and internal documents. It also includes abstract entities such as competencies and job. descriptions, all interconnected within the graph. This structure facilitates the discovery of hidden relationships and the derivation of optimal action recommendations, ranging

from personnel selection for specific machine issues to maintenance strategy suggestions.

- **FU5 - AI-enhanced Decision Making:** This unit leverages the KG to suggest actions using statistical and similarity-based learning algorithms. These algorithms are instrumental in identifying potential causes of machine faults and appropriate maintenance measures. A reasoner is applied based on the KG to extract relations, and it is then possible to derive insights and recommendations for downstream algorithms such as a SLA.
- **FU6 - Human Interface:** The final unit focuses on providing tailored recommendations and support, emphasizing a human-centered design approach. ARCHIE prioritizes intuitive user interfaces and real-time, relevant information delivery, ensuring high user acceptance and operational efficiency. This unit also includes generic tools for maintenance tasks, competence assessments, and KPI predictions, demonstrating ARCHIE's versatility and added value in maintenance operations.

Collectively, these FUs can be employed individually or in tandem, maximizing ARCHIE's capabilities. Additionally, ARCHIE's modular design allows for integration into broader maintenance models such as PRIMA (Ansari et al., 2019), which, with its layered architecture including a data analytics and semantic layer, facilitates the integration of an assistance system such as ARCHIE. The FUs can be used individually for each use case. In combination, however, they exploit the full possibilities of ARCHIE. Moreover, ARCHIE can also be implemented as part of a comprehensive maintenance model such as PRIMA (Ansari et al., 2019). Therefore, the design principle shown in Figure 35 represents a modular structure that allows a partial validation of ARCHIE and its specific functionalities, as shown in Table 5.

Table 5: *Overview of FUs of ARCHIE and Their Specific Functions*

Functional Units	**Specific Functions**
1 - Data Acquisition	Multi-channel data acquisition, including virtual and physical sensors
2 - Data Preprocessing	Data processing, including speech-to-text.
3 - Data Analysis	Combination of analytics pipelines for structured and unstructured data, especially sensor data and textual data, including a novel SLA
4 - Data Storage	Semantic storage of sensor data and textual data, including relations and logical restrictions
5 - AI-enhanced Decision Making	Utilization of a first-order inference system with a downstream recommender model using the KGs
6 - Human-centered Support	Versatile interface that can be adapted to a wide range of maintenance workflows in planning and operations

4.1 FU1: Data Acquisition

Correct data acquisition plays a pivotal role for decision-making in maintenance planning and operations, particularly in the context of sensor and text data, which form the backbone of predictive maintenance strategies. Sensor data, derived from a multitude of devices embedded within machinery, provides real-time insights into equipment performance, enabling early detection of potential faults (Al-amri et al., 2021). These sensors, ranging from temperature to vibration detectors, generate structured data that can be continuously monitored and analyzed for anomalies indicative of impending malfunctions (Müller et al., 2020). Concurrently, text data, often sourced from maintenance logs, operator reports, and technical manuals, offer a rich repository of unstructured information captured in virtual sensors. When processed using advanced TLP techniques, this textual information can reveal patterns and insights that are not immediately apparent from sensor data alone (Biegel et al., 2022).

Moreover, the integration of image and video data, though less prevalent, is gaining traction in maintenance. When converted into text through advanced transformer models, these data forms can supplement sensor and text data, providing a more comprehensive understanding of the equipment health index. For instance, image data can be used to detect visible signs of wear or corrosion, while video data can capture dynamic anomalies in machinery operation (Müller et al., 2020). The conversion of these visual data forms into textual descriptions enriches the data pool and makes it amenable to the same analytical techniques used for text data, thereby creating a more integrated and robust maintenance data ecosystem.

4.1.1 Data Acquisition of Structured Data

Based on a literature review by Saini et al. (2022) it can be deduced that much of the machinery utilized in contemporary industrial settings comprises rotating components, including gears, bearings, and shafts, as extensively documented in the literature (Cui et al., 2017; Cui et al., 2019; H. Li et al., 2015; Y. Li et al., 2018; R. Liu et al., 2018; Lu et al., 2019; Shen et al., 2013). These machines often operate under strenuous and challenging environmental conditions, elevating the likelihood of fault development.

Table 6: *Sensor Technology and Its Proven Applications Symptoms (Saini et al., 2022)*

		Technology					
		Vibration	Lubrication	Wear	MCA	Infrared	Acoustics
Application	Generator	✓	✓	✓	×	✓	✓
	Turbine	✓	✓	✓	×	✓	✓
	Pump	✓	✓	✓	✓	✓	✓
	Electric Motor	✓	✓	✓	✓	✓	✓
	Engine	✓	✓	✓	×	✓	✓
	Fan	✓	✓	✓	✓	✓	✓
	Gearbox	✓	✓	✓	×	✓	✓
	Cranes	✓	✓	✓	✓	✓	✓
	Electric Circuit	×	×	×	✓	✓	✓
	Transformer	×	✓	×	✓	✓	✓

The rotating parts within these machines are prone to a range of faults, such as gear tooth chipping, surface pitting, gear root cracks, defects in the bearing's inner and outer races, ball defects, cage defects, shaft bending, misalignment, and imbalance (Saini et al., 2022). The fault of any single component in these machines has the potential to result in the complete shutdown of the entire machine or asset. Therefore, early fault detection is crucial to prevent catastrophic failures and monitor the progression and intensity of these faults over time, as emphasized in various studies (Y. Li et al., 2019). This proactive approach to fault diagnosis is essential for maintaining operational continuity and ensuring safety in industrial environments. Data acquisition involves collecting and storing data using various sensors attached to machinery. This process is fundamental to condition monitoring and serves as an initial step in machinery prognostics. Common sensors used in this process include accelerometers, acoustic emission sensors, infrared thermometers, and current sensors. Vibration sensors, in particular, are widely employed for diagnosing faults in bearings (Goyal et al., 2017; Hoang & Kang, 2019) and gears. Acoustic emission sensors are effective in detecting early-stage faults in bearings (Z. Liu et al., 2018) and gears (Wong et al., 2017) especially when machines operate at low speeds and in environments with low-frequency noise. Bearing faults (Goyal et al., 2017) can be identified using specialized approaches, such as damage to the outer race, inner race, roller elements, multi-fault damage, and wear. Table 6 summarizes various sensor technologies and their applications for general mechanical components in a typical plant, whereas

Table 7 outlines each technology's capability in detecting symptoms related to faults. Although ARCHIE offers the possibility to include all shown sensor types, it has been shown that vibration and acoustics can be included (Biegel et al., 2022).

Table 7: *Summary of General Symptoms of Faults and Sensor Technologies That Can Detect Those Symptoms (Saini et al., 2022)*

		Technology					
		Vibration	Lubrication	Wear	MCA	Infrared	Acoustics
Symptoms	Wear	✓	×	✓	×	×	✓
	Heating	✓	✓	✓	×	✓	×
	Impact	✓	×	✓	×	×	✓
	Corrosion	×	✓	✓	×	×	×
	Fatigue	✓	✓	✓	×	×	×

4.1.2 Data Acquisition of Unstructured Data

The data acquisition layer, leveraging virtual sensors, plays a crucial role in aggregating data from diverse sources, including maintenance reports and internal enterprise wikis. Maintenance reports, particularly from digital shift books[1] (DSBs), are rich in details about the equipment, the executed maintenance tasks, and any challenges or anomalies encountered during these processes. Additionally, the enterprise wiki serves as a repository of comprehensive information on equipment specifications, detailed maintenance procedures, and established best practices.

To enhance the process of locating similar maintenance records, the system employs a similarity search mechanism, which is based on the principle of cosine similarity (Banik, 2018), as outlined in Equation 1. In this context, A represents a new maintenance entry, while B denotes an existing entry in the system. The foundation of the cosine similarity calculation is a term-document matrix, which is constructed using the term frequency-inverse document frequency (TF-IDF) method (Banik, 2018) as described in Equation 3. In this matrix, t signifies the frequency of a specific term, n indicates the total count of documents, and df is the number of documents that contain the term in question.

$$Cos(A, B) = \frac{\sum_{i=1}^{n} A_i \times B_i}{\sqrt{\sum_{i=1}^{n} A_i} \times \sqrt{\sum_{i=1}^{n} B_i}} \quad (3)$$

$$TF - IDF = \left(\frac{t}{n}\right) \times \log\left(\frac{n}{df + 1}\right) \quad (4)$$

The term-document matrix is structured with rows representing key terms extracted from all documented maintenance events and columns populated with unique identifiers for these events. Each cell within this matrix contains the frequency of a particular term, adjusted by the size of the input data. This matrix not only facilitates the identification of similar maintenance entries but also enhances the speed and accuracy of the query process, as found in the study by (Ansari et al., 2021). By leveraging these advanced computational techniques, the data acquisition layer significantly improves the maintenance management system’s capability to swiftly and precisely match new entries with existing records, thereby streamlining maintenance operations and knowledge sharing within the enterprise.

[1] Digital shift books facilitate the digital management of shift handovers and the documentation of activities such as maintenance. They can be implemented as a standalone solution or as part of an ERP program, such as SAP-PM©.

4.2 FU2: Data Preprocessing

Data preprocessing plays a crucial role in the field of industrial maintenance, particularly in the context of PdM strategies. Maintenance tasks are often guided by large volumes of data collected from various virtual and physical sensors and machine controls in industrial settings. However, these data can be noisy, incomplete, and inconsistent, leading to inaccurate predictions and inefficient maintenance decisions. Therefore, effective data preprocessing techniques, such as data cleaning, normalization, outlier detection, and feature extraction, are essential to ensure the quality and reliability of these data. The impact of data preprocessing on the used enterprise data can be measured in data entropy and dark data, as shown in Table 4. Correctly prepared data is the basis for accurate predictions of the subsequent FU 3. The effort required to prepare the data correctly is also the most time-consuming part of an industrial data science project (Geron, 2022). Data preprocessing is a foundational step in transforming raw data into meaningful insights, which is vital for the effective implementation of advanced maintenance strategies in industrial environments (Ansari et al., 2019).

4.2.1 Data Preprocessing of Structured Data

Mechanical components are often designed for variable and sometimes very low-speed operations, which poses challenges for analyzing them using only raw signals, as obtained in Section 4.1.1. The low-energy, low-frequency signals generated under these conditions are often obscured by strong environmental noise, making it difficult to identify any hidden faulty signals. To address this, various signal denoising techniques with subsequent feature processing have been developed, tested, and applied, with their effectiveness well-documented in existing literature. Signal preprocessing methods are categorized based on their diagnostic utility (Di Wu et al., 2022; Saini et al., 2022) into (i) time-domain, (ii) frequency-domain, and (iii) time-frequency domain methods, each of which is elaborated in the following sub-sections. The advantages and disadvantages of the techniques and their applications are outlined in

Table **8**.

- **Time-Domain Signal Processing:** This is typically employed for diagnosing faults that generate periodic shocks or peaks in the signal. It is considered to be the simplest feature extraction technique since it requires no transformation. Time-domain techniques analyze variations in amplitude over specific time intervals. Key time-domain processing techniques include the amplitude envelope, root mean square energy, and zero crossing rate (Krishnamurthi et al., 2022). However, its effectiveness is limited in variable speed applications. Leaman et al. (2021) explored a method for calculating time-domain averages in planetary gearboxes. Further, Tang et al. (2020) examined statistical indicators such as root mean square and standard deviation to differentiate between faulty and healthy states in planetary gearboxes. Additionally, Q. Feng et al. (2023) investigated how continuously varying load conditions affect the vibration signals of gearboxes.

- **Frequency-Domain Analysis:** This examines physical signals or time-series data concerning frequency components, often using the Fourier transform method (Kheirrouz et al., 2022). It identifies the distribution of signals within specified frequency bands over a range of frequencies; however, it often struggles to accurately locate faults within machine sub-components due to its limited scope in the time domain. Conversely, different machine components exhibit distinct frequencies during operation. Therefore, frequency domain methods are employed to address the limitations of time-domain methods, offering precise fault diagnosis and condition monitoring of mechanical components.

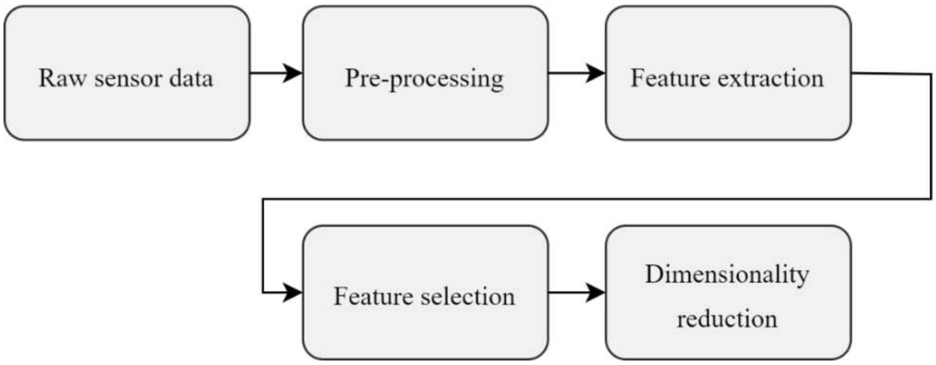
Table 8: *Advantages, Disadvantages, and Applications for the Preprocessing of Structured Sensor Data*

Method	Advantage	Disadvantage	Applications
Time-Domain Method	Simple, efficient; detects abrupt changes and transients	Limited frequency analysis; may miss periodic features	Monitoring vibrations; detecting sudden operational changes
Frequency Domain Method	Identifies frequency components; diagnoses periodic issues	Loses time context; computationally intensive; less effective for non-stationary signals	Analyzing vibration spectra; identifying wear and misalignment
Time-Frequency Domain Method	Captures non-stationary and transient signals; detailed insight	High computational load; complex interpretation	Early fault detection; monitoring transients; advanced acoustic analysis

- **Time-Frequency Domain Methods:** These are applied to non-stationary signals, where analysis techniques such as short-time Fourier transform and wavelet transform are suitable for analyzing those non-stationary and non-linear signals (Pandiyan et al., 2020). The drawback of this advanced technique is that it averages out the amplitude of frequencies over time when converting signals from the time domain, making it challenging to analyze systems whose dynamics change over time. Time-frequency domain signal processing methods are utilized to address this, as they can represent signals in both time and frequency domains. Various time-frequency domain methods, such as Wigner–Ville distribution and wavelets, have been applied. K. Feng et al. (2023) analyzed instantaneous speed and vibration in gearboxes operating under non-stationary conditions, applying a time-frequency method to detect faults in automobiles.

The primary goal of feature processing is to extract relevant information that correlates with increasing machine fault severity. This process involves three key steps: (i) preprocessing, (ii) feature extraction, (iii) feature selection, and (iv) dimensionality reduction (Hastie et al., 2001), Figure 36.

Figure 36: *Data Pre-Processing Pipeline of Structured Data*



- **Preprocessing:** It represents the first phase in the domain of signal analytics, in which the initial, unfiltered data obtained from physical sensors undergo techniques such as filtering, denoising, and normalization to mitigate noise, suppress unwanted frequency components, and standardize the range of measurements.
- **Feature Extraction:** This involves deriving various features to assess machine health. These features fall into the categories of time domain, frequency domain, and time-frequency domain. Time-domain features are divided into dimensional (e.g., mean and standard deviation) and dimensionless types (e.g., skewness and kurtosis). Frequency domain features (e.g., mean frequency and frequency center) capture information not present in time-domain features. Time-frequency domain features, like energy entropy,

are used under non-stationary conditions. Various methods, including wavelet packet decomposition, are applied for feature extraction in different scenarios, such as bearing fault diagnosis and gearbox fault detection (Cui et al., 2019).

- **Feature Selection:** This identifies sensitive features that indicate machine health. Feature selection methods include filters (e.g., relief and information gain), wrappers (assessed by classifier performance), and embedded methods (Hoang & Kang, 2019). These techniques help select features with high diagnostic value for faults in gearboxes and other components.
- **Dimensionality Reduction:** This creates features with fewer dimensions using linear (e.g., principal component analysis) and non-linear methods (e.g., kernel function-based method). This step is crucial for managing complex, multi-dimensional feature sets, aiding in the diagnosis of faults in various machinery such as wind turbines and gearboxes. Researchers have developed specific features for this purpose, such as statistical features and energy-based metrics (T. Xie et al., 2023).

4.2.2 Data Preprocessing of Unstructured Data

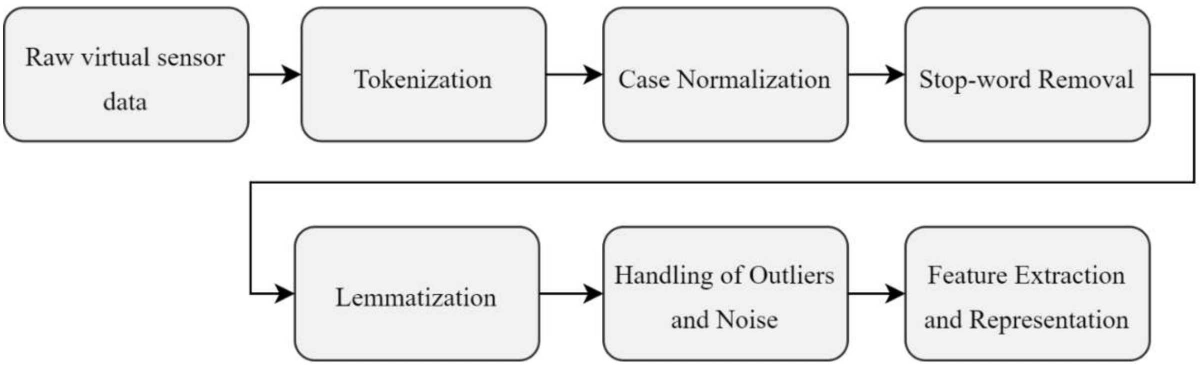
The process of transforming raw data from text via virtual sensors into a format that is both machine-readable and interpretable is a critical step in enhancing various predictive applications. The underlying research hypothesis posits that effectively converting DSB entries into a structured format can significantly improve predictions across several domains (Ansari et al., 2021). These include automated recommendations for documentation, accurate downtime prediction, and identifying the best-fit maintenance technician for specific tasks.

While focusing exclusively on the preprocessing of textual data, it is essential to understand the current techniques. These techniques are critical in transforming raw text data into a format that is not only machine-readable but also conducive to sophisticated TLP and ML Figure 37.

- **Tokenization**: The process of tokenization has evolved with the development of subword tokenization methods, which are integral to foundational models (C. Zhou et al., 2023) and models such as BERT and GPT-3 (Brown et al., 2020; Devlin et al., 2018; Peters et al., 2018). These methods break down text into smaller units, capturing the nuances of language more effectively than traditional tokenization.
- **Case Normalization:** Lowercasing text is a standard preprocessing step. However, recent advancements in case-sensitive NLP models, such as the T5 framework by (Raffel et al., 2019), have led to a reevaluation of this step, considering the contextual importance of case information in certain applications.

- **Stop-word Removal:** The practice of removing stop words (commonly used words with minimal contextual significance) has been debated. Rogers and Augenstein (2020) discussed its limited impact on the performance of DL models, suggesting that stop-word removal might not be as critical as previously thought in the context of advanced NLP models.
- **Lemmatization**: Context-aware lemmatization techniques have seen significant improvements. Modelssuch as ELMo and BERT offer more sophisticated approaches to lemmatization, considering the context in which words are used to determine their base form (Devlin et al., 2018; Peters et al., 2018).
- **Handling of Outliers and Noise**: In textual data, outliers and noise can manifest as irrelevant or misclassified information. Advanced NLP techniques and models have been developed to better identify and handle these anomalies, ensuring cleaner and more relevant datasets for training and analysis (Al Sharou et al., 2021).
- **Feature Extraction and Representation:** The transformation of text data into a format suitable for ML models often involves feature extraction techniques such as TF-IDF or embeddings generated by models like BERT and GPT-3. These techniques have significantly evolved, allowing for more nuanced and contextually rich representations of text data (Brown et al., 2020; Devlin et al., 2018).

Figure 37: *Data Preprocessing Pipeline of Unstructured Data*



In addition to the aforementioned preprocessing steps, vectorization (Alessandro Moschitti et al., 2014; Pennington et al., 2014) is pivotal in transforming textual data into a format suitable for ML and TLP applications. Vectorization refers to the conversion of text into numerical vectors, which ML algorithms can process. This is a crucial process that translates the inherently non-numeric nature of text into a form that algorithms can understand and analyze. A practical example would be to transfer all the words used in a maintenance report into columns of a table. Thus, all the records of the maintenance reports used can be compiled by setting 0 and 1 in a row.

Two primary methods of vectorization are widely used in recent literature: bag-of-words (BoW) and word embeddings (Vajjala et al., 2020). The BoW model, as described by Mikolov et al. (2017), represents text as an unordered collection of words, disregarding grammar and word order but keeping multiplicity. This method, however, often leads to high-dimensional and sparse matrices, which can be computationally intensive. Even so, as can be observed from the DPO example, there are more 0s than 1s in such a row because at least the records use the entire available vocabulary, which is called a sparse vector. Therefore, a lot of matrix operations are needed and lead to computationally intensive operations.

Conversely, word embeddings provide a more sophisticated approach, capturing connections between words. They represent words in a continuous vector space where semantically similar words are mapped to proximate points. This method captures more contextual information than BoW. The advent of word embeddings, particularly through models like Word2Vec (Mikolov et al., 2017) and GloVe (Pennington et al., 2014) has revolutionized how text is represented, enabling more nuanced and sophisticated analysis of large text corpora. Word embeddings significantly influence transformers and large language models by providing a foundational numerical representation of words in continuous vector spaces. These embeddings capture semantic and syntactic relationships between words, enabling the models to understand context and meaning. In transformer architectures, word embeddings serve as the initial input that is processed through attention mechanisms, allowing the model to weigh the relevance of different words in a sequence. Effective word embeddings enhance the ability of large language models to model language patterns, comprehend context, and generate coherent and contextually appropriate responses.

Building upon the described data acquisition techniques for textual data in maintenance, there has been an increasing emphasis on incorporating multimedia data sources such as images (Jiawei Xie et al., 2020), videos, and speech (Q. Li et al., 2019). These forms of data, while inherently different from traditional text-based inputs, offer a wealth of information that can significantly enhance the maintenance process. For instance, images and videos can capture visual cues of equipment wear or damage that might not be easily described in text. Advanced image processing techniques and video analytics can be employed to extract critical information from these visual data sources. Similarly, speech data, often collected from verbal reports or comments by maintenance personnel, provides valuable insights into equipment conditions and maintenance experiences (J. Wang et al., 2022).

To integrate these varied data types into the existing text-centric framework, they are converted into textual data using sophisticated algorithms. Image data are processed through image-to-text transformers and image-to-text conversion techniques, and videos can be translated to text via the use of chained encoder models to translate visual information into descriptive text. Moreover, speech data undergo a transformation process via speech-to-text algorithms, converting verbal reports into written form employing hidden units for encoder-decoder (Z. Zhang et al., 2022). Once converted, this textual data is then subjected to the same analytical processes as traditional text data, such as TF-IDF analysis, to ensure consistency and comparability across all data types. This integration broadens the scope of data available for maintenance analysis and ensures a more comprehensive and multi-dimensional approach to equipment monitoring and fault diagnosis, leveraging the strengths of diverse data formats.

4.3 FU3: Data Analysis

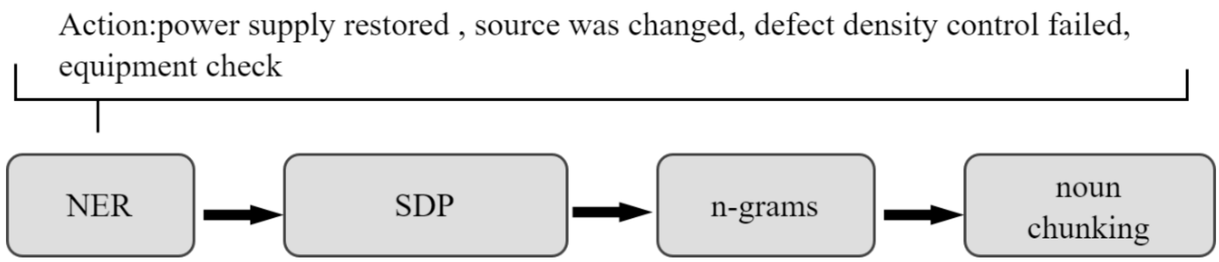
The data analysis layer aims to analyze the information from the preprocessing in such a manner that the causes of faults can be identified for both maintenance activities and machine data. These data represents the concepts that are later semantically linked and required for AI-based decision-making.

4.3.1 Text Mining

The text mining pipeline (TMP), which serves as an intermediary between raw data collection and advanced analytical operations, is intricately designed to refine raw data obtained from the initial data acquisition phase. It employs a series of sophisticated techniques, including data cleaning, normalization, and transformation. These methods are pivotal in mitigating noise and reducing redundancy in the dataset, thereby enhancing the quality of the information for further processing (Gharatkar et al., 2017).

Central to the TMP's functionality is the extraction of relevant entities from unstructured text data. This is achieved through an array of advanced text mining techniques (Madreiter et al., 2021). Named entity recognition and semantic dependency parsing are employed for identifying and categorizing key elements within the text (K. B. Cohen & Verspoor, 2013). The TMP utilizes n-grams and noun-chunking strategies to meticulously dissect the text, ensuring a comprehensive extraction of all pertinent information from sources such as maintenance reports, as illustrated in Figure 38.

Figure 38: *Workflow of the Text Mining Pipeline Based on (Kohl & Ansari, 2023)*



The pipeline incorporates sophisticated TLP tools, notably spaCy[2] and XLM-RoBERTa (Raffel et al., 2019). These tools are instrumental in parsing and understanding the context within the text (Raffel et al., 2019). In the subsequent stages of the TMP, repetitive tasks are identified and filtered out to avoid redundancy. Moreover, the process involves the resolution of company-specific abbreviations into their full forms, enhancing the clarity and comprehensibility of the extracted data. In line with data protection regulations, personal names are meticulously extracted to ensure adherence to privacy standards. The TMP process can be extended by using large language models that can interpret semantically incomplete sentences and derive further events from them. However, these large language models still need to be trained with company-specific or at least domain-specific data so that the large language model can handle the wording used (Kohl, Eschenbacher, et al., 2024).

The extracted entities encompass a wide range of information, including but not limited to equipment names, maintenance activities, and detailed accounts of issues or challenges encountered during maintenance procedures. This rich dataset forms the basis for subsequent analytical processes.

In the downstream phase of the TMP, a classification algorithm is employed to further categorize the processed data. Logistic regression, a statistical method, is utilized to ascertain whether the extracted entities pertain to tasks or other categories (Doe et al., 2018). This classification is crucial for streamlining the data and preparing it for more nuanced analyses and decision-making processes.

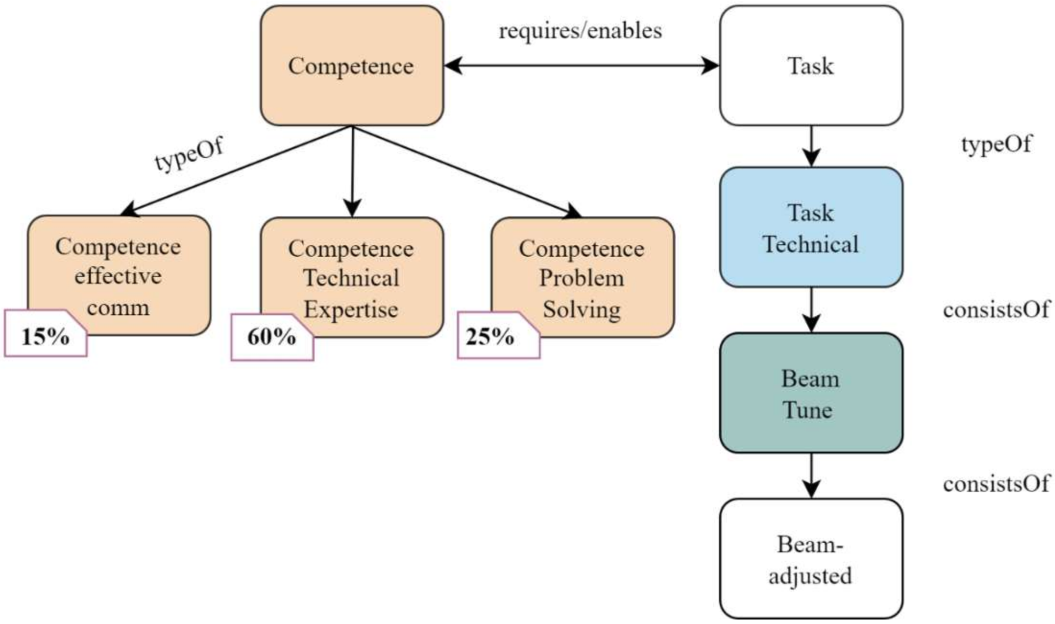
4.3.2 Statistical Learning Algorithm for Competence Extraction

The SLA is designed to transform tasks into the necessary competencies based on their distributions. In this thesis, competence refers to an individual's capacity to perform job responsibilities, in contrast to competency, which focuses on an individual's actual performance on a particular task, as defined by the European Commission et al. (2017). SLA plays a pivotal

[2] spaCy - Industrial-strength Natural Language Processing in Python. Retrieved 09.03.2025, from https://spacy.io/

role in interpreting maintenance reports and internal enterprise wiki data, converting extracted entities into competencies required for maintenance tasks. The algorithm achieves this by mapping these entities to related concepts within a KG, as illustrated in Figure 39.

Figure 39: *Representation of the Semantic Connections Used for SLA*



The foundation of the SLA is the ESCO framework (European Commission et al., 2017), which provides a comprehensive classification of tasks in a semantic structure, connecting jobs, skills, qualifications, and competencies. The advantage of the ESCO classification is that it is a Europe-wide standard. However, national or international schemes such as O*NET (U.S. Department of Labor, 2024) can also be used as long as they are available in a semantic structure. The underlying principle of the SLA is that the frequency of performing a technical task correlates with its routineness. Tasks performed more frequently are considered routine and typically require less communication and problem-solving. Additionally, the versatility of a technical task in addressing various problem causes is a key consideration.

The algorithm employs initial probability θ and dispersion ε, as outlined in the study by Mihaylov and Tijdens (2019) to form the basis of its calculations. These initial values, specific to each occupation, are instrumental in determining the probability of technical competence. This is calculated using Equation 5, where $\sum_k^n T_k$ represents the sum of occurrences of one technical task, and $\sum_k^m T_k$ denotes the sum of all technical tasks.

Further, the probability of a technical task θ_{T_i} is utilized to ascertain the dispersion of the task ε_{T_i}, as per Equation 6. This leads to the calculation of the probability of technical competence $P_{TE(U_i|T_i)}$ for a specific cause, as detailed in Equation 7. This calculation involves taking the sum of occurrences of a cause for the technical task $\sum_i^n U_i$ and dividing it by the sum of all causes for the technical task $\sum_k^n U_k$.

Recent advancements in statistical learning and data analysis have emphasized the importance of such algorithms in extracting meaningful insights from complex datasets. Studies by Joe Qin et al. (2021) and Parks et al. (2020) highlight the growing application of statistical learning in various fields, including technical analysis and competence mapping. However, to the author's best knowledge, no study deals with the automatic mapping of described tasks to competencies using statistical methods. This lack underscores the SLA's potential and novelty in enhancing the accuracy and efficiency of competence identification in maintenance tasks, contributing significantly to the field of workforce optimization and task management.

$$\theta_{T_i} = \theta + \varepsilon * \frac{\sum_k^n T_k}{\sum_k^m T_k} \tag{5}$$

$$\varepsilon_{T_i} = \left(1 - \theta_{U_i}\right) \tag{6}$$

$$P_{TE(U_i|T_i)} = \theta_{T_i} + \varepsilon_{T_i} * \frac{\sum_i^n U_i}{\sum_k^n U_k} \tag{7}$$

$$P_{EC(U_i|T_i)} = \omega_{EC}(100 - P_{TE(U_i|T_i)}) \tag{8}$$

Based on the probability of technical competence (Mihaylov & Tijdens, 2019), SLA extends its analysis to encompass communication $P_{EC(U_i|T_i)}$and problem-solving $P_{PS(U_i|T_i)}$ competences, as delineated in Equation 8. A critical boundary condition in this context is $\omega_{EC} + \omega_{PS} = 1$, where $\omega_{EC}$and $\omega_{PS}$are the respective weights for communication and problem-solving, derived from the framework established by Mihaylov and Tijdens (2019).

This approach allows for a nuanced translation of general technical competencies into specific internal company competencies. This translation process necessitates the construction of a matrix of weights, $M_{TE\,(U_i|C_i)}$, as specified in Equation 9. In this matrix, each column corresponds to a particular cause, while the rows represent the company-specific technical sub-competencies. The weights within this matrix, ω_{STE_i,U_i}, are meticulously determined through iterative processes involving domain experts, managers, and shop floor personnel across various workshops.

The product of this matrix, $M_{TE\,(U_i|C_i)}$, and the probability of technical competence, $P_{EC(U_i|T_i)}$, yields a value between 0 and 1, representing the level of technical competence,

$L_{TE}(U_i|C_i)$, as shown in Equation 10. To align these continuous result values with the ordinal scale of company-specific competencies, a step size s is employed, as detailed in Equation 11.

$$M_{TE}(U_i|C_i) = \begin{bmatrix} \omega_{STE_i,U_i} & \cdots & \omega_{STE_j,U_i} \\ \vdots & \ddots & \vdots \\ \omega_{STE_i,U_j} & \cdots & \omega_{STE_j,U_j} \end{bmatrix} \begin{bmatrix} \tau_{U_i} \\ \vdots \\ \tau_{U_n} \end{bmatrix} \quad (9)$$

$$L_{TE}(U_i|C_i) = P_{TE}(U_i|C_i) * M_{TE}(U_i|C_i) \quad (10)$$

$$L_{TE}(U_i|C_i) \rightarrow \begin{cases} L_{STE_i} < s \rightarrow Level\ 1 \\ s \leq L_{STE_i} < s * 2 \rightarrow Level\ 2 \\ L_{STE_i} \geq s * 2 \rightarrow Level\ 3 \end{cases} \quad (11)$$

Recent advancements in the field of competence modeling and workforce analytics (Akimov et al., 2023; Khang et al., 2023; Treviño-Elizondo & García-Reyes, 2023) have underscored the importance of such detailed and iterative approaches in competence assessment. A study by Hernandez-de-Menendez et al. (2020) highlights the integration of statistical learning in competence analysis, emphasizing the need for precision and adaptability in competence models. This study advocates for the incorporation of diverse inputs from various organizational levels to enhance the accuracy and relevance of competence assessments.

Furthermore, the application of weighted matrices in competence evaluation, as explored by Kohl and Ansari (2023) demonstrates the effectiveness of this approach in capturing the multifaceted nature of competencies. This method allows for a more dynamic and context-specific assessment, aligning competencies with organizational needs and individual capabilities.

4.3.3 Anomaly Detection

Anomaly detection is essential to map sensor anomalies to certain maintenance tasks necessary to resolve these anomalies, or the underlying equipment failure. As the primary focus of this thesis was competence-based planning using maintenance task information, the anomaly detection follows state-of-the-art methods described by Goodfellow et al. (2016). The anomaly detection process consists of four steps: (i) constructing an autoencoder, (ii) model training, (iii) detection of anomalies, and (iv) model evaluation (Pang et al., 2022).

- **Constructing an Autoencoder:** Autoencoder should have an input layer matching the number of features in the data, several hidden layers for encoding and decoding, and an output layer that reconstructs the input data. For model compilation, an optimizer like Adam is used for efficient training (Hastie et al., 2001). The loss function is calculated based on the mean squared error (MSE), as it effectively measures the reconstruction error. A prerequisite for the autoencoder is the defined Euclidean space $X = \mathbb{R}^m$ and $Z = \mathbb{R}^n$ with $m > n$ and the encoder and decoder. For $x \in X$ $x = E_\phi(x)$ is written. Formally, the autoencoder E_ϕ can be defined as a multilayer perceptron (10), a form of neural network where is an element-wise activation function, W is a weight matrix, b is a bias vector, and x the independent variable.

$$E_\phi(x) = \sigma(W_x + b) \tag{12}$$

- **Model Training:** In the model fitting phase, the autoencoder is trained on the training dataset, where it learns to reconstruct the input data, thereby using the same data as both the input and the target output. Concurrently, a subset of the data is utilized as a validation set during the validation phase. This practice is essential for monitoring the model's performance and mitigating the risk of overfitting, ensuring that the model generalizes well to new, unseen data. The training function (11) of the autoencoder is a simple tuple of two functions, quality and task. Where the task is defined by a distribution μ_{ref} over X and a reconstruction quality $d(x, x')$. Therefore, how much x differs from x can be measured. With those, the loss function can be defined as.

$$L(\theta, \phi) = E_{x \sim \mu_{ref}}[d(x, D_\theta(E_\phi(x)))] \tag{13}$$

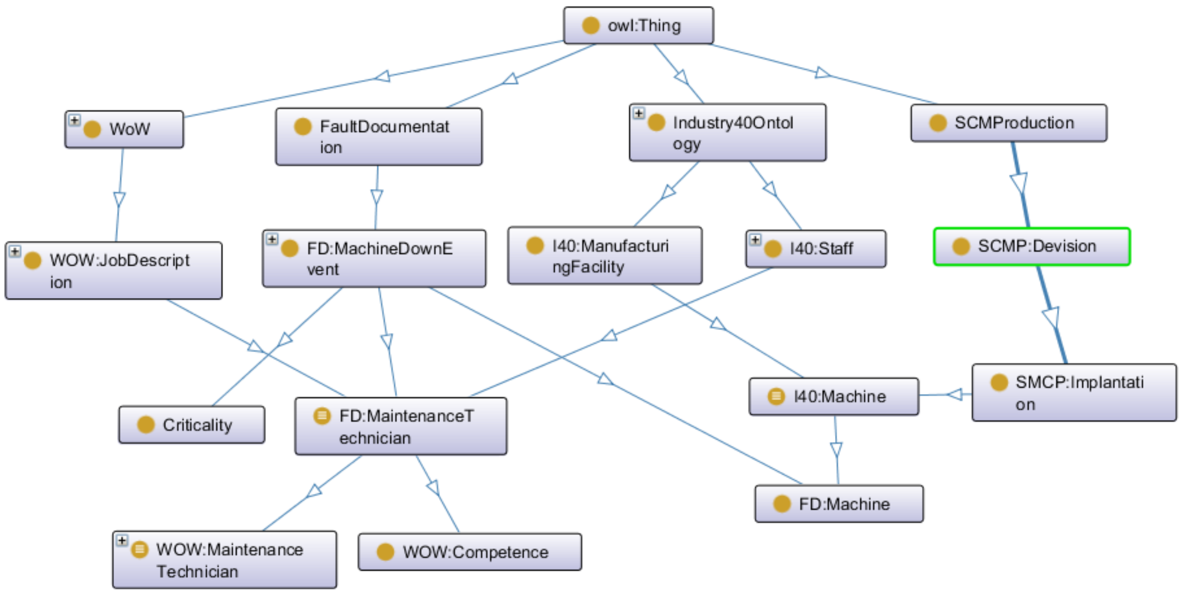
- **Detecting Anomalies:** For detecting anomalies, the MSE between the test data MSE_{test} and its reconstruction $MSE_{reconstruction}$is calculated. The error function F(x), as presented in Equation 12, can then be interpreted, where a specific value between $[0..1]$ is chosen to indicate an error. The formulation of the approach is as follows:

$$F(x) = \frac{MSE_{reconstruction}}{MSE_{test}} \tag{14}$$

4.4 FU4: Semantic Data Storage

The semantic storage module is based on a preprocessed collection of multi-structured data and expert input to label key characteristics. The first step involves establishing the formal framework used for mapping the preprocessed data. This framework is based on three important ontologies: the Industry 4.0 Context Ontology (I40CO) (Giustozzi et al., 2018), the World of Work Ontology (WOW) (Ansari, Khobreh, et al., 2018) and the Fault Documentation Ontology (FD) (Woods et al., 2023).

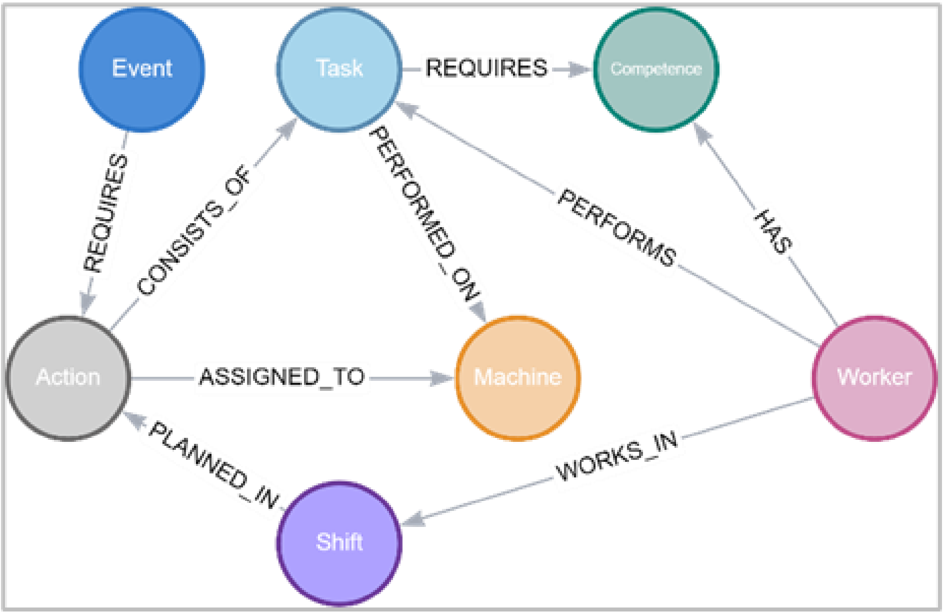
Figure 40: *Overview of I40CO, FD, and WOW*



The second step involves utilizing semantic extension—that is, using extracted keywords through named entity recognition and mapping them to classes within the chosen domain ontologies. These ontologies, selected and created with shared concepts, ensure coherence across different data types.

For example, machine-related concepts are consistently linked across all ontologies. A full merge technique is employed for ontology integration, where similar concepts are merged to resolve interoperability issues, and then the three ontologies are combined along those mapped concepts (i.e., entities). The I40CO is used for general manufacturing concepts, while the WOW covers processes, tasks, workers, and competencies. Depending on the application, additional ontologies or modifications to the I40CO may be necessary, as illustrated in Figure 40. In the third step, sensor and text data are utilized to populate the class properties of the individuals in the KG. The classes identified in the previous step are interconnected using formal relationships defined by verbs, and their properties are determined using available data, such as structured machine data and technical documents.

Figure 41: *Connected Main Concepts Implemented in Neo4j for Population (Kohl & Ansari, 2023)*



The preprocessed and analyzed data is then stored in a KG, implemented in Neo4j (Neo4j, 2024), as shown in Figure 41.

Table 9: *Example of Industrial KG Applications*

Enterprise	Product	Highlight
Siemens	Siemens MindSphere[3]	• KG generation from maintenance orders • Uses SKOS thesaurus and formal ontology • Integrates language models for thesaurus expansion
Johnson Controls	OpenBlue Platform[4]	• Utilizes KG and BrickSchema • Integrates various building systems • Enables informed decision-making based on comprehensive building data
Schneider Electric	Stardog[5]	• Leverages GraphDB with BrickSchema • Connects disparate building system data • Facilitates predictive maintenance optimization

[3]Siemens Industries. Kai Industrial Knowledge Graph. Retrieved 09.03.2025, from https://www.siemens.com/global/en/company/stories/research-technologies/artificial-intelligence/artificial-intelligence-industrial-knowledge-graph.html

[4] Johnson Controls. Digital Vault. Retrieved 09.03.2025, from https://www.johnsoncontrols.co.th/en/digital-solutions/digital-vault

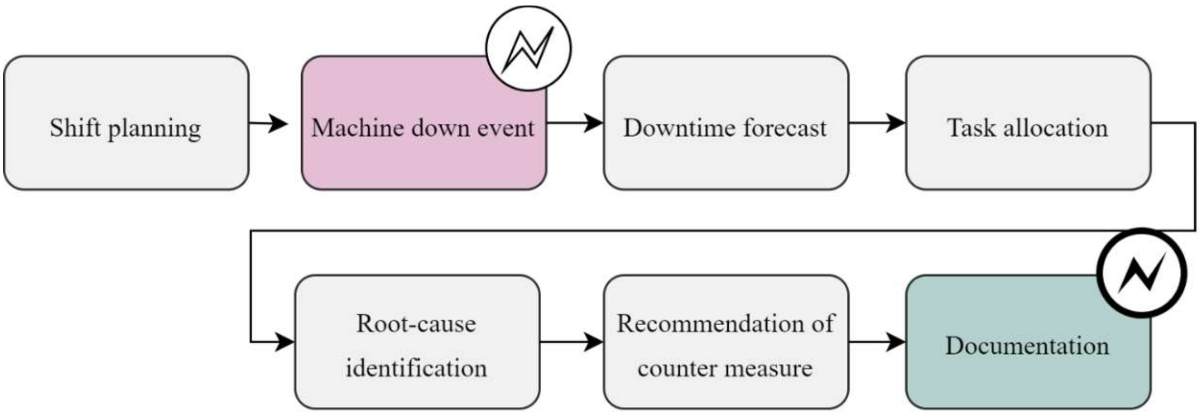
[5] Schneider Electric. Stardog. Retrieved 09.03.2025, https://www.stardog.com/company/customers/schneider-electric/

This KG serves as a flexible and scalable data repository, structuring data and their relationships. It models relevant entities such as machines, maintenance procedures, workers, and their roles and competencies, using unique resource identifiers for knowledge base connections. Semantic links between these entities enable complex queries, such as identifying required competencies for specific tasks. The KG's development involved a collaborative approach, combining expert knowledge from maintenance professionals with data-driven analysis of machine and DSB data. As Table 9 shows, this approach ensured a comprehensive and relevant KG structure, reflective of real-world maintenance scenarios.

4.5 FU5: AI-enhanced Decision Making

The AI-enhanced decision-making layer in ARCHIE aims to provide information for problem-solving tasks based on data from the previous layers. Typical problem-solving tasks in maintenance planning and operations include (i) shift planning, (ii) downtime forecasting, (iii) task allocation for finding the best-fit maintenance technician, and (iv) root-cause identification and subsequent recommendation of the correct solution, as illustrated in Figure 42. All tasks focus on recommending one piece of one or another. Moreover, the order of the applications also represents the temporal processing of the ARCHIE applications. The novel approach of ARCHIE is to include textual data as well as competence information from the SLA in combination with enterprise resource planning (ERP) data for these tasks.

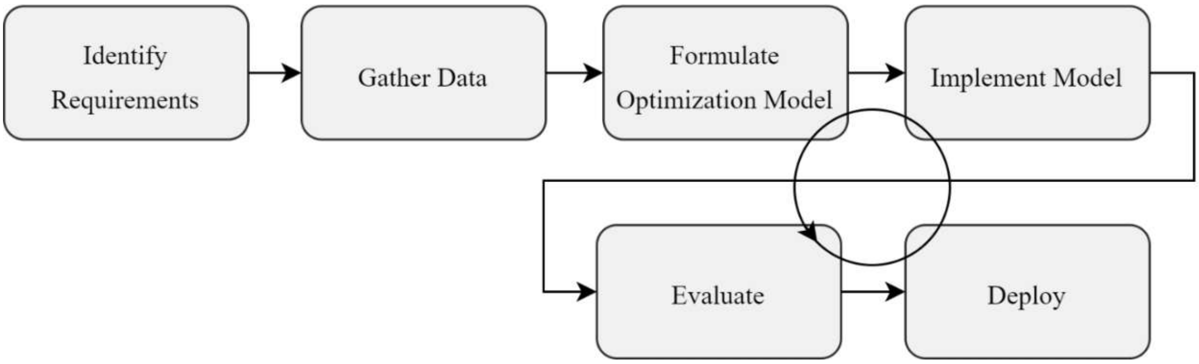
Figure 42: *Process Model of the FU AI-Enhanced Decision-Making*



4.5.1 Shift Planning

Shift planning in maintenance is the process of organizing and scheduling maintenance personnel into shifts to ensure continuous operational coverage and optimal resource utilization. It involves assigning technicians with the appropriate qualifications and competencies to specific time slots, considering factors such as equipment availability and workload balancing. Shift planning employs discrete optimization through linear programming, a method well-suited for tasks where computing time is not a primary concern. This approach aligns with the overall objectives of efficient shift scheduling and is implemented using the open-source Python library, SciPy, known for its robust capabilities in scientific computing (Mitsotakis, 2023). The steps depicted in Figure 43 are described in more detail below, with a focus on the formulation of the optimization model.

Figure 43: *Shift Planning Pipeline*



The shift planning algorithm considers a range of constraints, including the number of shifts, adherence to legal requirements such as overtime regulations, holiday entitlements, and sick leave provisions, as well as the required and available competencies of the workforce. These considerations are crucial for ensuring compliance and operational efficiency (Smith & Foster, 2019). The algorithm's design is informed by predictions from previous steps, integrating data-driven insights into the planning process.

A key aspect of this approach is the use of binary decision variables—namely, M_{ij}, E_{ij}, N_{ij}, representing the allocation of workers to morning, evening, and night shifts, respectively—and for maintenance technicians on leave or sick. This binary framework allows for a clear and concise representation of shift assignments (Jones et al., 2021).

The objective function, which the algorithm seeks to minimize, is the sum of the vectors $\boldsymbol{WP}$ and $\boldsymbol{WS}$ (see Equations 15–18). These vectors represent the weighted competencies of the workers, both primary $\boldsymbol{wp}$ and secondary $\boldsymbol{ws}$, multiplied by the sum of the shifts they are working. This formulation, as detailed in Equations 15 and 16, aims to optimize the allocation of competencies across shifts, ensuring that each shift has the necessary skill sets while also considering worker availability and preferences (Chen et al., 2021).

$$\boldsymbol{WP} = \boldsymbol{wp}\left(M_{ij} + E_{ij} + N_{ij} + L_{ij}\right) \qquad (15)$$

$$\boldsymbol{WS} = \boldsymbol{ws}\,(M_{ij} + E_{ij} + N_{ij} + L_{ij})) \qquad (16)$$

$$min\left(\sum_{i=0}^{n}\sum_{j=0}^{m}\boldsymbol{WP} + \boldsymbol{WS}\right) \qquad (17)$$

$$\sum_{i=0}^{n}\sum_{j=0}^{m} M_{ij} + E_{ij} + N_{ij} + L_{ij} = 1 \qquad (18)$$

Recent studies in optimization algorithms and workforce management have highlighted the importance of incorporating fairness, flexibility, and worker preferences into shift scheduling, leading to improved worker satisfaction and operational efficiency (Gaal et al., 2023; Uhde et al., 2020). Additionally, the integration of ML models for predicting workforce availability and competence requirements has further enhanced the effectiveness of shift planning algorithms (Xiong Wang et al., 2022).

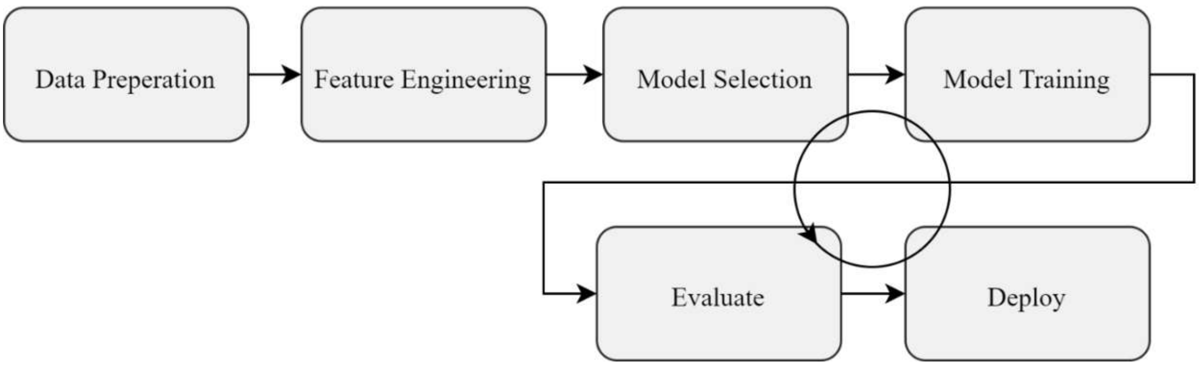
In the context of workforce scheduling, the primary objective is to attain an optimal mix of competencies across various shifts, ensuring that the required competencies for each shift are met without assigning overqualified personnel. This approach aims to maximize workforce efficiency while avoiding the underutilization of highly skilled workers. The challenge lies in balancing the competence requirements of each shift with the optimal utilization of the workforce's skill set, constrained by legal requirements, as shown in Equation 16.

The goal is to achieve an optimal competence mix by satisfying required competencies while not scheduling overqualified personnel for respected shifts. Additionally, constraints ensure that a worker, for instance, can only be assigned to one shift per day, as shown in Equation 15.

4.5.2 Downtime Forecast

Downtime forecasting in maintenance is the process of predicting future machine downtimes when assets, machines, or equipment will be unavailable due to scheduled maintenance or unscheduled maintenance, i.e., breakdowns. The focus in downtime prediction is to forecast the expected downtime of assets as promptly as possible, typically following the generation of a machine fault report. This forecasting is crucial for efficient maintenance planning and potentially reordering production schedules to minimize impact (Ansari et al., 2021).

Figure 44: *Downtime Forecast Pipeline*



The downtime forecast pipeline, as shown in Figure 44, starts with the initial preparation by extracting key textual elements (e.g., nouns, noun chunks, and verbs). This is followed by feature engineering, which involves the vectorization of noun chunks and verbs from the machine fault reports. This transformation into a machine-readable format is essential for describing and converting the contained knowledge into a form suitable for ML models. Recent advancements in TLP have highlighted the effectiveness of vectorization in enhancing the accuracy of predictive models (Smith et al., 2021; Johnson & Zhang, 2019).

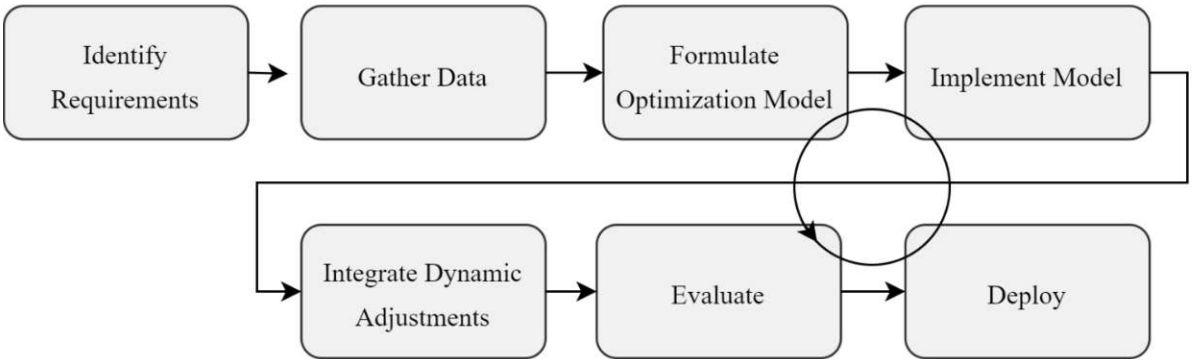
Once vectorized, these machine fault reports and associated downtime data serve as training input for various ML models, with downtime being the target variable. In addressing this regression problem, two vector representations are commonly employed: BoW and TF-IDF. While BoW offers a straightforward approach that can lead to high-dimensional data, TF-IDF provides a more nuanced representation by considering word frequencies (Pennington et al., 2014).

For the regression models, a range of ML algorithms are optimized, including random forest, k-nearest neighbor, support vector machine, and DNNs. Each of these models has its strengths and must be chosen based on the specific characteristics of the data and the problem at hand. Random forest is known for its robustness and ability to handle non-linear data, while k-nearest neighbor is effective for smaller, less complex datasets (Hastie et al., 2001). Support vector machine is preferred for its effectiveness in high-dimensional spaces (Jieyu Xie & Wan, 2023), and DNNs are utilized for their ability to model complex, non-linear relationships in large datasets (Deutsch et al., 2017).

4.5.3 Task Allocation

Task allocation in maintenance is the process of assigning specific maintenance tasks to individual technicians or teams based on their qualifications, competencies, and availability. It involves matching the requirements of each task with the appropriate technicians to ensure efficient execution, optimal resource use, and reduction of MTTR. In this thesis, task allocation was conducted through the use of genetic algorithms (GAs), which are emphasized for their efficacy in generating solutions that adhere to stringent time constraints. This capability is particularly advantageous in shift management, where GAs facilitate two critical functions: (i) the assignment of planned tasks at the commencement of the shift and (ii) the agile rescheduling of tasks as ad-hoc requirements arise. The implementation of GA for this purpose is adeptly handled using PyGAD, a renowned open-source Python library that specializes in GA (Gad, 2023). The task allocation pipeline, as illustrated in Figure 45, comprises seven sub-steps, the first two of which concern the collection of requirements and data. The following discussion focuses on the formulation of the optimization model, its implementation, and its dynamic adjustment.

Figure 45: *Task Allocation Pipeline*



$$\text{Minimize } \sum_{w\in W}\sum_{j\in J} x_{wj} z_{wj} \quad (19)$$

$$\text{begin subject to } \sum_{w\in W} = z_{wj}, \forall j \in J \quad (20)$$

$$\sum_{j\in J} = \boldsymbol{c_{wj}} * r_{wj} z_{wj} \leq b_w, \forall w \in W \quad (21)$$

$$z_{wj} \in \{0,1\}, \forall w \in W, \forall j \in J \quad (22)$$

In this setting, the set of workers, denoted as $j = \{1,2,\ldots,n\}$, is dynamically adjusted by excluding those already assigned to tasks, thereby optimizing the allocation process. Concurrently, the set of jobs, represented as $j = 1,2,\ldots,n$, is defined for each worker $w \in W$ and job $j \in J$.

The algorithm incorporates several additional variables to refine the task allocation process:

- x_{wj}: This variable represents the cost of assigning job j to worker w. It encompasses various factors such as time, effort, or resource utilization, which are crucial for optimizing task allocation.
- r_{wj}: This indicates the resources available for worker w to perform job j, ensuring that tasks are assigned based on resource availability.
- $\boldsymbol{c_{wj}}$: This denotes the competence required by worker w to perform job j, aligning job requirements with worker skills.
- b_w: This represents the resource capacity of worker w, considering the maximum workload a worker can handle.
- z_{ij}: This binary variable indicating the occupation status of the worker, essential for tracking task assignments.

Equations 21 and 20 in the model ensure that each job is assigned to only one worker, thus preventing the over-exceedance of total resources. Moreover, Equations 21 and 22 establish necessary boundary conditions for the resource capacity and occupation status of workers, ensuring a balanced and feasible allocation of tasks.

Lastly, the individual workers and their properties are updated based on feedback from executed shifts and task allocations. This step is crucial as it incorporates the concept of balanced manpower planning, which takes the workload and experience increment per worker into consideration. Such an approach fosters continuous improvement in planning results, as it adapts and evolves based on real-world feedback and outcomes (Abidi et al., 2022; Converso et al., 2023).

This dynamic and feedback-oriented approach to task allocation, underpinned by the robust capabilities of GAs, represents a significant advancement in shift management, not only optimizing task allocation based on current requirements but also evolving to continually enhance planning efficacy.

In the task allocation model, Equations 17 and 18 ensure that each job is assigned to only one worker, effectively managing resource utilization and preventing over-exceedance. These equations are crucial for maintaining a balanced workload and align with recent research in workforce optimization (Reinhold et al., 2019).

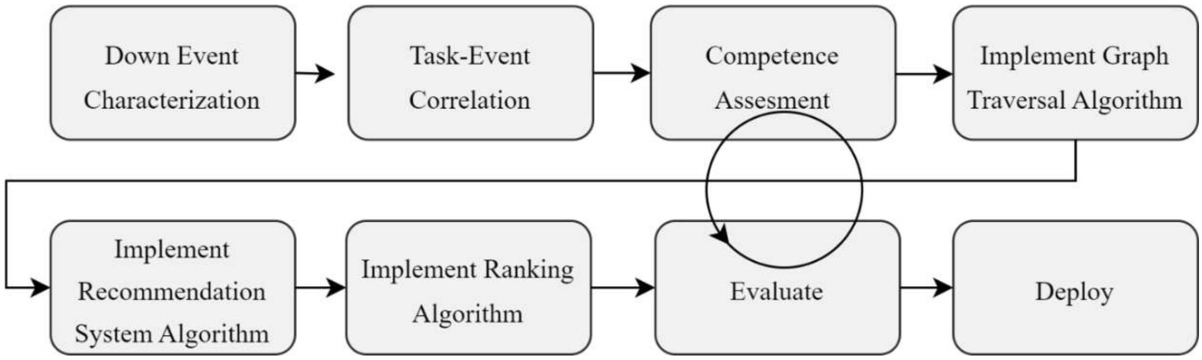
Equations 21 and 22 establish boundary conditions for the resource capacity and occupation status of workers. While Equation 21 ensures that the resource capacity of each worker is not exceeded, aligning with studies on sustainable workload management, Equation 20 manages the binary occupation status of workers, simplifying the complexity of task allocation.

These equations, integral to the GA framework, demonstrate the model's robustness in handling workforce scheduling challenges, ensuring efficient, equitable, and sustainable task allocation.

4.5.4 Root Cause Identification

Root-cause identification in maintenance is the process of determining the fundamental underlying cause of a fault or problem in equipment or systems, which involves analyzing incidents to pinpoint the primary issue that must be addressed to prevent recurrence. For this, the KG of the semantic data storage is required in combination with the algorithmic recommendation of maintenance tasks. This process is predicated on the structured interrelations among entities such as workers, machines, down events, actions, tasks, and competencies within the KG. The recommendation mechanism is a multi-step process involving graph-based querying and algorithmic reasoning, as follows (Xiang Wang et al., 2019). The root-cause identification pipeline comprises eight sub-steps, as illustrated Figure 46. The subsequent discussion focuses on the first six steps.

Figure 46: *Root-Cause Identification Pipeline*



Upon the occurrence of a machine down event, the KG is interrogated to extract pertinent data, encompassing the specific machine type, the nature of the malfunction, and any historical precedents. This interrogation is formalized through a query $Q_d$that retrieves a subset of the graph G relevant to the down event. The KG encapsulates a mapping of maintenance tasks to various down events. A query Q_t is executed against G to enumerate potential tasks T that are pertinent to the identified down event. Each worker node in the KG is annotated with competence attributes. A competence matching query Q_c is employed to ascertain workers W whose skill sets align with the requirements of the tasks T identified in the preceding step.

Next, the implementation of different algorithms—namely, graph traversal and recommendation algorithms—aligns suitable tasks with qualified workers, as well as a ranking algorithm to provide ranked recommendations. This involves three primary algorithmic approaches:

Figure 47: *Depth-First Search Algorithm*

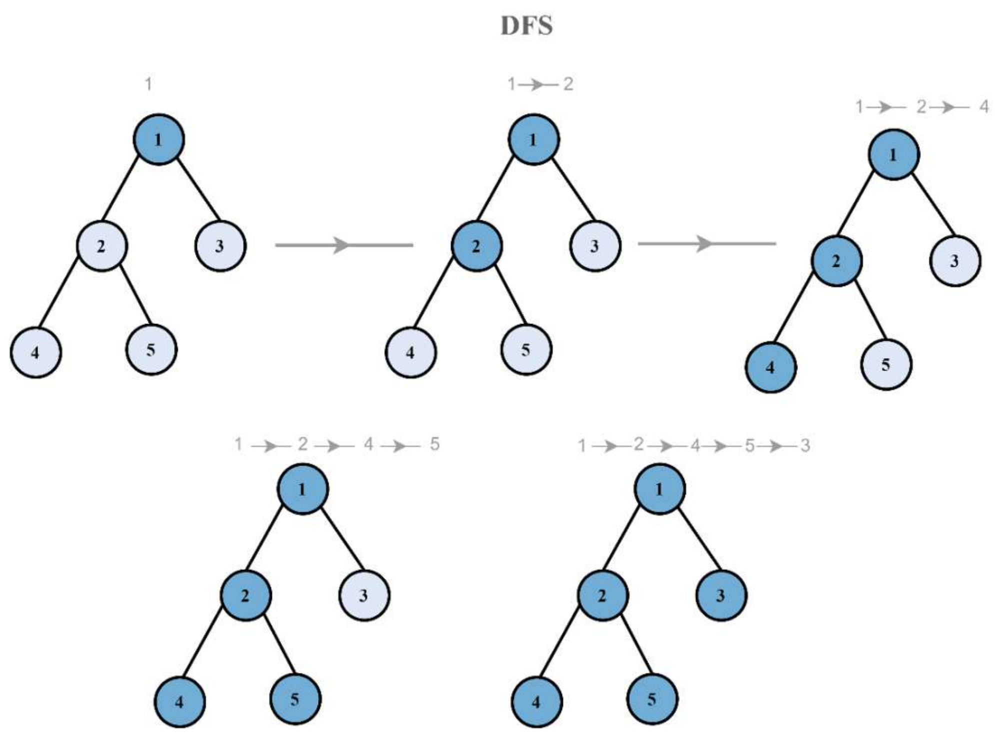
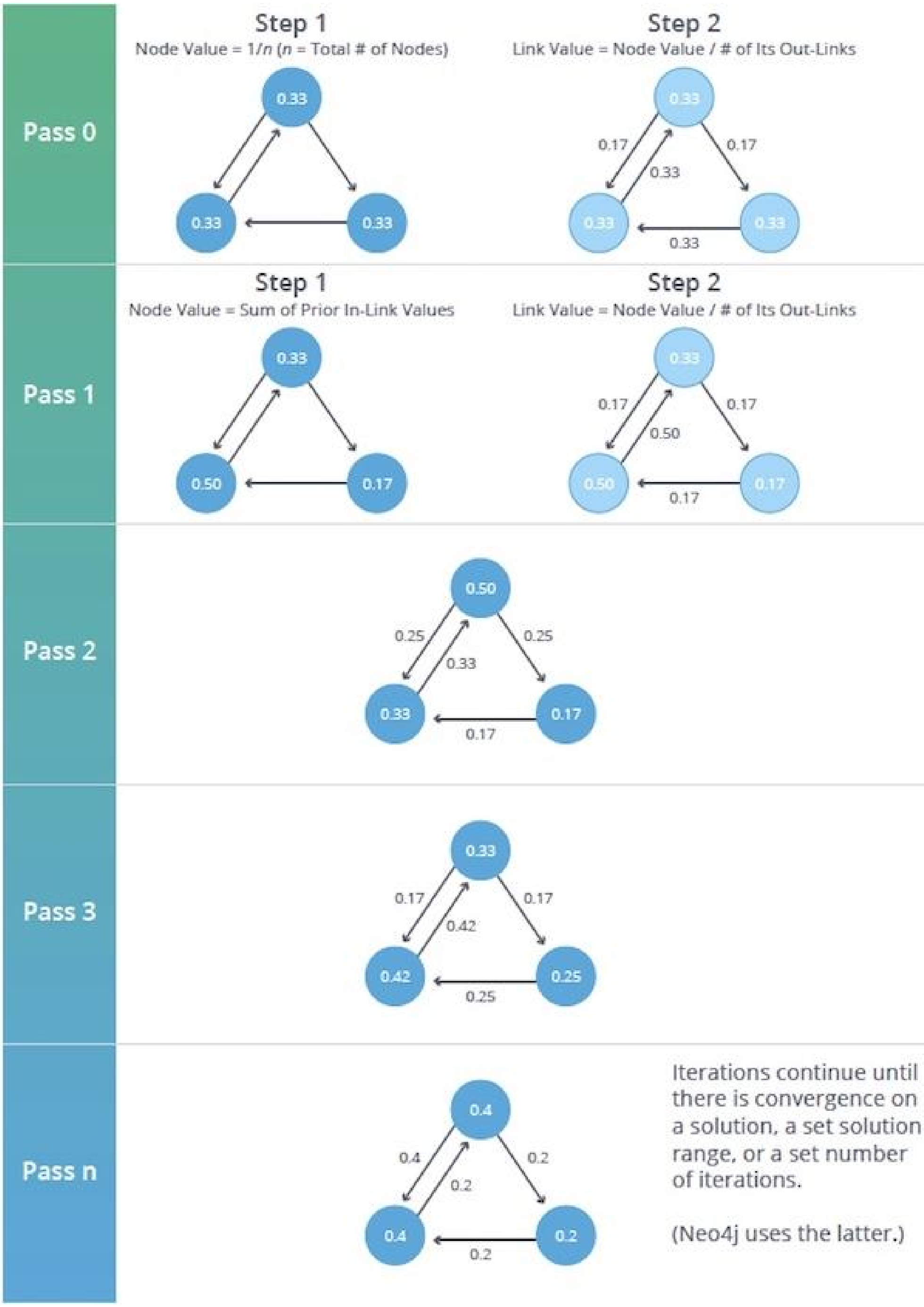


Figure 48: *Procedure of the PageRank Algorithm*



- **Graph Traversal Algorithms:** Algorithms such as depth-first search (DFS), as shown in Figure 47, or breadth-first search (BFS) are utilized for graph navigation. DFS can be represented as a recursive function $DFS(v, G)$, where v is a vertex and G is the graph. These are explored as far as possible along each branch before backtracking, thereby mapping the connections between down events, tasks, and worker competencies.
- **Recommendation System Algorithms:** To match tasks with workers, algorithms similar to collaborative filtering or content-based filtering are adapted. In the KG context, a content-based approach can be particularly effective, leveraging semantic

relationships. This can be mathematically represented as a function $R(u, i)$, where u is a user `worker` and i is an item `task`, with the output being a relevance score based on the user's profile and item attributes.

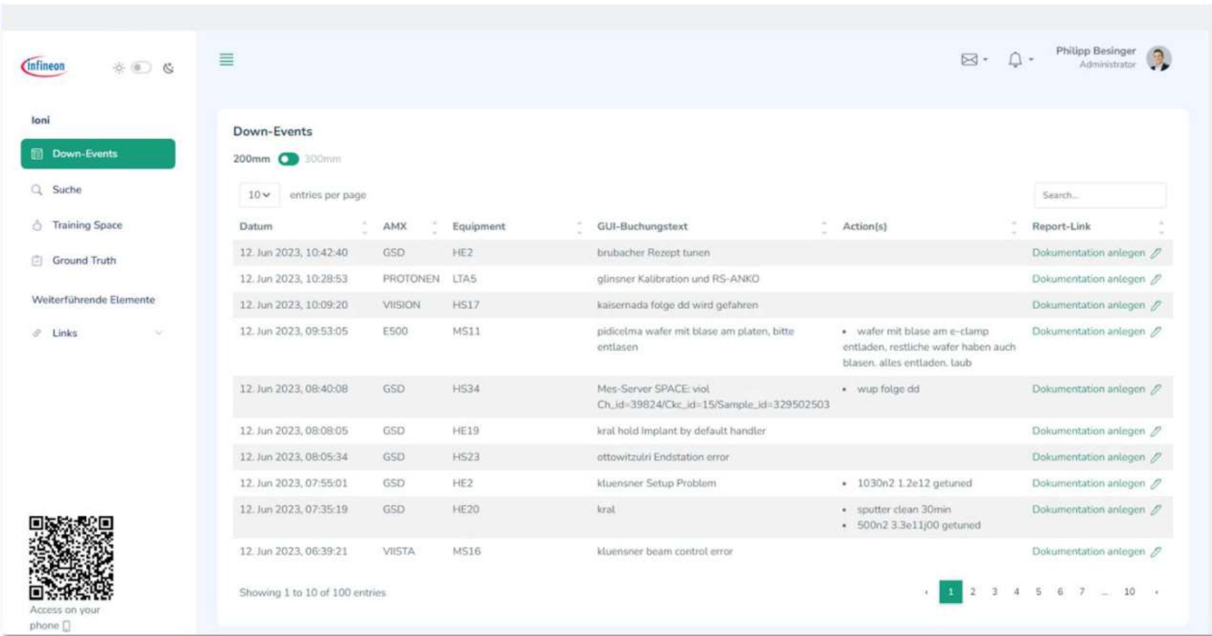
- **Ranking Algorithms:** To prioritize *tasks*, algorithms like PageRank are employed. PageRank is defined $PR(u) = \frac{1-d}{N} + d \sum_{v \in B_u} \frac{PR(v)}{L(v)}$, where $PR(u)$is the PageRank of task u, N is the total number of tasks, B_u is the set of tasks that link to u, $L(v)$ is the number of outbound links on task v, and d is a damping factor, as illustrated in Figure 48.

By integrating these algorithmic methodologies, the KG becomes a potent tool for the recommendation of maintenance tasks, ensuring optimal task allocation to workers based on their competencies and the specific requirements of the maintenance scenario. The final steps are then evaluation and deployment.

4.6 FU6: Human-Interface

The human interface (HI) layer is the central interface for the human users of ARCHIE. The interface is designed to meet the needs of maintenance planners and technicians, as noted by Cimini(Cimini et al., 2020).

Figure 49: *Basic Structure of ARCHIE Human Interface*

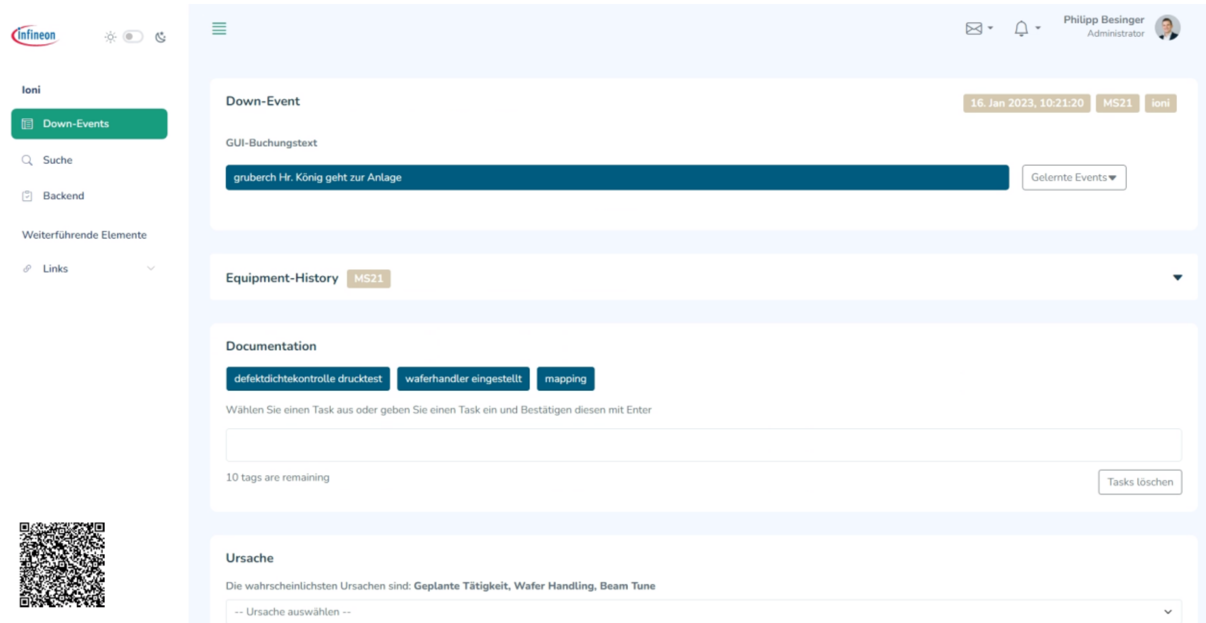


This layer is multifaceted and serves several functions. First, it assists maintenance planners by providing comprehensive shift schedules, thereby streamlining the scheduling process (Ansari et al., 2023). Additionally, it provides both maintenance planners and technicians with real-time access to critical information such as machine status and KPIs, thus improving decision-making and operational efficiency. The ARCHIE HI consists of three primary building blocks: a list of down events, a semantic search, a planning space, and a training space, as illustrated Figure 49.

4.6.1 Down Events

The down events section provides a list of all currently open planned and unplanned maintenance orders. A distinction is made according to production type and plant. Each open down event can be selected by maintenance planners or technicians. However, ARCHIE automatically suggests the most suitable maintenance technician. This technician is supported by suggestions for possible root causes during problem-solving and later documentation of countermeasures.

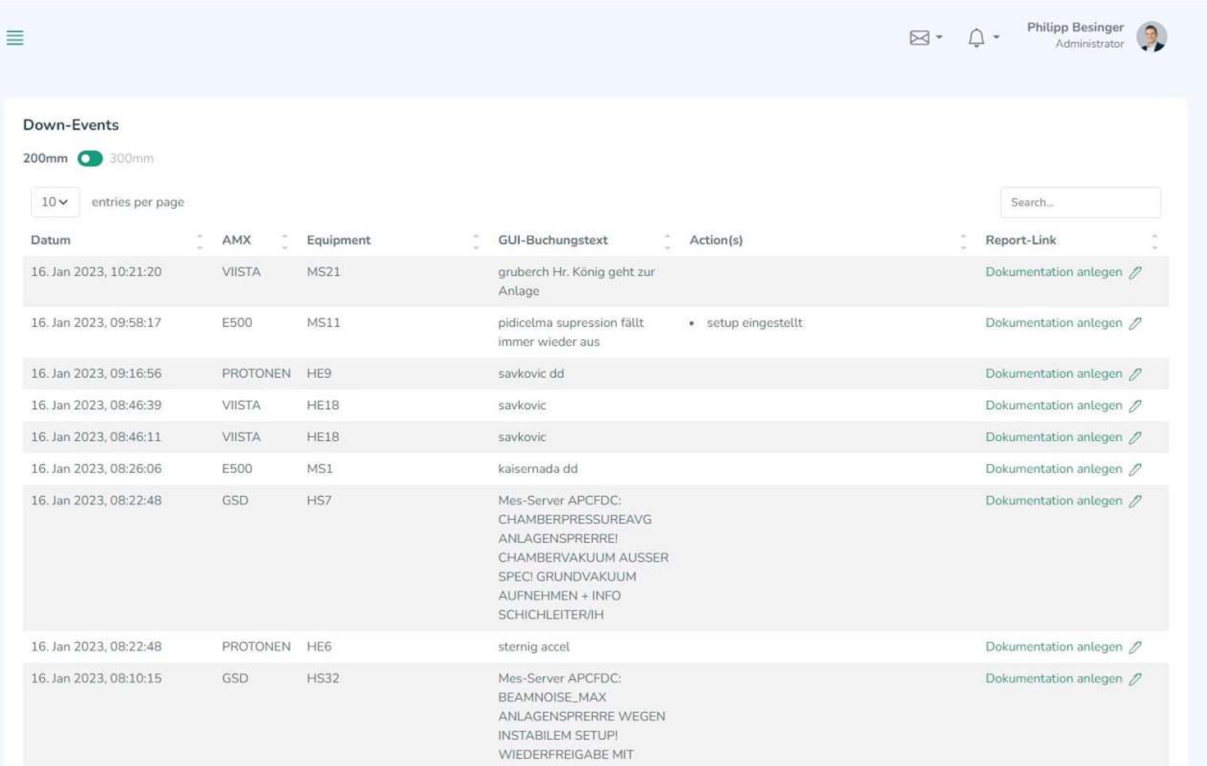
Figure 50: *Basic Structure of ARCHIE Down Event Section*



4.6.2 Search

The search function is an important tool for deeper root cause assignment, as it provides a fine-grained final conclusion about the entire stored data, as shown in Figure 51. This allows field service technicians to find precise background information or related cases for the current problem. By using the search function, maintenance technicians leverage the semantic data storage. The KG interlinks countermeasures with uniform resource identifiers in the enterprise wiki, allowing maintenance technicians to seamlessly access a wealth of expert knowledge and detailed procedures stored within the enterprise wiki, directly related to each proposed task (Kohl & Ansari, 2023).

Figure 51: *ARCHIE Search*



4.6.3 Planning Space

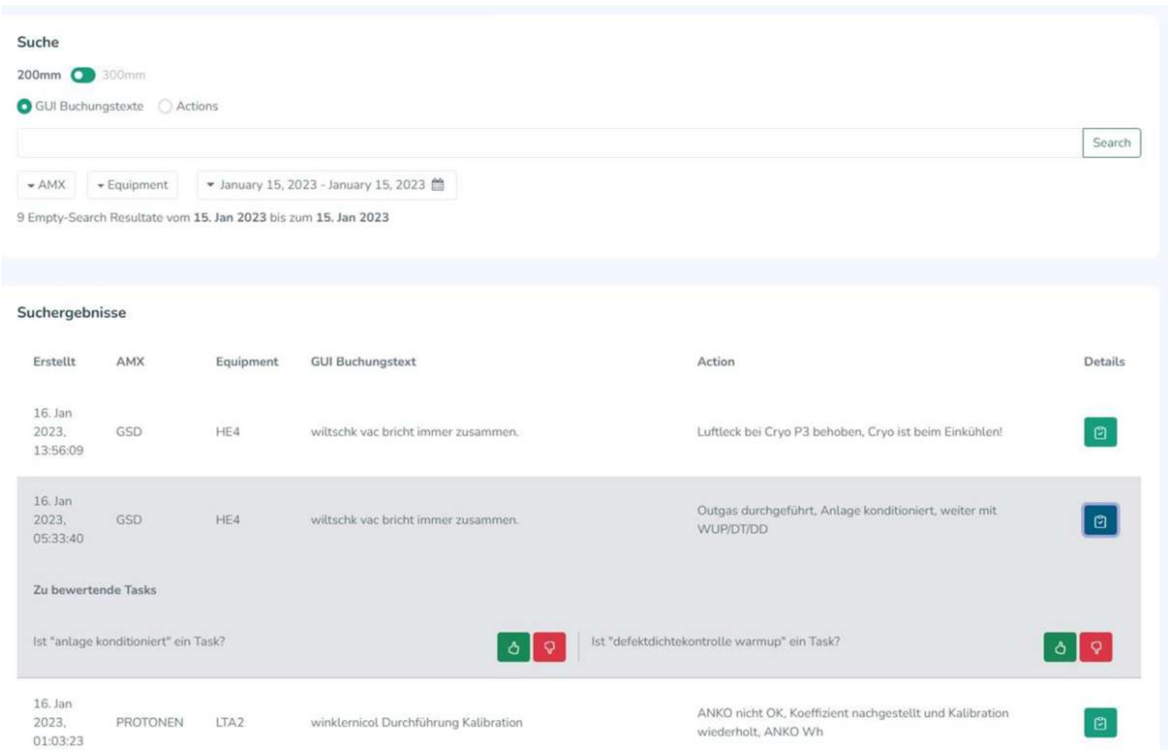
The planning space provides an overview of equipment and personnel availability, skill distribution, and planned jobs for each maintenance technician, as shown in Figure 52. This layer serves numerous functions. First, it assists maintenance planners by providing comprehensive shift plans, streamlining the scheduling process (Ansari et al., 2023). Further, it offers real-time access to critical information, such as machine status and KPIs, to both maintenance planners and technicians, thus enhancing decision-making and operational efficiency. The planning space also provides detailed information on available competencies, competence distributions, and competence development for the available maintenance technicians (Kohl, Fuchs, Berndt, et al., 2021).

Figure 52: *ARCHIE Planning Space*

4.6.4 Training Space

The training space allows users to influence the TMP and SLA algorithms, as shown in Figure 53. Here, planners and technicians can correct terms and competencies, thus influencing the weights in the respective algorithms. Since a cascading approach is also used for the TMP when utilizing large language models, users of maintenance chatbots can also use the training space. A crucial part of applying ARCHIE is customizing the HI to the specific needs of the domain or enterprise (Bhattacharya et al., 2023).

Figure 53: *ARCHIE Training Space*



This involves understanding the roles, competencies, and workflows of maintenance planners and technicians. User interviews, surveys, and observations can be crucial in this phase. It can be reasonably deduced that the following steps must be carried out (GRLER, 2022; Pohl & Rupp, 2021).

- **Defining Functional Requirements:** Based on the user analysis, the functional requirements of the HI are defined. For an enterprise in the semiconductor industry, this might include real-time display of machine status, maintenance scheduling, root cause analysis workflows, and access to an enterprise wiki.
- **Creating User Personas and Scenarios:** User personas representing different user groups, such as maintenance planners and technicians with varying competence levels, must be designed. Based on these personas, scenarios that depict how they would interact with the system can be created.
- **Sketching and Wireframing:** In the initial step, low-fidelity sketches and wireframes of the HI are created to outline the layout of the interface, including the placement of elements like task lists, machine status indicators, and navigation menus.
- **Incorporating User Feedback:** A main focus must be on conducting usability testing with actual users or representatives of the user groups. The feedback gathered on the usability and functionality of the prototype must be iterated back into the design.
- **Final Interface:** After several iterations and refinements, the final HI can be designed. This should include high-fidelity visual elements, consistent color schemes, and typography that aligns with the ergonomic and aesthetic requirements of the users.
- **Accessibility and Compliance:** It must be ensured that the HI is accessible to all users, including those with disabilities, and complies with relevant legal and ethical standards, such as data protection and privacy laws.

4.7 Interim Conclusion: A Framework for Cognitive Assistance Systems

The framework presented in this chapter establishes a robust structure for the integration of CAIS in Industry 4.0 environments. ARCHIE is carefully designed to incorporate advanced AI and ML techniques, addressing the research questions presented in the thesis. This thesis focused on the selected gray box approach, which emphasizes the white box storage and reasoning capabilities of traceable semantic data, and on the black box-oriented ML approaches at the edges, which are used for processing heterogeneous data sources from virtual and physical sensors. It describes methods for extracting actionable insights from textual data, thereby improving the efficiency and effectiveness of maintenance planning. By integrating competency-based shift scheduling and task assignment, the framework ensures optimal use of HR by aligning maintenance activities with personnel expertise and availability.

To address the research gap identified in Section 4.3, the framework proposes a modular and adaptable architecture to facilitate seamless integration and real-time application in dynamic industrial environments. However, empirical validation and iterative refinement through practical implementation remain critical. Therefore, the framework emphasizes the importance of error-free and time-saving documentation processes, which directly addresses research questions regarding the support and efficiency of maintenance reporting systems. This comprehensive approach provides not only theoretical foundations but also practical guidelines for implementing cognitive maintenance systems and highlights the need for further research to validate and improve the proposed methods in real-world applications.

A detailed examination of the basic functionalities of ARCHIE, the AI methods used, and the impact of the respective FUs on maintenance is shown in Table 10. In summary, it can be said that it is primarily the FUs of data analysis, semantic data storage, and AI-enhanced decision making that make ARCHIE so innovative.

Table 10: *Overview of Core Functionality, AI Methods Used, and Implications for Maintaining ARCHIE*

Method	Core Function	AI Method Used	Impact on Maintenance
Data Acquisition	Collection of heterogeneous data sources	No AI method applied	Interfaces open system for maximum generalizability
Data Preprocessing	Preparing raw data for algorithmic processing	Statistical methods such as feature extraction and principal component analysis	Automatic preparation of data for the next step in the analysis process
Data Analysis	Extracting information and detecting anomalies in virtual and physical sensor signals	Combination of TLP, DL, and SLA	Linking of textual information for maintenance planning and in combination with anomaly detection for root cause identification.
Semantic Data Storage	Semantically related information storage	Semantic storage of extracted information in a KG	Enables heterogeneous data sources to deliver semantically linked knowledge
AI-enhanced Decision Making	Reasoning and recommendation of information	DL in combination with reasoning capabilities	Combination of conclusions with subsequent competency-based action recommendations
Human Interface	Displaying of information	No AI method applied	Competence-based provision of information

5 Application and Evaluation in the Semiconductor Industry

The application of ARCHIE is demonstrated in the context of semiconductor manufacturing. The semiconductor industry is at the forefront of technological innovation, propelling advancements in electronics, computing, telecommunications, and a plethora of other sectors. The semiconductor industry is distinguished by its rapid innovation cycles, high capital investment, and intricate manufacturing processes, which collectively demand a high level of precision and expertise operating on a 24/7 schedule. The production of semiconductor devices involves complex processes that transform raw silicon wafers into highly integrated circuits with millions or billions of transistors. This transformation requires not only advanced technological equipment but also a workforce with specialized competencies in various disciplines. Another specialty is the industry's rapid pace of technological evolution, often described by Moore's Law, predicts the doubling of transistors on a microchip approximately every two years. This relentless push for miniaturization and enhanced performance necessitates continuous innovation in manufacturing processes and equipment.

One of the primary specialties of the semiconductor industry is its reliance on ultra-clean manufacturing environments. Additionally, the industry's rapid pace of technological evolution, often described by Moore's Law, which predicts the doubling of transistors on a microchip approximately every two years. This relentless push for miniaturization and enhanced performance necessitates continuous innovation in manufacturing processes and assets, resulting in complex competence requirements for maintenance planners and technicians to maintain high asset availability levels. Moreover, the industry also operates under strict quality control measures, given the high cost of defects and the critical role semiconductors play in modern electronics. Semiconductor manufacturing is a sequential process that includes wafer fabrication, doping, photolithography, etching, deposition, and packaging. Each of these steps requires precise control and coordination to ensure the final product meets the desired specifications. Ion implantation is a critical process in semiconductor manufacturing used to modify the electrical properties of silicon wafers, which involves accelerating ions of dopant elements and implanting them into the substrate material. The ions penetrate the wafer's surface, creating regions of n-type or p-type semi-conductivity, which are essential for forming transistors and other semiconductor devices. Ion implantation offers precise control over dopant concentration and depth profiles, which is crucial for the fabrication of modern integrated circuits with nanometer-scale features. It allows for uniform doping over large wafer areas and can be finely tuned to achieve specific electrical characteristics.

Maintenance of ion implantation machines is particularly challenging due to their complexity and the precision required in their operation. Therefore, specialized competencies are needed to perform maintenance tasks effectively. Hence, technicians must have deep knowledge of high-vacuum systems, electromagnetic fields, high-voltage components, and materials compatibility. Unexpected downtime can have significant implications due to the high cost of equipment and the potential for production delays.

Furthermore, the integration of these machines into automated production lines means that maintenance personnel must also be proficient in control systems and software diagnostics. Predictive maintenance strategies are increasingly employed, utilizing data analytics and sensor information to anticipate failures before they occur, thereby minimizing unplanned downtime. Accordingly, plant availability in the semiconductor industry is the central KPI by which maintenance departments are measured, and employees are, therefore, evaluated by MTTR.

In the investigated case of ion implantation, textual data is stored in an application called eBook, which acts as a digital service backbone, and machines are continuously monitored via advanced process control (APC). The APC system monitors machine parameters in the background using external sensors and data from the machine control system. Afterward, it analyzes these data using statistical methods, especially time-domain signal processing. If any of the monitored parameters exceed their specified limit values, the system issues an automated warning, error message, or even a maintenance order. Therefore, the APC reports data back to the fault documentation and MES. In the presented use case, a distinction is made between planned and unplanned maintenance. The fixed schedule of planned PM is not based on time windows available at short notice but rather is rigidly applied throughout the year. Unscheduled maintenance can be triggered by machine operators or by the APC. A so-called machine down event is triggered if the APC or the machine operator detects anomalies or malfunctions. An anomaly, as defined in ISO 13372:2012, is a deviation from normal operating conditions that may indicate a potential issue, whereas a malfunction is an actual failure of equipment or a system to perform its required function, defined in ISO 13849-1:2023. Anomalies are identified using process indicators and associated corridors that are exceeded. Until now, each department (e.g., etching or ion implantation) has its own maintenance team. The maintenance staff assigns themselves via eBook to such a machine down event and repairs the machine fault, leading to unequal utilization of the maintenance staff at shift levels—that is, there are not enough maintenance technicians (e.g., in the ion implantation department), while not all are deployed in the etching department. On an individual level, there is a risk of overestimating or underestimating one's own competencies. At the factory level, ARCHIE enables shift planning, while on an operational level, it enables the assignment of machine down events to maintenance technicians based on their competencies rather than formal department allocation, as well as competence-based support through the identification of root causes and appropriate countermeasures. The described setting allows for competence-based planning of shifts based on planned maintenance and considering historical data of unplanned maintenance events while utilizing available technicians across departmental borders. ARCHIE improves maintenance in various areas through its gray-box approach, making an important contribution to AI-based maintenance and the aggregation of maintenance planners and technicians, as shown in Table 11.

Table 11: *Influence of ARCHIE on Maintenance*

Method	As-Is State	As-To-Be State
Shift Planning	Rigid, department-level scheduling; inflexible; potential understaffing during critical times	Flexible, factory-level planning; competence-based scheduling; optimal coverage across departments
Task Allocation	Self-assignment; risk of misjudging competencies; confined by departments	Competence-based task assignment; matches tasks to suitable technicians; improved efficiency
Competence Management	No systematic assessment; relies on individual initiative; possible skill mismatches	Systematic competence evaluation; tailored assignments; supports skill development
Downtime Management	Inefficient response; potential delays due to self-assignment and silos	Proactive technician assignment; reduced MTTR; improved asset availability

Method	As-Is State	As-To-Be State
Prescriptive Analytics	Limited integration of prescriptive maintenance; reactive actions dominate	Full integration of prescriptive analytics; proactive maintenance; minimized unplanned downtime
Resource Allocation	Confined within departments; inefficient; uneven workloads	Dynamic, competence-based allocation across departments; improved productivity
Root Cause Identification	Limited support; depends on personal experience; issues may recur	Provides competence-based information; effective diagnostics; reduces problem recurrence
Data Integration	Fragmented data sources; underutilized information; hampers decision-making	Integrates data using semantic technologies; enhances accessibility; informs planning and decisions

5.1 Data Acquisition

In the first step, three major data sources are identified: MES, DSB, and HR databases. The MES system consists of production orders, machine sequences, and tact times. The available data go back seven years and contain approximately 330,000 maintenance reports. Fault documentation, conducted in a mix of German and English-German, provides machine down events with corresponding machine, EQUIPMENT that the down event is assign to, at which point of time, TIME_STAMP, the machine goes out of order, including a written description called GUI_BUCHUNGSTEXT and a categorization of the event DOWN_EVENT. After the assignment of a maintenance technician and the consequent maintenance actions, a description of the performed tasks is created, referred to as ACTION, which is uniquely defined by CREATED and the performing maintenance technician CONTACT_PERSON, as shown in Table 12.

This information is used to

- plan initial shifts,
- determine machine faults and its criticality with the help of APC data,
- classify the faults according to required competencies,
- determine the best-fit maintenance technician for the task,
- recommend appropriate tasks for the maintenance technician.

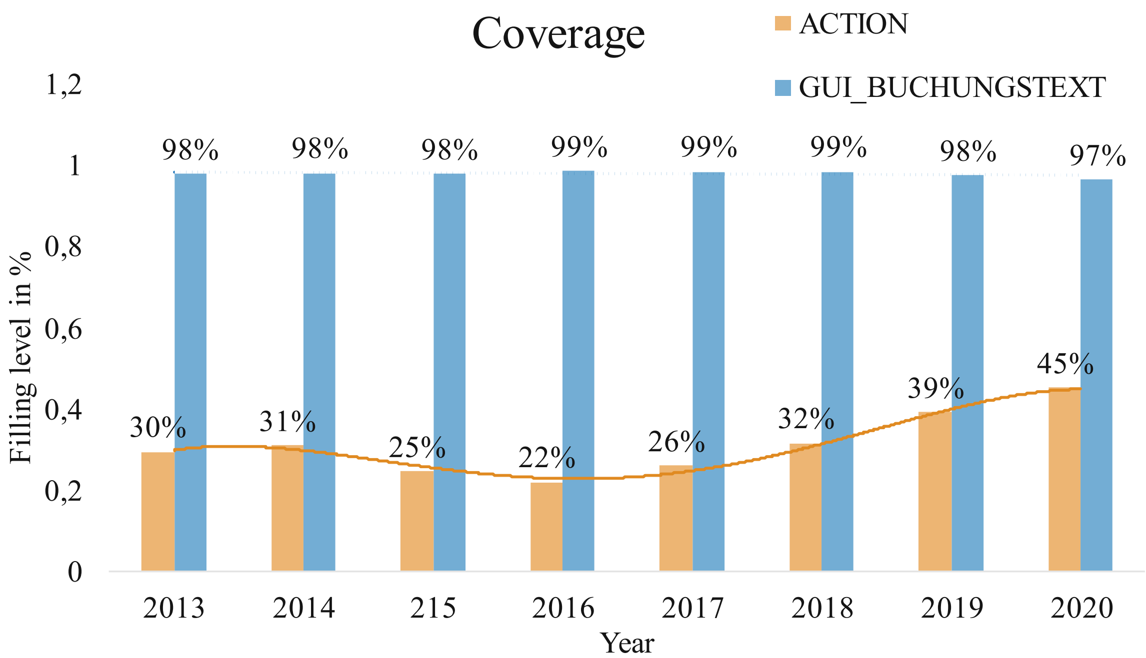
Furthermore, data sources from HR data in ERP systems include job and competence descriptions of maintenance staff.

Table 12: *Maintenance Logs Documenting Down Events and Actions Taken Actions for Specific Equipment*

EQUIPMENT	TIME_STAMP	GUI_BUCHUNGSTEXT	DOWN_EVENT	ACTION	CREATED	CONTACT_PERSON
Equipment_1	24.10.2018	Operator_1 temperature error	ST	TCU reset and filled -> Temperature back to normal Technician_1	24.10.2018 12:45	Technician_1
Equipment_2	24.10.2018	Operator_2 ETCH TIME Main ETCH	ST	MHz Generator Reset performed Technician_2	25.10.2018 13:24	Technician_2

A thorough data quality assessment was carried in accordance with Pipino et al. (2002) and Taleb et al. (2018), which revealed that the following quality aspects should be examined: (i) statistical analysis of the data, (ii) coverage of the data of the present process, (iii) completeness, (iv) redundancy, (v) freedom, (vi) interpretability and noise, and (vii) accuracy. In this context, (i) basic statistical parameters such as distribution, mean, and variance are collected; (ii) it is checked whether the data covered the maintenance process as documented in the company documents; (iii) the completeness of the data sources (i.e., empty cells) is verified; (iv) check for redundant entries were conducted, v) check for heterogeneity of entries are performed; (vi) check for interpretability and the number of unique words used in relation to all words are made; and (vii) checks for accuracy of entries are conducted (i.e., how well they describe the activity). Based on the coverage, as shown in Figure 54, it is clear that there is a corresponding ACTION for every GUI_BUCHUNGSTEXT in only around 27% of cases, suggesting that only a quarter of the data can be used for further analysis.

Figure 54: *Coverage of ACTION for GUI_BUCHUNGSTEXT*



5.2 Data Preprocessing

Following preprocessing, entities of unstructured text data sources are mapped to the I40CO, WOW, and FD concepts for the population of the KG.

Section 5.2 on provides a detailed account of the techniques used to prepare raw data for integration into ARCHIE. The section addresses the challenges of handling multi-structured data, including structured sensor outputs and unstructured textual inputs such as maintenance logs. Key processes discussed include data cleaning, normalization, and feature extraction to ensure data quality and consistency. Section 5.2.1 describes methods for handling time-series data, including outlier detection, smoothing, and standardization to ensure compatibility across datasets, whereas Section 5.2.2 focuses on TLP techniques to process maintenance logs, including tokenization, stemming, and vectorization for effective text analysis. These subsections collectively outline a systematic approach to preparing data, enabling the effective application of predictive and cognitive maintenance models.

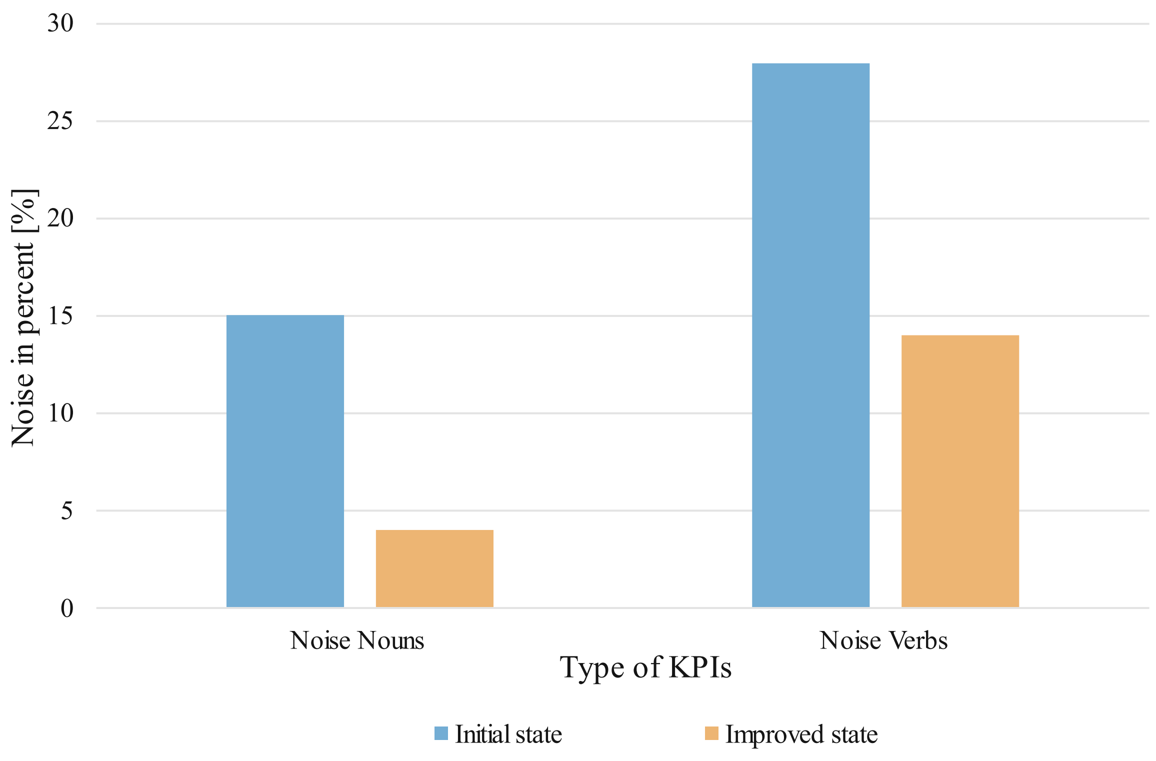
5.2.1 Preprocessing of Structured Data

In this step, the outliers downtime in the dataset—specifically, multi-day downtimes, were removed, . This is particularly important in this dataset, as some of the machines have served as spare parts stores for identical machines over long periods. Therefore, their respective downtime is not characterized by the complexity of the problem but by the use of the machine as a spare part storage. As for the APC data, as they already come with time-domain signal processing, a principal component analysis was performed to manage the multi-dimensional feature set (Biegel et al., 2022).

5.2.2 Preprocessing of Unstructured Data

Concerning unstructured, textual data, each GUI_BUCHUNGSTEXT and ACTION was first tokenized, and words were standardized. This primarily concerns abbreviations and common spelling mistakes. However, character strings (e.g., "->") were also either removed or replaced by corresponding words. These steps were taken to reduce the number of unique words in the dataset and allow for better predictions later. Preprocessing leads to a reduction of unique words, which, in turn, reduces the noise in the text, where the ratio of the total frequency of all identified words per unique word is called noise. With the introduction of large language models, preprocessing is still necessary, although abbreviations and character strings are less of a challenge. However, the black box character of the preprocessing pipeline becomes larger.

Figure 55: *Improvements Through the AI-Enhanced Methodology*



This results in a reduction of noise in nouns by 385% and in verbs by 203% in the use case, as shown in Figure 55. The strong reduction of nouns is particularly noteworthy here, as illustrated in Table 13. This can be attributed to the high number of abbreviations, bilingual terms, and individual spelling variants. In the next step, stop words were removed, and lemmatization was used. To create a processable feature representation, sentences were vectorized using TF-IDF.

Table 13: *Noise in Unstructured Data*

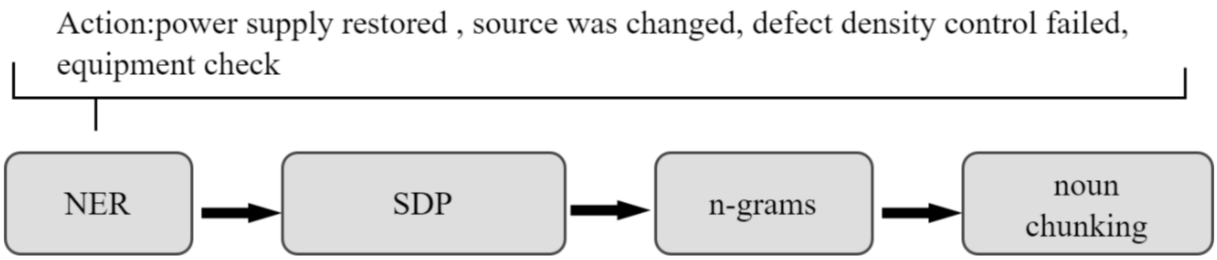
	Nouns		Verbs	
	Unique nominations	Frequency	Unique nominations	Frequency
Original	2,434	255,860	480	117,137
Domain Dictionary	200	92.022	126	102,970

5.3 Data Analysis

5.3.1 Text Mining

Once the dataset has undergone a comprehensive pre-processing and feature engineering phase, specific entities can be extracted from the text through the application of TLP techniques. As detailed in Section 4.3.1, the TMP is employed for this purpose, utilizing domain-specific methodologies tailored to the maintenance context. A noteworthy enhancement to this procedure is the creation of a domain-specific dictionary, referred to as the "ground truth." This dictionary improves the accuracy of named entity recognition by providing precise definitions and contextual understanding of domain-relevant terms, thereby ensuring a higher degree of accuracy in identifying key entities. The outcomes of the TMP facilitate the extraction of entities, which are subsequently incorporated into the KG. The KG offers a structured representation of the extracted knowledge, linking identified entities with their semantic relationships to enable comprehensive reasoning and analysis. If an entity is not identified by the named entity recognition component, a logistic regression model is employed as a supplementary classification mechanism. This model categorizes unidentified entities into predefined classes, including tasks, names, equipment, and other relevant categories, thus ensuring that no valuable information is overlooked.

Figure 56: *Applied Text Mining Pipeline, based on Kohl and Ansari (2023)*



A crucial output of this process is the extraction of *TASKS* from the *ACTION* field, which is defined as the fundamental units of a countermeasure (see Table 14). Tasks represent individual maintenance activities and are of great consequence for downstream applications, such as planning, task allocation, and root-cause identification. Similarly, this procedure is extended to include data from GUI_BUCHUNGSTEXTE, which enables the methodology to be applied across a range of textual sources within the industrial environment. This guarantees a unified methodology for entity extraction and *TASK* representation, thereby enhancing the robustness and applicability of the system in diverse maintenance scenarios.

By integrating both TMP and logistic regression-based classification, this approach not only enriches the KG but also lays the foundation for advanced analysis and decision-making in CAS. The structured representation of tasks and entities enables improved tracking, resource allocation, and competence-based task assignment, significantly contributing to the efficiency of maintenance operations (see Table 14).

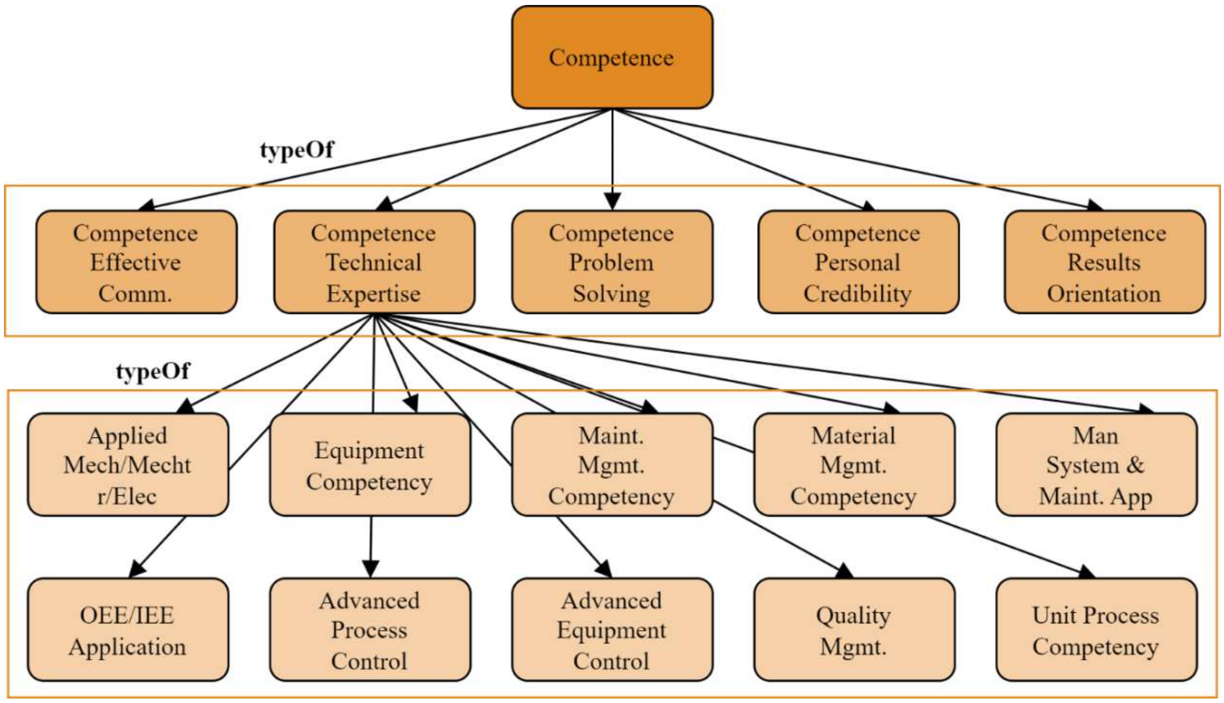
Table 14: *With TMP Transformed ACTION Column to TASK*

EQUIPMENT	ACTION	TASKS	CREATED	CONTACT_PERSON
Equipment_1	TCU reset and filled -> Temp back to normal Technician_1	[TCU reset, TCU befüllt]	24.10.2018 12:45	Technician_1
Equipment_2	Mhz Generator reset performed Technician_2	[Mhz Generator reset]	25.10.2018 13:24	Technician_2

5.3.2 Statistical Learning Algorithm

Based on the generic formulation of the KG and the underlying ontologies, particularly WOW, the KG was extended by company-specific competencies and their competence , see Figure 57. This enables a semantic mapping from *ACTION* to *TASK* to *Competence* and their distributions. For the STA, the ESCO entry 7421 (Mihaylov & Tijdens, 2019) was used as the starting value for θ_{T_i}, as well as the dispersion value e. Based on the *calculated competence in technical* expertise $P_{TE(U_i|T_i)}$, the *competence effective communication* and *problem-solving* were determined using Mikhaleov's distribution values.

Figure 57: *Expansion of Semantic Data Storage to Include Enterprise-Specific Sub-Competencies and Their Levels*



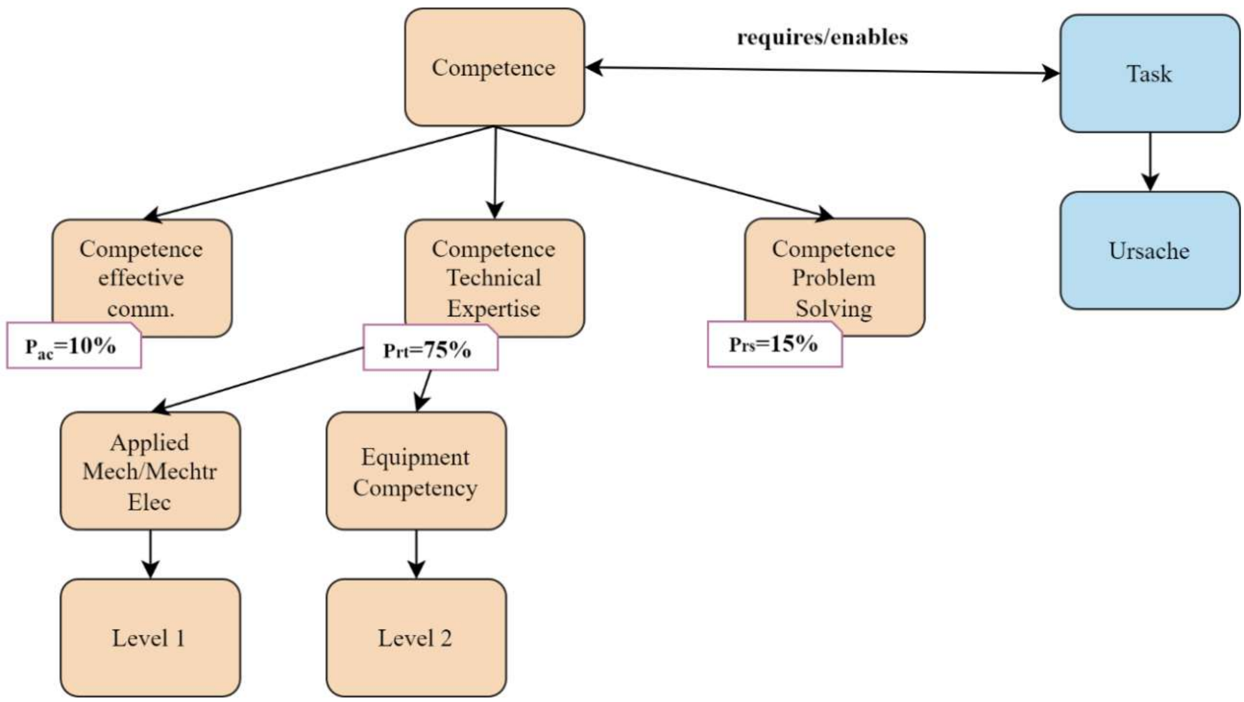
Based on these generic competencies, a weighting matrix $M_{TE}(U_i|C_i)$ was used to infer company-specific *Equipment Competence* and *Applied Mechanical Competence* using $L_{TE}(U_i|C_i)$ and its levels s. It should be noted that ω_{STE_i,U_i} and τ_{U_i} were determined in expert workshops and fine-tuned based on important TASKS. The result of the STLA can be seen in Table 15, where a competence distribution is provided for each activity and action. This distribution of competencies was carried out for all historical *ACTIONS* and *TASKS* and, thus, forms the basis for subsequent shift planning and task allocation.

Table 15: *Distribution of Competencies for ACTION and TASK*

EQUIPMENT	TASKS	ACTION Competence $[P_{EC}, P_{TC}, P_{PS}]$	TASKS $[P_{EC}, P_{TC}, P_{PS}]$	CREATED	CONTACT PERSON
Equipment_1	[TCU resetet, TCU filled]	[0.1, 0.75, 0.15]	[[0.12, 0.77, 0.11], [0.1, 0.8, 0.1]]	24.10.2018 12:45	Technician_1
Equipment_2	[Mhz Generator reset]	[0.12, 0.77, 0.11]	[0.12, 0.77, 0.11]	25.10.2018 13:24	Technician_2

The distribution of competencies can then be used to semantically infer from the task the original cause to competence and then the different sub-competencies. Then. in the use case, the technical competence $P_{TE(U_i|T_i)}$, for instance, is used to infer enterprise-specific sub-competencies and their levels, as shown in Figure 58.

Figure 58: *Representation of the Working of the SLA*



For anomaly detection, a DL-based anomaly detection system, based on Pang et al. (2022), was implemented. The data used stems from APC systems, loaded in a pandas DataFrame[6], to identify any irregularities in the process. The APC data included parameters such as ion beam current, dose, and energy.

[6] Pandas is a software for the Python suited for data manipulation and analysis.

Code 1: *Loading Data into Python, Scaling, and Performing a Train-Test Split*

```
import pandas as pd
from sklearn.preprocessing import MinMaxScaler

# Load data
apc_data = pd.read_csv('ion_implantation_data.csv')

# Normalize data
scaler = MinMaxScaler()
apc_data_scaled = scaler.fit_transform(apc_data)

# Split data
split_index = int(len(apc_data_scaled) * 0.8)
train_data = apc_data_scaled[:split_index]
test_data = apc_data_scaled[split_index:]
```

An autoencoder suited for high-dimensional time-series data was chosen based on the study by Pang et al. (2020). The model includes long short-term memory layers, considering the sequential nature of the data, using the RLU function with the Adam optimizer and MSE for loss see Code 2.

Code 2: *Design of the Autoencoder Model*

```
from keras.models import Sequential
from keras.layers import LSTM, Dense, RepeatVector, TimeDistributed

# Autoencoder model
model = Sequential([
    LSTM(128, activation='relu', input_shape=(train_data.shape[1],
train_data.shape[2]), return_sequences=True),
    LSTM(64, activation='relu', return_sequences=False),
    RepeatVector(train_data.shape[1]),
    LSTM(64, activation='relu', return_sequences=True),
    LSTM(128, activation='relu', return_sequences=True),
    TimeDistributed(Dense(train_data.shape[2]))
])
model.compile(optimizer='adam', loss='mse')
```

The autoencoder was trained on the normalized training data. The reconstruction error on the test set was established and used as a threshold. The exemplary used threshold must be coordinated with domain experts and production quality requirements see Code 3.

Code 3: *Setting of Threshold Using MSE*

```
# Predict on test data
predicted = model.predict(test_data)
mse = np.mean(np.power(test_data - predicted, 2), axis=1)

# Define a threshold
threshold = np.quantile(mse, 0.99)

# Detect anomalies
anomalies = mse > threshold
```

The implemented DL model for anomaly detection, based on APC data, proves highly effective, with an F1 score of 77.5%. The autoencoder model, especially with its long short-term memory layers, was adept at identifying irregular patterns in the time-series data, ensuring the quality and reliability of the semiconductor manufacturing process.

5.4 Semantic Data Storage

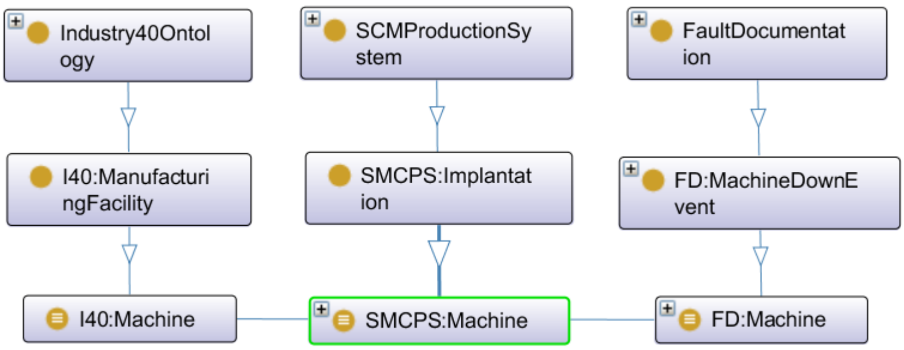
The ARCHIE KG builds upon well-established ontological frameworks, including the generic WOW and I4.0 ontologies, as well as the FD. These foundational ontologies provide a robust starting point for structuring the semantic relationships necessary for cognitive maintenance systems. However, for the specific application in semiconductor manufacturing, this framework requires enrichment with the Semiconductor Production System Ontology (SMCPS), as illustrated in Figure 4. Both the SMCPS and FD ontologies were developed in collaboration with domain experts, ensuring they accurately reflect the specialized requirements and nuances of the semiconductor production context.

To integrate these ontologies into a unified framework, a full merge is conducted. This results in a single, cohesive ontology that consolidates concepts, entities, and relationships across domains. The merging process involves aligning and harmonizing overlapping or related concepts, ensuring semantic consistency and eliminating redundancies. The most significant merges are elaborated below to illustrate the methodology and its implications.

A critical point of integration is the machine entity, which serves as the connecting class across the FD, SMPC, and I4.0 ontologies. The mapping ensures that concepts such as SMPC:machine, FD:machine, and I4.0:machine are semantically equivalent ($SMPC\!:machine \equiv FD\!:machine \equiv I4.0\!:machine$), as depicted in Figure 59. This alignment enables seamless interoperability between fault documentation, production processes, and broader Industry 4.0 concepts. By linking the machine entities, the KG can reason about equipment functionality, faults, and associated maintenance activities in a unified manner. Further enhancements include the incorporation of process-specific attributes from the SMCPS ontology, such as wafer handling, lithography, and etching stages, into the existing structure. These attributes are mapped to the corresponding entities in the I4.0 and FD ontologies to ensure consistency. For example, $SMPC\colon WaferHandlng$ is mapped to a sub-class of I4.0:Process, while their associated faults are integrated into $FD\colon Failure$. The unified ontology not only enriches the semantic capabilities of the ARCHIE KG but also provides a domain-specific layer that addresses the unique requirements of semiconductor manufacturing. This integration facilitates advanced reasoning, enabling tasks such as fault diagnosis, root cause analysis, and PdM to be performed with greater accuracy and contextual awareness.

The process of merging these ontologies highlights the importance of domain collaboration, as the nuanced understanding of domain experts ensures that the ontology structure remains both comprehensive and practical. The resulting unified ontology forms the backbone of the ARCHIE KG, allowing it to function as a powerful tool for knowledge representation, reasoning, and decision support in complex industrial environments.

Figure 59: *Merge of I40CO, FD, and WOW*



5.5 AI-Enhanced Decision Making

Section 5.5 explores the integration of IAI-based decision-making processes within the ARCHIE framework, highlighting how AI-driven models leverage structured and unstructured data to augment maintenance planners and technicians, enabling proactive and informed decisions. Key aspects include the implementation of DL, optimization algorithms, and probabilistic reasoning to enhance decision-making, reduce downtime, and improve asset availability. Thus, this section underscores the transformative potential of IAI in maintenance.

5.5.1 Shift Planning

In the first step, the LPM is defined by the objective function. The objective function aims to optimize the distribution of competencies over all shifts based on the historical need for competencies, considering factors such as competence level matching and workload balance, as seen in Equations 13–15. Efficiency is gauged through a combination of factors, including the alignment of technician competence levels with the specific requirements of each shift and the equitable distribution of workload. The objective function, thus, encapsulates a multi-faceted approach, aiming to optimize both operational efficacy and the well-being of the workforce.

A set of constraints is integral to the model, ensuring that each shift is staffed with an adequate number of maintenance technicians possessing the requisite skills. These constraints are derived from the detailed requirements of the use case process, which may include technical competence, certain equipment-specific certifications, and safety certifications. The generic constraints of the LPM are then extended by various use case-specific constraints, e.g., minimum and maximum number of maintenance technicians required for each shift see Code 4. The linear programming solver, as detailed in Code 4, is employed to navigate through the solution space defined by the objective function and constraints. This solver iteratively evaluates potential shift allocations, seeking the configuration that minimizes the objective function while adhering to all constraints. The optimization process involves a linear model capable of handling the complexity and scale of the problem, leading to a mathematically optimal solution. Although this results in higher runtime, the shift schedule is not a critical time process. Upon identifying the optimal shift allocation, the solution is then translated into a practical shift schedule. This schedule details the specific assignments of technicians to various shifts, ensuring efficiently staffed operations. Here, the average competence requirements from the previous year, within a floating time window, are used. The translation of the optimized solution into a real-world schedule involves considerations such as shift timings, technician availability, and contingency plans for unforeseen absences or equipment downtime. The presented shift planning algorithm was discussed in (Ansari et al., 2023; Kohl & Ansari, 2023). The planning algorithm has since been adapted and implemented in ARCHIE based on input from domain experts and the requirements of the use case.

Code 4: *Implementation of the Linear Programming Algorithm, Including Constraints*

```
from scipy.optimize import linprog

# Example data (to be replaced with real data)
num_technicians = 10
num_shifts = 5
availability = [[1 if technician is available else 0 for shift in
range(num_shifts)] for technician in range(num_technicians)]
competence_levels = [competence_level_of_technician for technician in
range(num_technicians)]
shift_requirements = [required_competence_level_for_shift for shift in
range(num_shifts)]

# Objective function: maximize competence levels while meeting shift
requirements
c = [-competence for competence in competence_levels for _ in
range(num_shifts)]

# Constraints
A = []
b = []

# Each shift must be covered
for shift in range(num_shifts):
    constraint = [1 if i % num_shifts == shift else 0 for i in
range(num_technicians * num_shifts)]
    A.append(constraint)
    b.append(shift_requirements[shift])

# Each technician works at most one shift
for technician in range(num_technicians):
    constraint = [1 if i // num_shifts == technician else 0 for i in
range(num_technicians * num_shifts)]
    A.append(constraint)
    b.append(1)

# Solve the linear programming problem
res = linprog(c, A_eq=A, b_eq=b, bounds=(0, 1))

# Translate the solution into a shift schedule
schedule = [None if x < 0.5 else i // num_shifts for i, x in
enumerate(res.x)]
print("Shift Schedule:", schedule)
```

5.5.2 Downtime Forecast

As proposed in Section 4.2.2, ML models are employed for downtime forecasting using data and maintenance reports. These models were meticulously trained on an existing dataset, with the creation of individual parameter grids tailored to each model's unique characteristics. This approach allowed for a more nuanced and effective optimization of each model.

The optimization process for each ML model involved the utilization of both BOW and TF-IDF vectorization techniques. This was coupled with rigorous 10-fold cross-validation on the training dataset, which constituted 75% of the total dataset. Such a methodological approach ensures that the models were not only trained on a substantial portion of the data but also validated across multiple subsets, enhancing their generalizability and robustness.

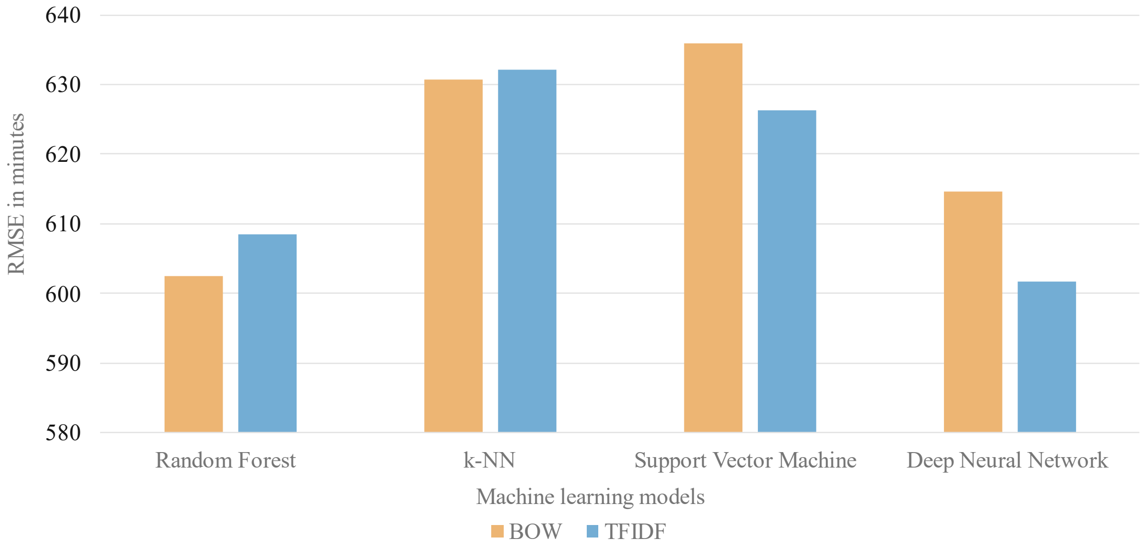
Subsequent to the training and validation phases, the optimal parameter combinations for each model were systematically evaluated and ranked against one another, as illustrated in Figure 60.This comparative analysis was pivotal in determining the most effective model for downtime prediction. The DNN model, when integrated with TFIDF vectorization, emerged as the superior model, exhibiting a mean MAE of 442.5 minutes across the 10-fold cross-validation, indicating its high precision in predicting downtime durations.

Furthermore, the mean RMSE of the DNN model has been calculated to be 605.7 minutes, with a notable standard deviation of 2,251.8 minutes within the use-case dataset. This RMSE value, while higher than the MAE, provides an additional layer of insight into the model's predictive capabilities, particularly regarding its sensitivity to larger errors.

Given its exemplary performance on the test set, the DNN model was subsequently integrated into the AI-enhanced methodology for the specific use case., signifying an improvement in the application of AI and ML techniques in industrial settings, particularly in the realm of PM. The ability of the DNN model to accurately predict machine downtime not only enhances operational efficiency but also contributes to opportunistic maintenance planning, ultimately leading to reduced machine downtime and associated costs.

In summary, the application of sophisticated ML models, especially the DNN combined with TF-IDF vectorization, represents a substantial leap forward in the field of PM. The rigorous training, validation, and comparative analysis of these models underscore the potential of AI-enhanced methodologies in transforming industrial operations. The presented downtime forecast algorithm was discussed in (Ansari et al., 2021). The downtime forecast algorithm has since been adapted and implemented in ARCHIE based on input from domain experts and the requirements of the use case.

Figure 60: *Comparison of the RMSE Value of the Best ML Models*



5.5.3 Task Allocation

The current shift plan, combined with the machine operator's fault report and the shift book-entry metadata, allows for selecting the selection of the most suitable maintenance technician for the problem. The first step is to filter the dataset by machine and type of fault. The filtered dataset is then analyzed using the words employed in the problem description. The similarity between the words used in the machine fault report and the number of similar words is calculated. Consequently, the technician with the highest similarity value for the given problem is selected, allowing for the assignment of the most suitable maintenance technician.

The application of GAs offers a methodical approach to optimize task allocation. This process involves several key steps. The first phase in the GA process entails generating an initial population of potential task allocation solutions. Each individual in this population represents a unique allocation of tasks to the available pool of maintenance technicians and equipment, considering the constraints and requirements of the maintenance process, such as necessary competencies and safety regulations, including a one-minute reaction time for the GA. Notably, the one-minute reaction time for the GA was an important requirement from domain experts, as slower reaction times for ad-hoc task allocation would render ARCHIE unsuitable for their daily operations.

A crucial component of the GA is the fitness function. This function evaluates each potential solution in the population based on several criteria, such as total process time, efficiency of resource utilization, and adherence to predefined process constraints. The fitness function thus quantifies the effectiveness of each task allocation strategy, as implemented in Code 5.

Code 5: *Defining the Genetic Algorithm*

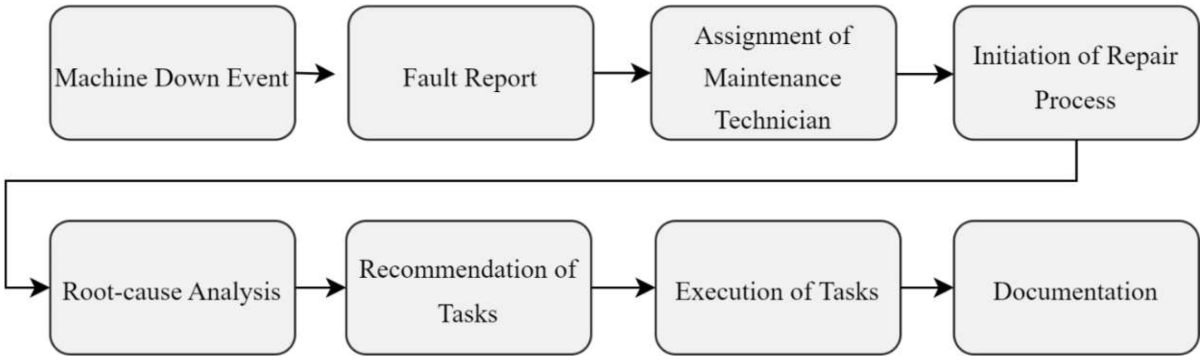
```python
import pygad
import numpy as np

def fitness_func(solution, solution_idx):
    # Fitness function to evaluate each solution
    # Example: Minimize process time and maximize resource efficiency
    process_time, resource_efficiency = calculate_metrics(solution)
    fitness = -process_time + resource_efficiency
    return fitness

ga_instance = pygad.GA(num_generations=50,
                       num_parents_mating=10,
                       fitness_func=fitness_func,
                       sol_per_pop=100,
                       num_genes=len(tasks),
                       gene_space=[list(range(len(technicians)))])
```

In the GA operations step, a selection method, such as tournament selection, is implemented to identify the most promising solutions from the population (Z. Zhou et al., 2020). This method ensures that solutions with higher fitness scores have a greater probability of being selected for the next generation, thereby guiding the GA toward more optimal solutions. To generate new solutions and maintain genetic diversity within the population, crossover and mutation operations are applied. Crossover combines features from two parent solutions to create offspring, while mutation introduces random changes to some solutions, helping to explore new areas of the solution space. The GA is iteratively run until a predefined convergence criterion is met. This criterion could be a maximum number of generations or a target fitness level, ensuring that the algorithm terminates upon achieving satisfactory optimization. The task allocation is implemented as detailed in Section 4.5.3. Within the training iterations of the model, the approach could satisfy appropriate task allocation in 81% of the cases as compared to traditional methods.

5.5.4 Task Recommendation with Root Cause Identification

Task recommendations for documentation are generated as described in the methodology, see Figure 61. A typical maintenance process starts with the machine fault report issued by the machine operator (see Figure 61). The operator can enter additional metadata in addition to the data automatically taken from the DSB. In the following step, the machine operator is presented with the five most frequent fault descriptions, which drastically reduces the time needed for documentation while simultaneously increasing the quality of the short texts written. Based on the operator's input, the next step involves the maintenance technician, referred to as the worker in the code, starting to repair the fault. For this purpose, further metadata is filled in first, and the maintenance technician is provided with aligned word recommendations for the countermeasure, which also enable fast fault detection. Additionally, the technician can enrich his/her report with images. For more complex problems where the word recommendations are insufficient, the search function can be used to access similar error messages with corresponding documented solutions, including relevant pictures, immediately. This is achieved by implementing the task recommendation algorithm.

Figure 61: *Task Recommendation and Root-Cause Identification Pipeline*



First, a DFS algorithm is used to thoroughly explore the KG. The algorithm starts at a selected node, the `down_event` and explores as far as possible along each branch before backtracking. This method is particularly effective for discovering all possible paths and relationships for potential task-worker pairings and dependencies, see Code 6.

Code 6: *Implement DFS for Graph Exploration*

```
def dfs(graph, start, visited=None):
    if visited is None:
        visited = set()
    visited.add(start)
    for (start, end) in graph.out_edges(start):
        if end not in visited:
            dfs(graph, end, visited)
    return visited
```

In the following step, content-based filtering is employed to match `TASK` with `WORKER` based on their `COMPETENCE`. This approach involves creating a profile for each maintenance technician based on their competencies, extracted from the KG, and then matching these profiles with the competencies connected to different tasks. The algorithm evaluates the suitability of tasks for a maintenance technician by comparing the competencies required for each task with those possessed by the technician. This technique ensures that tasks are recommended to maintenance technicians who have the requisite competence, thereby reducing MTTR for task execution (Ansari et al., 2021), see Code 7.

Code 7: *Implement Content-Based Filtering for Task Matching*

```python
def content_based_filtering(graph, worker):
    worker_competences = [edge[1] for edge in graph.edges(worker)]
    suitable_tasks = []
    for task in tasks:
        task_required_competence = [edge[1] for edge in graph.edges(task)]
        if any(competence in worker_competences for competence in
task_required_competence):
            suitable_tasks.append(task)
    return suitable_tasks
```

In the concluding step, PageRank (Thalhammer & Rettinger, 2016), is used to rank tasks based on their importance and relevance within the KG, based on a given `down_event`. The algorithm assigns a numerical weighting to each task, with higher weights indicating greater importance or relevance. This weighting is determined by the number and quality of edges (relationships) connected to the task node in the KG. By employing PageRank, the task allocation can prioritize tasks, ensuring that more promising tasks or those with higher correlation between required competencies by the `task` and available `competencies` by the `worker` are recommended first, see Code 8.

Code 8: *Applying PageRank for Task Ranking*

```python
def rank_tasks(graph, tasks):
    pr = nx.pagerank(graph)
    task_ranks = {task: pr[task] for task in tasks if task in pr}
    ranked_tasks = sorted(task_ranks, key=task_ranks.get, reverse=True)
    return ranked_tasks
```

The implemented algorithms are then combined into an integrated approach ensures that the task recommendation system is not only accurate in matching tasks with workers but also effective in prioritizing tasks based on their importance within the maintenance workflow.

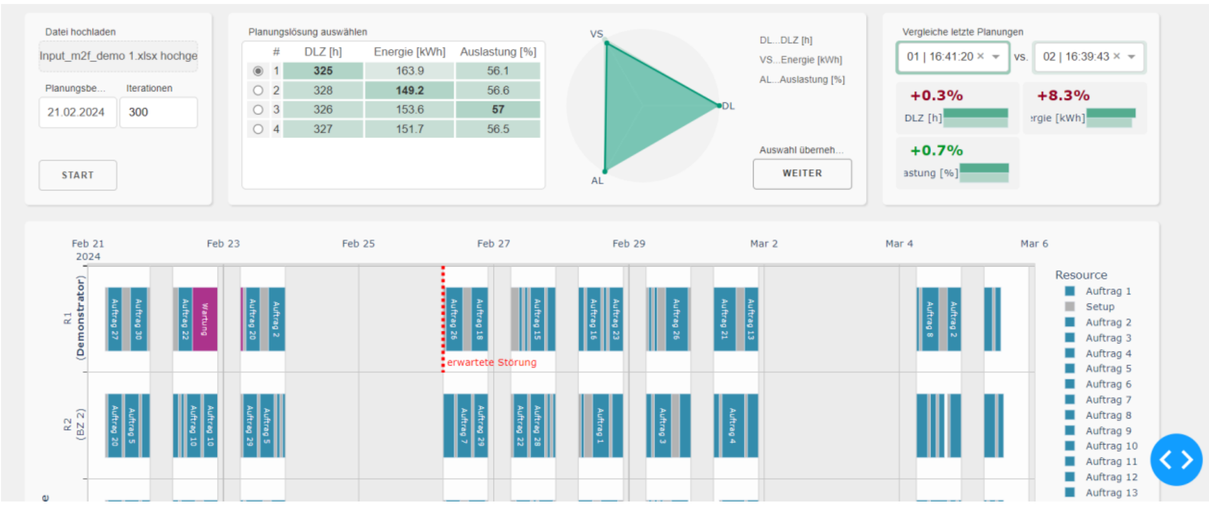
5.6 Human-Interface

In accordance with the requirements established in a workshop, a prototype user interface has been meticulously created, as shown in Figure 45.The development of this mock-up and the resulting user interface strictly followed the guidelines outlined in Section 4.6. The development was an iterative process, involving feedback from domain experts, see Figure 62. Once developed, the interface underwent a comprehensive evaluation during an on-site workshop.

Figure 62: *Development Process of ARCHIE*

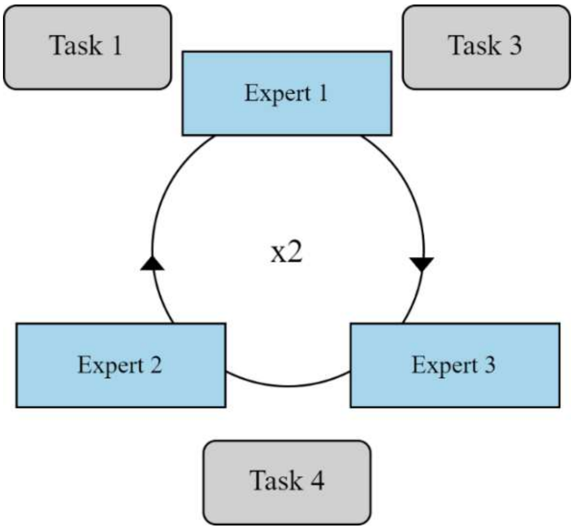
This evaluation involved a cohort of experts divided into three distinct teams consisting of business professionals, managers, and IT personnel. The participants were tasked with performing three specific activities: documenting maintenance tasks, searching for analogous errors, and evaluating task assignments. These activities were performed in two different ways, as shown in Figure 63 and 64.

Figure 63: *Planning Space*



In the latter part of the workshop, the participants engaged in in-depth discussions, referred to as “deep dives,” which provided valuable insights into their user experience. Additionally, the participants were presented with a detailed technical explanation of the algorithm’s functionality. Noteworthy here is the planning logic, which was developed in addition to the functionalities shown in Section 4.6 and allows planning to be based on the target variables of energy, throughput time, and capacity utilization.

Figure 64: *Use Case Evaluation Process*



The evaluation process culminated with the administration of a PSSUQ (Jordan et al., 1996), designed to measure user satisfaction with the system's usability, using the provided Graphical User Interface.

This questionnaire consisted of19 questions covering various aspects such as the user's overall satisfaction with the system (OVERALL), the usefulness of the system (SYSUSE), the quality of the information provided (INFOQUAL), and the quality of the user interface (INTERQUAL). Responses were recorded on a seven-point Likert scale ranging from 1 ("I strongly agree") to 7 ("I strongly disagree"), following Lewis (1995).

Table 16: *PSSUQ Evaluation Round 1*

Nr.	Question	Group 1	Group 2	Group 3	Total
1	Overall, I am satisfied with the ease of use of this system.	2	2	3	2.3
2	The system is easy to understand.	2	2	3	2,3
3	I was able to successfully complete the tasks and scenarios with this system.	2	2	2	2,0
4	I was able to complete the tasks and scenarios quickly with this system.	2	7	2	3.7
5	I was able to complete the tasks and scenarios efficiently with this system.	2	7	5	4.7

Nr.	Question	Group 1	Group 2	Group 3	Total
6	I felt comfortable using the system.	1	2	2	1.7
7	It was easy to learn/understand how to use the system.	1	1	3	1,7
8	I believe that I can quickly become productive with this system.	1	2	4	2,3
9	The system provided error messages that clearly indicated how to fix problems.	7	7	7	7.0
10	If I made a mistake while using the system, I was able to fix it easily and quickly.	7	7	7	7,0
11	The information provided by the system was clear.	7	7	7	7,0
12	It was easy to find the information I needed.	7	2	7	5.3
13	The information provided by the system was easy to understand.	2	2	7	3,7
14	The information helped me a lot in completing the tasks and scenarios.	7	2	7	5,3
15	The layout of the information on the system's screens was clear.	2	3	5	3,3
16	The user interface of this system was appealing.	2	2	3	2,3
17	I liked the operation of the user interface.	2	3	3	2.72
17	I liked the operation of the user interface.	2	3	3	2,7
18	The system has all the functions and features I expect.	3	3	3	3.0
19	Overall, I am satisfied with this system.	3	2	3	2,7

The initial results, shown in Table 16, indicated general satisfaction among users but also highlighted areas of concern, particularly in questions 9–12. After iterative improvements to the system and an expansion of the user group to nine groups, the survey was conducted again. The subsequent results, presented in Table 16, revealed significant improvements in responses to questions 9–12. In particular, while there was a slight decrease in the INTERQUAL aspect, there were significant improvements in the OVERALL, SYSUSE, and INFOQUAL categories.

Table 17: *Comparison of Evaluation Rounds*

Key Figure	Total Round 1	Total Round 2	Difference
OVERALL – Q1-19	3,7	3.33.3	-0.4
SYSUSE – Q1-8	2.6	2,3	-0.3
INFOQUAL- Q9-15	5,5	3.9	-1.6
INTERQUAL - Q16-18	2,7	3,0	0.3

The data presented in Figure 55 demonstrates a consistent reduction in MTTR across all months in *t2* compared to *t1*. The average reduction in MTTR across the seven months is around 20%. The observed reduction in MTTR is consistent with the expected benefits of a CAS, which underlines the qualitative results of the PSSUQ study. Nevertheless, based on the PSSUQ results, the interface and usability show areas for improvement. Accordingly, a maintenance chatbot with a redesigned interface was implemented, enabling communication with maintenance planners and operators in natural language (Kohl, Eschenbacher, et al., 2024).

In this context, the CAS acts as a self-service tool that enables staff to resolve problems independently or with more specific guidance from the system. This self-service approach streamlines the maintenance process and leads to less machine downtime, thereby increasing availability and OEE. While the data suggests a positive correlation between the CAS and reduced MTTR, it is crucial to acknowledge limitations. It can be seen that the correct distribution of skills among maintenance staff, including targeted support for them in the solution process, leads to a reduction in MTTR; however, other influencing factors cannot be completely ruled out. Learning curves of maintenance staff and continuous process improvements are not taken into account in the current study.

5.7 Interim Conclusion: Application in the Semiconductor Industry

Applying the cognitive maintenance framework in the semiconductor industry illustrates its significant practical relevance and potential impact. The case studies and examples demonstrate how AI-driven maintenance can effectively reduce downtime, improve equipment reliability, and support maintenance technicians with real-time data and intelligent recommendations.

This application addresses the research gap highlighted in Section 4.3 by emphasizing the need for industry-specific adaptations of the CAS. While the framework provides a generalizable approach, the unique challenges and stringent requirements of the semiconductor industry necessitate tailored solutions. The application also addresses the research questions by demonstrating how cognitive systems can be designed and implemented in a real-world context. For example, the framework's ability to provide background information and recommended solutions to existing machine faults illustrates the practical utility of AI-enhanced maintenance support systems.

The application underscores the importance of designing cognitive systems that are not only technically robust but also adaptable to the specific operational contexts of different industries. The example of the semiconductor industry illustrates how the integration of such systems can lead to significant improvements in maintenance efficiency and effectiveness, demonstrating the practical implications of the theoretical constructs outlined in Chapter 5. This detailed exploration of the application in the semiconductor industry confirms the potential of CAS to transform industrial maintenance practices while also pointing to the need for ongoing research to address specific industry challenges and optimize the implementation of these advanced technologies.

6 Conclusion and Outlook

This dissertation presents a data-driven framework for competence-based maintenance planning, illustrated through the example of the semiconductor industry but not limited to this context. Based on a comprehensive analysis of the current state of research and existing technologies, an AI-driven maintenance philosophy has been developed, including a gray-box AI approach, which was used for integrating cognitive systems and AI into maintenance processes. The derived goals formed the basis for the development of the ARCHIE framework, which aims to optimize maintenance activities through the use of physical and virtual sensor data. The developed framework was implemented, evaluated, and discussed in a simulation-based approach in real use cases in semiconductor production. Both the technical possibilities and the organizational framework conditions for the effective implementation of CAS were examined. It was found that the ARCHIE model enables a reduction of MTTR by up to 20%, increasing asset availability and thus the productivity and OEE of production systems.

This work makes an important contribution to coping with the high flexibility and reaction requirements placed on maintenance planning and operations today while simultaneously contributing to increasing the technical availability of systems. This supports objective 1 by identifying the components of a human-centered CAS that enables adaptive and competence-based maintenance planning and execution. Moreover, it contributes to solving problem statement 1 by addressing inefficiencies in task allocation and shift planning through competence-based planning. The sub-research question 1 is addressed in particular by integrating flexible shift planning and task assignment capabilities into the system, allowing for dynamic responses to operational changes.

By integrating unstructured data and cognitive learning processes into maintenance management, a significant step has been taken toward the implementation maturity of modern maintenance strategies. This aligns with objective 2 by enhancing the AI-driven architecture to leverage heterogeneous data, including unstructured inputs such as textual maintenance logs. It directly addresses the research problem by improving shift planning, fault identification, root-cause identification, and documentation through cognitive systems.

Furthermore, it ties to the thesis' research question by utilizing diverse data sources to provide maintenance planners and technicians with actionable, skills-based insights, thereby improving plant availability and reducing mean time to repair. Additionally, this aligns with objective 3 by providing real-time, detailed fault information and actionable recommendations for technicians and planners, thus addressing sub-research question 2. Finally, by reducing downtime and optimizing resource utilization, the system contributes directly to enhancing technical availability, supporting sub-research question 3 by improving documentation processes, reducing errors, and saving time through AI-enhanced decision-making tools.

In the following summary, the added value of the developed model is comprehensively presented, the research questions are answered, and an outlook on future research needs is given.

6.1 Application

Focusing more deeply on the application of the ARCHIE framework within the semiconductor industry provides an opportunity to highlight the specific benefits, challenges, and impacts associated with its deployment in a high-tech manufacturing environment. In summary, the benefits of ARCHIE for maintenance execution and planning can be described as follows, based on four main points.

- **Integration of Data Sources:** The semiconductor industry is characterized by complex manufacturing processes involving expensive and sensitive equipment. The ARCHIE framework is designed to integrate both structured and unstructured data from various sources, including machine sensors, maintenance logs, and operator inputs into a unifying KG. This integration allows for a comprehensive view of the manufacturing process, enabling more accurate fault diagnosis and maintenance forecasting.
- **Multi-Modal Data Analytics:** One of the primary advantages of implementing ARCHIE in semiconductor manufacturing is its capability for data analysis. It enables anomaly detection on sensor data as well as task extraction on DSB data, allowing for early fault identification, classification, and reliable MTTR estimation.
- **Competence Extraction:** The distinctive feature of ARCHIE is its utilization of SLA to infer competencies utilized from extracted tasks based on the ESCO framework. This approach provides insights into the maintenance tasks performed on the shop floor and the skills required to perform them. The distinctive feature of the SLA is its transferability to other application areas, as a scalable basic architecture based on the ESCO is employed. Furthermore, the underlying KG enables further analysis of the tasks, skills, and employee structure.
- **Competence-Based Task Allocation:** Another significant aspect of the ARCHIE framework is its ability to enhance maintenance shift planning and task allocation based on the competencies of available technicians. By analyzing past performance data and individual competence profiles, the system can assign the best-fitting personnel to

specific maintenance tasks. This optimizes the effectiveness of maintenance interventions, ensuring that the most qualified technicians handle complex issues.

- **Impact on Production Efficiency and Equipment Lifespan:** The holistic use of ARCHIE in the production environment enables a significant improvement in MTTR of around 20%. This quantitative study is supported by a very positive evaluation by experts in a PSSUQ.

6.2 Discussion of Research Questions

As seen in Chapter 3, no existing approaches were identified that deal with the semantic linking of maintenance data with the aim of supporting maintenance execution and planning as a cognitive assistance system. Although a large number of research papers deal with specific sub-topics of CAS in maintenance, most current approaches are based on idealized assumptions or do not comprehensively consider many factors in maintenance. However, a holistic view is essential, as only consistent data processing will meet with acceptance among operational and planning maintenance personnel. In addition, a robust, generalizable, and scalable approach is required, which can effectively deal with the heterogeneous data landscape and the ever-changing requirements of modern maintenance.

In this context, the overarching question was "What information can be extracted from textual data to improve the efficiency and effectiveness of maintenance planning?" The TMP presented in Section 4.3.1 demonstrates that individual, inseparable activities can be extracted from maintenance activities, which can be used for skills extraction. This competence information can subsequently be utilized for a rethought through (i) shift planning, (ii) downtime forecasting, (iii) task assignment, and (iv) task recommendation. In particular, the sub-questions set out in Section 1.2 can be answered as follows:

- **Sub-Research Question 1:** Competence-based shift planning and task assignment can be effectively integrated into maintenance planning by utilizing a systematic approach that aligns the specific competencies of maintenance personnel with the demands of individual tasks. According to the principles outlined in Chapter 4, this can be achieved through the use of TMP (FU3) and SLA (FU3) by extracting tasks from DSB entries and mapping those to concepts of machine down events, action descriptions, as well as worker profiles to capture the steps and competencies needed. Furthermore, the categorization of maintenance tasks based on their complexity and required competence is at the heart of SLA. This detailed information can then be used for shift planning and task assignment (FU5). The implementation of advanced scheduling algorithms to match tasks with the most suitable technicians based on their competence profiles is also crucial.

- **Sub-Research Question 2:** Maintenance technicians and planners can be supported by integrating semantically interlinked data, which can be reasoned over and used in further analysis for recommendations. This is achieved by establishing a KG as a centralized database that collects and stores real-time and historical data on machine performance, maintenance logs, and technician reports. The KG enables fast scalable reasoning and recommendations over relevant information needed for diagnosing and solving faults. Furthermore, the FU 3 facilitates data analytics, including the forecasting of downtime and MTTR. These tools enable the identification of patterns and the prediction of potential faults, thereby providing planners and technicians with proactive recommendations and KPIs for an optimized maintenance workflow.
- **Sub-Research Question 3:** The integration of automated data capture and processing technologies, combined with human-centered system design (FU6), can facilitate error-free and time-saving documentation of maintenance reports in industrial practice. Automated data collection is implemented through the use of physical and virtual sensors (FU1) to collect data from machinery and equipment during operation and maintenance data from DSB. Furthermore, digital reporting is enabled by task recommendation (FU5). This utilizes reasoning and ML to streamline the creation and submission of maintenance reports. These tools can pre-populate forms, with data directly from the automated collection systems, reducing the time required for technicians to fill out reports and ensuring that the information is accurate.

6.3 Limitations and Future Research

Despite its successes, the framework encounters several limitations. The integration of unstructured data, while innovative, also presents challenges in data quality and consistency. The variability in how data is recorded and the often incomplete logs can affect the performance of the AI models used in ARCHIE. Furthermore, the implementation complexity and the need for substantial initial training of the system to achieve optimal functionality are significant hurdles. These factors can lead to resistance from human operators, who may be skeptical of replacing traditional methods with a system perceived as less transparent or reliable. This is addressed in research problem 2 and its supporting sub-research question 2.

Therefore, future research on the framework should focus on enhancing the robustness and user-friendliness of the ARCHIE framework. Improving the AI's ability to manage diverse and imperfect data inputs will increase the system more reliable and applicability across various industrial settings beyond the semiconductor industry. Additionally, further studies could investigate the integration of more advanced predictive analytics tools to enhance the system's predictive capabilities. There is significant potential in the adaptability of current and emerging foundation models, particularly production foundation models, which provide a generalizable solution to the challenges of adopting predictive solutions for individual use cases, especially for structured data. Increasing the transparency of the AI decision-making process and offering more intuitive user interfaces could foster greater user acceptance and facilitate the transition from traditional practices.

From a broader perspective, the current scientific landscape requires an intensive discussion on how AI approaches can enhance the driving forces of our time—productivity and sustainability. In addition to integrated maintenance planning and data-driven productivity improvements, the research field should expand its focus to include self-healing machines and the cognitive and physical aspects of engineering and workplace design necessary for this, supporting research problem 2 and its guiding sub-research question 2. Although public discourse is strongly moving in the direction of full automation through AI, the near and medium-term future of maintenance planners and technicians will continue to rely heavily on HR and their roles in fostering a sustainable working environment. Therefore, a strong research emphasis should be placed on augmenting human competencies, particularly with the rise of generative AI, specifically large language model. This research focus aligns with the main research question of this thesis.

In conclusion, this PhD thesis discussion could encapsulate the effectiveness and innovative aspects of the ARCHIE framework while critically addressing the challenges and opportunities for enhancement. This balanced discussion not only highlights the contributions of the research but also establishes a clear path for future work in the field.

7 Appendix I - Detailed Insights to the Systematic Literature Review

Table 18: *Result of the Literature Review*

Article	Type	Periodical
Bagheri et al. (2015)	Conference Paper	Advances in Service-Oriented and Cloud Computing
Kuo et al. (2017)	Journal Article	Journal of Systems Architecture
Li and Wang (2017)	Journal Article	Advances in Manufacturing
Apiletti et al. (2018)	Conference Paper	IEEE SPA/IUCC/BDCloud/SocialCom/ SustainCom
Butte et al. (2018)	Journal Article	Computers & Industrial Engineering
Chacada et al. (2018)	Conference Paper	IEEE 23rd International Conference on Emerging Technologies and Factory Automation
Candanedo et al. (2018)	Book Section	Knowledge Management in Organizations
Cho et al. (2018)	Book Section	Advances in Production Management Systems: Smart Manufacturing for Industry 4.0
Hwang et al. (2018)	Conference Paper	Thirteenth International Conference on Digital Information Management
Mashhadi et al. (2018)	Journal Article	Procedia Manufacturing
Massaro et al. (2018)	Journal Article	International Journal of Artificial Intelligence & Applications
Paolanti et al. (2018)	Journal Article	IEEE/ASME International Conference on Mechatronic and Embedded Systems and Applications
Schmidt and Wang (2018)	Journal Article	Procedia Manufacturing
Sezer et al. (2018)	Conference Paper	IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC)
Strauß et al. (2018)	Conference Paper	IEEE International Conference on Big Data (Big Data)
Susto et al. (2018)	Journal Article	Control Engineering Practice

Article	Type	Periodical
Uhlmann et al. (2018)	Journal Article	Procedia Manufacturing
Bergmann and Klein (2019)	Conference Paper	Proceedings of the 16th International Conference on Informatics in Control, Automation and Robotics
Brueno and Vita (2019)	Conference Paper	IEEE International Conference on Smart Computing
Calabrese et al. (2019)	Journal Article	Procedia Manufacturing
Cao et al. (2019)	Journal Article	Procedia Computer Science
Einabadi et al. (2019)	Journal Article	IFAC-PapersOnLine
Hesser and Markert (2019)	Journal Article	Manufacturing Letters
Him et al. (2019)	Conference Paper	Asia-Pacific Signal and Information Processing Association Annual Summit and Conference
Lee et al. (2019)	Journal Article	Procedia CIRP
Lindemann et al. (2019)	Journal Article	Procedia CIRP
Nguyen and Medjaher (2019)	Journal Article	Reliability Engineering & System Safety
Pintoa and Cerquitelli (2019)	Journal Article	Procedia Computer Science
Traini et al. (2019)	Journal Article	IFAC-PapersOnLine
Zenisek et al. (2019)	Journal Article	Computers & Industrial Engineering
Angadi et al. (2020)	Journal Article	IFAC-PapersOnLine
Angadi et al. (2020)	Journal Article	Procedia Manufacturing
Aydemir and Acar (2020)	Journal Article	Journal of Manufacturing Systems
Bodo et al. (2020)	Conference Paper	IEEE International Instrumentation and Measurement Technology Conference
Calabrese et al. (2020)	Journal Article	Information
Caroff et al. (2020)	Journal Article	Procedia Manufacturing
Chen et al. (2020)	Journal Article	IEEE Access
Fernandes et al. (2020)	Book Section	Distributed Computing and Artificial Intelligence, 16th International Conference
Fila et al. (2020)	Journal Article	Procedia Computer Science
Foukalas (2020)	Journal Article	Computers & Electrical Engineering
Joung et al. (2020)	Journal Article	Procedia Manufacturing
Kiangala and Wang (2020)	Journal Article	IEEE Access

Article	Type	Periodical
Kozłowski et al. (2020)	Journal Article	Expert Systems with Applications
Luo et al. (2020)	Journal Article	Robotics and Computer-Integrated Manufacturing
Koca et al. (2020)	Conference Paper	International Conference on Development and Application Systems
Patel et al. (2020)	Conference Paper	IEEE International Conference on Big Data
Pierleoni et al. (2020)	Conference Paper	International Conference on Intelligent Engineering and Management (ICIEM)
Rivas et al. (2020)	Book Section	Conference on Soft Computing Models in Industrial and Environmental Applications
Ruiz-Sarmiento et al. (2020)	Journal Article	Engineering Applications of Artificial Intelligence
Yu et al. (2020)	Journal Article	IEEE Transactions on Industrial Informatics
Al-amri et al. (2021)	Conference Paper	Applied Sciences
Ansari et al. (2021)	Journal Article	CIRP Annals
Kohl, Ansari, and Sihn (2021)	Conference Paper	WGAB Conference
Bourezza and Mousrij (2021)	Journal Article	Advances in Intelligent Systems and Computing
Brügge et al.Dogan and Birant (2021)	Conference Paper	Expert Systems with Applications
Drakaki et al. (2021)	Conference Paper	Procedia Computer Science
Ehtesham Rasi and Sohanian (2021)	Journal Article	Journal of Modelling in Management
ElMaraghy et al. (2021)	Journal Article	CIRP Annals
Gerdenitsch et al. (2021)	Book Section	Smart Technologies for Precision Assembly
Huber et al. (2021)	Conference Paper	HMD Praxis der Wirtschaftsinformatik
Kohl, Ansari, and Sihn (2021)	Conference Paper	WGAB Conference on Competence development and Learning Assistance
Ladj et al. (2021)	Journal Article	Journal of Manufacturing Systems
Leaman et al. (2021)	Conference Paper	Acoustics Australia
Shin et al. (2021)	Journal Article	Energy
Zirpins et al. (2021)	Journal Article	Journal of Manufacturing Systems

Article	Type	Periodical
Samatas et al. (2021)	Conference Paper	IEEE World AI IoT 2021
Drakaki et al. (2021)	Journal Article	Procedia Computer Science
Masrour et al. (2021)	Journal Article	Advances in Intelligent Systems and Computing
Partovi et al. (2021)	Journal Article	Advances in Service-Oriented and Cloud Computing
Zufle et al. (2021)	Conference Paper	IEEE 19th International Conference 2021
Maktoubian et al. (2021)	Conference Paper	Forest
Zhai et al. (2021)	Journal Article	Journal of Manufacturing Systems
Passath et al. (2021)	Conference Paper	Tehnički glasnik
Pech et al. (2021)	Journal Article	Sensors
Rossini et al. (2021)	Conference Paper	IEEE International Conference 2021
Leaman et al. (2021)	Conference Paper	Acoustics Australia
Sihn et al. (2021)	Conference Paper	WGAB Conference
Curry et al. (2022)	Book Chapter	Technologies and Applications for Big Data Value
Quandt et al. (2022)	Conference Paper	Procedia CIRP
Abidi et al. (2022)	Conference Paper	Sustainability
D'Amico et al. (2022)	Conference Paper	CIRP Journal of Manufacturing Science and Technology
Raj et al. (2022)	Book Chapter	Applied Edge AI
Wang et al. (2022)	Conference Paper	Design and Operation of Production 2022
Velmurugan et al. (2022)	Conference Paper	Materials Today: Proceedings
Teern et al. (2022)	Conference Paper	BIR 2022 Workshops and Doctoral
Canciglieri et al. (2022)	Journal Article	IFIP Advances in Information and Communication Technology
Xia et al. (2022)	Journal Article	Journal of Manufacturing Systems
Zhang et al. (2022)	Journal Article	Journal of Manufacturing Systems
Cao et al. (2022)	Conference Paper	Robotics and Computer-Integrated Manufacturing
Liu et al. (2022)	Conference Paper	Robotics and Computer-Integrated Manufacturing
Liu et al. (2022)	Conference Paper	IEEE Transactions on Industrial Informatics
Haase et al. (2022)	Conference Paper	I40M (Industrie 4.0 Management)

Article	Type	Periodical
Singh et al. (2022)	Journal Article	Emerging Technologies in Computer Engineering
Biegel et al. (2022)	Conference Paper	SSRN Journal (SSRN Electronic Journal)
Balas et al. (2022)	Conference Paper	Communications in Computer and Information Science
Martinez-Gil et al. (2022)	Conference Paper	Procedia Computer Science
Shcherbakov et al. (2022)	Conference Paper	ACM Transactions on Cyber-Physical Systems
Keleko et al. (2022)	Conference Paper	AI and Ethics
Saini et al. (2022)	Conference Paper	Archives of Computational Methods in Engineering
Cheng et al. (2022)	Conference Paper	Sensors
Wellsandt et al. (2022)	Conference Paper	Annual Reviews in Control
Unal et al. (2022)	Book Chapter	Technologies and Applications for Big Data
Tanane et al. (2022)	Book Chapter	Product Lifecycle Management
Ansari et al. (2022)	Book Chapter	Supply Network Dynamics and Control
Li et al. (2022)	Book Chapter	Procedia CIRP
Shaheen et al. (2022)	Book Chapter	Processes
Divya et al. (2023)	Journal Article	Journal of Quality in Maintenance Engineering
Wang et al. (2023)	Conference Paper	Expert Systems with Applications
Bhattacharya et al. (2023)	Conference Paper	Systems
Ghansah et al. (2023)	Conference Paper	Construction Innovation
Rosati et al. (2023)	Journal Article	Journal of Intelligent Manufacturing
Oladapo et al. (2023)	Book Chapter	Data Analytics and Computational Intelligence
Pagano et al. (2023)	Conference Paper	Decision Analytics Journal
Soori et al. (2023)	Conference Paper	Sustainable Manufacturing and Service Economics
Molęda et al. (2023)	Conference Paper	Sensors
Wang et al. (2023)	Journal Article	IET Collaborative Intelligent Manufacturing
Gaal et al. (2023)	Conference Paper	WGAB Conference
Li et al. (2023)	Journal Article	Robotics and Computer-Integrated Manufacturing
Ansari et al. (2023)	Journal Article	CIRP Annals

Article	Type	Periodical
Ghobakhloo et al. (2023)	Journal Article	Journal of Cleaner Production
Runji et al. (2023)	Journal Article	International Journal of Precision Engineering and Manufacturing-Green Technology
Rivera et al. (2023)	Book Chapter	Studies in Big Data
Converso et al. (2023)	Journal Article	Applied Sciences
Zhang et al. (2023)	Journal	Advanced Engineering Informatics
Yitmen et al. (2023)	Book Chapter	Cognitive Digital Twins for Smart Lifecycle Management of Built Environment and Infrastructure
Murugiah et al. (2023)	Journal Article	International Journal of Communication Systems
Kohl et al. (2023)	Book Chapter	Procedia CIRP
Jan et al. (2023)	Conference Paper	Expert Systems with Applications
Zhang et al. (2023)	Conference Paper	Advanced Engineering Informatics
Teoh et al. (2023)	Book Chapter	IEEE Internet of Things Journal
Fordal et al. (2023)	Conference Paper	Advances in Manufacturing
Nayyar et al. (2023)	Conference Paper	Contributions to Environmental Sciences & Innovative Business Technology
Taqi et al. (2023)	Journal Article	Operations Management Research
Iftikhar et al. (2023)	Journal Article	Internet of Things
Khan et al. (2023)	Conference Paper	Green Technologies and Sustainability
Rajawat et al. (2023)	Journal Article	Electronics
Aboshosha et al. (2023)	Book Chapter	Scientific Reports
Xie et al. (2023)	Book Chapter	Big Data and Cognitive Computing
Pinciroli et al. (2023)	Conference Paper	Reliability Engineering & System Safety
Ensafi et al. (2023)	Journal Article	Journal of Building Engineering
Soori et al. (2023)	Journal Article	Cognitive Robotics
Xie et al. (2023)	Conference Paper	Renewable Energy
Javaid et al. (2023)	Journal Article	Cognitive Robotics
Dummer et al. (2023)	Conference Paper	Towards a Smart, Resilient, and Sustainable Industry
Nunes et al. (2023)	Journal Article	CIRP Journal of Manufacturing Science and Technology

Article	Type	Periodical
van Oudenhoven et al. (2023)	Journal Article	International Journal of Production Research
Hosseinzadeh et al. (2023)	Conference Paper	Manufacturing Letters
Siraskar et al. (2023)	Journal Article	Artificial Intelligence Review
Arista et al. (2023)	Journal Article	Journal of Manufacturing Systems
La Iglesia et al. (2023)	Conference Paper	Procedia CIRP
Abbasi et al. (2023)	Book Chapter	New Trends in Disruptive Technologies, Tech Ethics, and Artificial Intelligence
Li et al. (2023)	Journal Article	Engineering
Eswaran et al. (2023)	Conference Paper	Expert Systems with Applications
Feng, Zhang, et al. (2023)	Journal Article	Journal of Manufacturing Systems
Mathew et al. (2023)	Book Chapter	New Horizons for Industry 4.0
Mezgebe et al. (2023)	Conference Paper	Scientific African
Nagy et al. (2023)	Journal Article	Applied Sciences
Azari et al. (2023)	Journal Article	IEEE Access
Bordegoni et al. (2023)	Journal Article	Journal of Computing and Information Science in Engineering
Mallioris et al. (2024)	Journal Article	CIRP Journal of Manufacturing Science and Technology
Mosbah et al. (2024)	Conference Paper	Lecture Notes in Computer Science
Islam et al. (2024)	Conference Paper	International Congress and Workshop on Industrial AI and eMaintenance 2023
Jaenal et al. (2024)	Journal Article	Engineering Applications of Artificial Intelligence
Chakroun et al. (2024)	Journal Article	Journal of Intelligent Manufacturing
Justus et al. (2024)	Journal Article	International Journal of System Assurance Engineering and Management
Mukherjee et al. (2024)	Journal Article	Computers & Industrial Engineering
Vimal et al. (2024)	Book Chapter	Environmental Footprints and eco-design of Products and Processes
Scaife (2024)	Conference Paper	Results in Engineering
Tian et al. (2024)	Journal Article	Expert Systems

Article	Type	Periodical
Kumar et al. (2024)	Conference Paper	International Congress and Workshop on Industrial AI and eMaintenance
Zheng et al. (2024)	Journal	Journal of Intelligent Manufacturing
Abdelillah et al. (2024)	Conference Paper	Model and Data Engineering
Paranitharan et al. (2024)	Book Chapter	Industry 4.0 Technologies
Marti-Puig et al. (2024)	Journal Article	Computers & Industrial Engineering
Arena et al. (2024)	Journal Article	Engineering Applications of Artificial Intelligence

8 List of Figures

9 List of Tables

10 References

Abdelillah, F. M., Nora, H., Samir, O., & Sidi-Mohammed, B. (2024). Hybrid Data-Driven and Knowledge-Based Predictive Maintenance Framework in the Context of Industry 4.0. In M. Mosbah, T. Kechadi, L. Bellatreche, & F. Gargouri (Eds.), *Lecture Notes in Computer Science. Model and Data Engineering* (Vol. 14396, pp. 319–337). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-49333-1_23

Abidi, M. H., Mohammed, M. K., & Alkhalefah, H. (2022). Predictive Maintenance Planning for Industry 4.0 Using Machine Learning for Sustainable Manufacturing. *Sustainability*, *14*(6), 3387. https://doi.org/10.3390/su14063387

Adamczyk, B. S., Szejka, A. L., & Canciglieri, O. (2020). Knowledge-based expert system to support the semantic interoperability in smart manufacturing. *Computers in Industry*, *115*, 103161. https://doi.org/10.1016/j.compind.2019.103161

Akimov, N., Kurmanov, N., Uskelenova, A., Aidargaliyeva, N., Mukhiyayeva, D., Rakhimova, S., Raimbekov, B., & Utegenova, Z. (2023). Components of education 4.0 in open innovation competence frameworks: Systematic review. *Journal of Open Innovation: Technology, Market, and Complexity*, *9*(2), 100037. https://doi.org/10.1016/j.joitmc.2023.100037

Al Sharou, K., Li, Z [Zhenhao], & Specia, L. (2021). Towards a Better Understanding of Noise in Natural Language Processing. In R. Mitkov & G. Angelova (Eds.), *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2021)* (pp. 53–62). INCOMA Ltd. https://aclanthology.org/2021.ranlp-1.7

Alammar, J., & Grootendorst, M. (2024). *Hands-On Large Language Models: Language Understanding and Generation*. O'Reilly Media, Inc.

Al-amri, R., Murugesan, R. K., Man, M., Abdulateef, A. F., Al-Sharafi, M. A., & Alkahtani, A. A. (2021). A Review of Machine Learning and Deep Learning Techniques for Anomaly Detection in IoT Data. *Applied Sciences*, *11*(12), 5320. https://doi.org/10.3390/app11125320

Alec Radford, Jeff Wu, Rewon Child, David Luan, Dario Amodei, & Ilya Sutskever (2019). Language Models are Unsupervised Multitask Learners. In https://api.semanticscholar.org/CorpusID:160025533

Alessandro Moschitti, Q. C. R. I., Bo Pang, G., & Walter Daelemans, U. o. A. (Eds.) (2014). *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Association for Computational Linguistics.

Ansari, F. (2019). Knowledge Management 4.0: Theoretical and Practical Considerations in Cyber Physical Production Systems. *IFAC-PapersOnLine*, *52*(13), 1597–1602. https://doi.org/10.1016/j.ifacol.2019.11.428

Ansari, F. (2020). Cost-based text understanding to improve maintenance knowledge intelligence in manufacturing enterprises. *Computers & Industrial Engineering*, *141*, 106319. https://doi.org/10.1016/j.cie.2020.106319

Ansari, F., Glawar, R., & Nemeth, T. (2019). PriMa: a prescriptive maintenance model for cyber-physical production systems. *International Journal of Computer Integrated Manufacturing*, *32*(4-5), 482–503. https://doi.org/10.1080/0951192X.2019.1571236

Ansari, F., Hold, P., & Khobreh, M. (2020). A knowledge-based approach for representing jobholder profile toward optimal human–machine collaboration in cyber physical production systems. *CIRP Journal of Manufacturing Science and Technology*, *28*, 87–106. https://doi.org/10.1016/j.cirpj.2019.11.005

Ansari, F., Hold, P., Mayrhofer, W., Schlund, S., & Sihn, W [Wilfried] (2018). AUTODIDACT: Introducing the Concept of Mutual Learning into a Smart Factory Industry 4.0. *International Association for Development of the Information Society*.

Ansari, F., Khobreh, M., Seidenberg, U., & Sihn, W [Wilfried] (2018). A problem-solving ontology for human-centered cyber physical production systems. *CIRP Journal of Manufacturing Science and Technology*, *22*, 91–106. https://doi.org/10.1016/j.cirpj.2018.06.002

Ansari, F., & Kohl, L. (2022). AI-Enhanced Maintenance for Building Resilience and Viability in Supply Chains. In A. Dolgui, D. Ivanov, & B. Sokolov (Eds.), *Springer Series in Supply Chain Management. Supply Network Dynamics and Control* (Vol. 20, pp. 163–185). Springer International Publishing. https://doi.org/10.1007/978-3-031-09179-7_8

Ansari, F., Kohl, L., Giner, J., & Meier, H. (2021). Text mining for AI enhanced failure detection and availability optimization in production systems. *CIRP Annals.* Advance online publication. https://doi.org/10.1016/j.cirp.2021.04.045

Ansari, F., Kohl, L., & Sihn, W [Wilfried] (2023). A competence-based planning methodology for optimizing human resource allocation in industrial maintenance. *CIRP Annals*, *72*(1), 389–392. https://doi.org/10.1016/j.cirp.2023.04.050

Apiletti, D., Barberis, C., Cerquitelli, T., Macii, A., Macii, E., Poncino, M., & Ventura, F. (2018). iSTEP, an Integrated Self-Tuning Engine for Predictive Maintenance in Industry 4.0. In *2018 IEEE Intl Conf on Parallel & Distributed Processing with Applications, Ubiquitous Computing & Communications, Big Data & Cloud Computing, Social Computing & Networking, Sustainable Computing & Communications (ISPA/IUCC/BDCloud/SocialCom/SustainCom)* (pp. 924–931). IEEE. https://doi.org/10.1109/BDCloud.2018.00136

Archer, L. B. (1984). Systematic method for designers. In, Cross, N. *Developments in Design Methodology*, 57–82.

Arinez, J. F., Chang, Q., Gao, R. X., Xu, C., & Zhang, J. (2020). Artificial Intelligence in Advanced Manufacturing: Current Status and Future Outlook. *Journal of Manufacturing Science and Engineering*, *142*(11), Article 110804. https://doi.org/10.1115/1.4047855

Atamuradov, V., Medjaher, K., Dersin, P., Lamoureux, B., & Zerhouni, N. (2017). Prognostics and Health Management for Maintenance Practitioners - Review, Implementation and Tools Evaluation. *International Journal of Prognostics and Health Management*, *8*(3). https://doi.org/10.36001/ijphm.2017.v8i3.2667

Ayad, S., Terrissa, L. S., & Zerhouni, N. (2018). An IoT approach for a smart maintenance. In *2018 International Conference on Advanced Systems and Electric Technologies (IC_ASET)* (pp. 210–214). IEEE. https://doi.org/10.1109/ASET.2018.8379861

Azari, M. S., Flammini, F., Santini, S., & Caporuscio, M. (2023). A Systematic Literature Review on Transfer Learning for Predictive Maintenance in Industry 4.0. *IEEE Access*, *11*, 12887–12910. https://doi.org/10.1109/ACCESS.2023.3239784

Balas, V. E., Sinha, G. R., Agarwal, B., Sharma, T. K., Dadheech, P., & Mahrishi, M. (Eds.). (2022). *Communications in Computer and Information Science*. *Emerging Technologies in Computer Engineering: Cognitive Computing and Intelligent IoT*. Springer International Publishing. https://doi.org/10.1007/978-3-031-07012-9

Banik, R. (2018). *Hands-on recommendation systems with Python: Start building powerful and personalized, recommendation engines with Python*. Packt Publishing.

Bhattacharya, M., Penica, M., O'Connell, E., Southern, M., & Hayes, M. (2023). Human-in-Loop: A Review of Smart Manufacturing Deployments. *Systems*, *11*(1), 35. https://doi.org/10.3390/systems11010035

Biedermann, H. (Ed.). (2016a). *Industrial Engineering und Management*. Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-12097-9

Biedermann, H. (2016b). Lean Smart Maintenance. In H. Biedermann (Ed.), *Industrial Engineering und Management* (pp. 119–141). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-12097-9_7

Biegel, T., Jourdan, N., Madreiter, T., Kohl, L., Fahle, S., Ansari, F., Kuhlenkötter, B., & Metternich, J. (2022). Combining Process Monitoring with Text Mining for Anomaly Detection in Discrete Manufacturing. *SSRN Electronic Journal.* Advance online publication. https://doi.org/10.2139/ssrn.4073942

Bonney, M. (2000). Reflections on production planning and control (PPC). *Gestão & Produção*, *7*(3), 181–207. https://doi.org/10.1590/S0104-530X2000000300002

Bordegoni, M., & Ferrise, F. (2023). Exploring the Intersection of Metaverse, Digital Twins, and Artificial Intelligence in Training and Maintenance. *Journal of Computing and Information Science in Engineering*, *23*(6), Article 060806. https://doi.org/10.1115/1.4062455

Bourezza, E. M., & Mousrij, A. (2021). Towards a Platform to Implement an Intelligent and Predictive Maintenance in the Context of Industry 4.0. In T. Masrour, A. Cherrafi, & I. El Hassani (Eds.), *Advances in Intelligent Systems and Computing. Artificial Intelligence and Industrial Applications* (Vol. 1193, pp. 33–44). Springer International Publishing. https://doi.org/10.1007/978-3-030-51186-9_3

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J [Jeffrey], Winter, C., . . . Amodei, D. (2020, May 28). *Language Models are Few-Shot Learners*. http://arxiv.org/pdf/2005.14165v4

Brügge, F., Taparia, A., & Lueth, K. L. (2023). *Predictive Maintenance & Asset Performance Market Report 2023–2028.*

Brügge, F., Wopata, M., Wegner, P., & Lueth, K. L. (2021). *Predictive Maintenance Market Report 2021–2026.* IoT Analytics.

Brundage, M. P., Sexton, T., Hodkiewicz, M., Dima, A., & Lukens, S. (2021). Technical language processing: Unlocking maintenance knowledge. *Manufacturing Letters*, *27*, 42–46. https://doi.org/10.1016/j.mfglet.2020.11.001

Buchgeher, G., Gabauer, D., Martinez-Gil, J., & Ehrlinger, L. (2021). Knowledge Graphs in Manufacturing and Production: A Systematic Literature Review. *IEEE Access*, *9*, 55537–55554. https://doi.org/10.1109/ACCESS.2021.3070395

Cachada, A., Moreira, P. M., Romero, L., Barbosa, J [Jose], Leitno, P., Gcraldcs, C. A., Deusdado, L., Costa, J., Teixeira, C., Teixeira, J., & Moreira, A. H. (2018). Maintenance 4.0: Intelligent and Predictive Maintenance System Architecture. In *2018 IEEE 23rd International Conference on Emerging Technologies and Factory Automation (ETFA)* (pp. 139–146). IEEE. https://doi.org/10.1109/ETFA.2018.8502489

Calabrese, M., Cimmino, M., Fiume, F., Manfrin, M., Romeo, L., Ceccacci, S., Paolanti, M., Toscano, G., Ciandrini, G., Carrotta, A., Mengoni, M., Frontoni, E., & Kapetis, D. (2020). SOPHIA: An Event-Based IoT and Machine Learning Architecture for Predictive Maintenance in Industry 4.0. *Information*, *11*(4), 202. https://doi.org/10.3390/info11040202

Canciglieri Junior, O., Noël, F., Rivest, L., & Bouras, A. (Eds.). (2022). *IFIP Advances in Information and Communication Technology*. *Product Lifecycle Management. Green and Blue Technologies to Support Smart and Sustainable Organizations*. Springer International Publishing. https://doi.org/10.1007/978-3-030-94335-6

Candanedo, I. S., Nieves, E. H., González, S. R., Martín, M. T. S., & Briones, A. G. (2018). Machine Learning Predictive Model for Industry 4.0. In L. Uden, B. Hadzima, & I.-H. Ting (Eds.), *Communications in Computer and Information Science. Knowledge Management in Organizations* (Vol. 877, pp. 501–510). Springer International Publishing. https://doi.org/10.1007/978-3-319-95204-8_42

Cao, Q., Zanni-Merk, C., Samet, A., Reich, C., Beuvron, F. B. de, Beckmann, A., & Giannetti, C. (2022). KSPMI: A Knowledge-based System for Predictive Maintenance in Industry 4.0. *Robotics and Computer-Integrated Manufacturing*, *74*, 102281. https://doi.org/10.1016/j.rcim.2021.102281

Carvalho, A., Chouchene, A., Lima, T., & Charrua-Santos, F. (2020). Cognitive Manufacturing in Industry 4.0 toward Cognitive Load Reduction: A Conceptual Framework. *Applied System Innovation*, *3*(4), 55. https://doi.org/10.3390/asi3040055

Carvalho, T. P., Soares, F. A. A. M. N., Vita, R., Da Francisco, R. P., Basto, J. P., & Alcalá, S. G. S. (2019). A systematic literature review of machine learning methods applied to predictive maintenance. *Computers & Industrial Engineering*, *137*, 106024. https://doi.org/10.1016/j.cie.2019.106024

Chen, B., Liu, Y [Yan], Zhang, C [Chunhua], & Wang, Z [Zhongren] (2020). Time Series Data for Equipment Reliability Analysis With Deep Learning. *IEEE Access*, *8*, 105484–105493. https://doi.org/10.1109/access.2020.3000006

Cheng, X., Chaw, J. K., Goh, K. M., Ting, T. T., Sahrani, S., Ahmad, M. N., Abdul Kadir, R., & Ang, M. C. (2022). Systematic Literature Review on Visual Analytics of Predictive Maintenance in the Manufacturing Industry. *Sensors (Basel, Switzerland)*, *22*(17). https://doi.org/10.3390/s22176321

Chicaiza, J., & Valdiviezo-Diaz, P. (2021). A Comprehensive Survey of Knowledge Graph-Based Recommender Systems: Technologies, Development, and Contributions. *Information*, *12*(6), 232. https://doi.org/10.3390/info12060232

Cimini, C., Pirola, F., Pinto, R., & Cavalieri, S. (2020). A human-in-the-loop manufacturing control architecture for the next generation of production systems. *Journal of Manufacturing Systems*, *54*, 258–271. https://doi.org/10.1016/j.jmsy.2020.01.002

Çınar, Z. M., Abdussalam Nuhu, A., Zeeshan, Q., Korhan, O., Asmael, M., & Safaei, B. (2020). Machine Learning in Predictive Maintenance towards Sustainable Smart Manufacturing in Industry 4.0. *Sustainability*, *12*(19), 8211. https://doi.org/10.3390/su12198211

Cohen, K. B., & Verspoor, K. (2013). Applied Text Mining. In W. Dubitzky, O. Wolkenhauer, K.-H. Cho, & H. Yokota (Eds.), *Encyclopedia of Systems Biology* (pp. 33–37). Springer New York. https://doi.org/10.1007/978-1-4419-9863-7_133

Cohen, Y., Naseraldin, H., Chaudhuri, A., & Pilati, F. (2019). Assembly systems in Industry 4.0 era: a road map to understand Assembly 4.0. *The International Journal of Advanced Manufacturing Technology*, *105*(9), 4037–4054. https://doi.org/10.1007/s00170-019-04203-1

Colledani, M., Magnanini, M. C., & Tolio, T. (2018). Impact of opportunistic maintenance on manufacturing system performance. *CIRP Annals*, *67*(1), 499–502. https://doi.org/10.1016/j.cirp.2018.04.078

Converso, G., Gallo, M., Murino, T., & Vespoli, S. (2023). Predicting Failure Probability in Industry 4.0 Production Systems: A Workload-Based Prognostic Model for Maintenance Planning. *Applied Sciences*, *13*(3), 1938. https://doi.org/10.3390/app13031938

Corallo, A., Crespino, A. M., Del Vecchio, V., Lazoi, M., & Marra, M. (2023). Understanding and Defining Dark Data for the Manufacturing Industry. *IEEE Transactions on Engineering Management*, *70*(2), 700–712. https://doi.org/10.1109/TEM.2021.3051981

Cui, L., Huang, J [Jinfeng], & Zhang, F. (2017). Quantitative and Localization Diagnosis of a Defective Ball Bearing Based on Vertical–Horizontal Synchronization Signal

Analysis. *IEEE Transactions on Industrial Electronics*, *64*(11), 8695–8706. https://doi.org/10.1109/TIE.2017.2698359

Cui, L., Huang, J [Jinfeng], Zhang, F., & Chu, F. (2019). HVSRMS localization formula and localization law: Localization diagnosis of a ball bearing outer ring fault. *Mechanical Systems and Signal Processing*, *120*, 608–629. https://doi.org/10.1016/j.ymssp.2018.09.043

Dalzochio, J., Kunst, R., Pignaton, E., Binotto, A., Sanyal, S., Favilla, J., & Barbosa, J [Jorge] (2020). Machine learning and reasoning for predictive maintenance in Industry 4.0: Current status and challenges. *Computers in Industry*, *123*, 103298. https://doi.org/10.1016/j.compind.2020.103298

Deutsch, J., He, M., & He, D. (2017). Remaining Useful Life Prediction of Hybrid Ceramic Bearings Using an Integrated Deep Learning and Particle Filter Approach. *Applied Sciences*, *7*(7), 649. https://doi.org/10.3390/app7070649

Devlin, J., Chang, M.-W., Lee, K [Kenton], & Toutanova, K. (2018, October 11). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. http://arxiv.org/pdf/1810.04805v2

Di Wu, Zhang, P., Yu, Z., Gao, Y., Zhang, H., Chen, H., Chen, S., & Tian, Y. (2022). Progress and perspectives of in-situ optical monitoring in laser beam welding: Sensing, characterization and modeling. *Journal of Manufacturing Processes*, *75*, 767–791. https://doi.org/10.1016/j.jmapro.2022.01.044

DIN 31051:2019-06. *DIN 31051:2019-06, Grundlagen der Instandhaltung* (DIN 31051:2019-06). Beuth Verlag GmbH.

Divya, D., Marath, B., & Santosh Kumar, M. B. (2023). Review of fault detection techniques for predictive maintenance. *Journal of Quality in Maintenance Engineering*, *29*(2), 420–441. https://doi.org/10.1108/JQME-10-2020-0107

Dogan, A., & Birant, D. (2021). Machine learning and data mining in manufacturing. *Expert Systems with Applications*, *166*, 114060. https://doi.org/10.1016/j.eswa.2020.114060

Drakaki, M., Karnavas, Y. L., Tzionas, P., & Chasiotis, I. D. (2021). Recent Developments Towards Industry 4.0 Oriented Predictive Maintenance in Induction Motors. *Procedia Computer Science*, *180*, 943–949. https://doi.org/10.1016/j.procs.2021.01.345

Dummer, W., Gaal, A., Sobottka, T., & Ansari, F. (2023). A Genetic Algorithm Approach for Medical Resident Scheduling in Austria. In Y. Borgianni, D. T. Matt, M. Molinaro, & G. Orzes (Eds.), *Lecture Notes in Networks and Systems. Towards a Smart, Resilient* *and Sustainable Industry* (Vol. 745, pp. 321–332). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-38274-1_27

Ehtesham Rasi, R., & Sohanian, M. (2021). A multi-objective optimization model for sustainable supply chain network with using genetic algorithm. *Journal of Modelling in Management*, *16*(2), 714–727. https://doi.org/10.1108/JM2-06-2020-0150

ElMaraghy, H., Monostori, L [Laszlo], Schuh, G [Guenther], & ElMaraghy, W. (2021). Evolution and future of manufacturing systems. *CIRP Annals*, *70*(2), 635–658. https://doi.org/10.1016/j.cirp.2021.05.008

European Commission, Directorate-General for Employment, Social Affairs, & Inclusion. (2017). *ESCO handbook – European skills, competences, qualifications and occupations*. Publications Office. https://doi.org/10.2767/934956

Färber, M., Bartscherer, F., Menne, C., & Rettinger, A. (2017). Linked data quality of DBpedia, Freebase, OpenCyc, Wikidata, and YAGO. *Semantic Web*, *9*(1), 77–129. https://doi.org/10.3233/SW-170275

Feng, K., Ji, J. C., Ni, Q., Li, Y [Yifan], Mao, W., & Liu, L. (2023). A novel vibration-based prognostic scheme for gear health management in surface wear progression of the intelligent manufacturing system. *Wear*, *522*, 204697. https://doi.org/10.1016/j.wear.2023.204697

Feng, Q., Zhang, Y., Sun, B., Guo, X., Fan, D., Ren, Y., Song, Y., & Wang, Z [Zili] (2023). Multi-level predictive maintenance of smart manufacturing systems driven by digital twin: A matheuristics approach. *Journal of Manufacturing Systems*, *68*, 443–454. https://doi.org/10.1016/j.jmsy.2023.05.004

Fensel, D., Şimşek, U., Angele, K., Huaman, E., Kärle, E., Panasiuk, O., Toma, I., Umbrich, J., & Wahler, A. (2020). *Knowledge Graphs*. Springer International Publishing. https://doi.org/10.1007/978-3-030-37439-6

Fila, R., Khaili, M. E., & Mestari, M. (2020). Cloud Computing for Industrial Predictive Maintenance Based on Prognostics and Health Management. *Procedia Computer Science*, *177*, 631–638. https://doi.org/10.1016/j.procs.2020.10.090

Fink, O., Netland, T., & Feuerriegel, S. (2021, June 21). *Artificial intelligence across company borders*. http://arxiv.org/pdf/2107.03912v1

Flatt, H., Koch, N., Rocker, C., Gunter, A., & Jasperneite, J. (2015, September 8–11). A context-aware assistance system for maintenance applications in smart factories based on augmented reality and indoor localization. In *2015 IEEE 20th Conference on*

Emerging Technologies & Factory Automation (ETFA) (pp. 1–4). IEEE. https://doi.org/10.1109/ETFA.2015.7301586

Foukalas, F. (2020). Cognitive IoT platform for fog computing industrial applications. *Computers & Electrical Engineering*, *87*, 106770. https://doi.org/10.1016/j.compeleceng.2020.106770

Gaal, A., Dummer, W., Sobottka(1, T., Ansari, F., & Schlund, S. (2023). Advancing Medical Resident Scheduling. In J. Deuse (Ed.), *How can industrial management contribute to a brighter future?* (pp. 115–134). GITO mbH Verlag. https://doi.org/10.30844/wgab_2023_7

Gad, A. F. (2023). PyGAD: an intuitive genetic algorithm Python library. *Multimedia Tools and Applications.* Advance online publication. https://doi.org/10.1007/s11042-023-17167-y

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, *35*(2), 137–144. https://doi.org/10.1016/j.ijinfomgt.2014.10.007

Gao, R. X., Wang, L [Lihui], Helu, M., & Teti, R. (2020). Big data analytics for smart factories of the future. *CIRP Annals*, *69*(2), 668–692. https://doi.org/10.1016/j.cirp.2020.05.002

Gartner. (2023). *Unstructured Data in Enterprise*. Research Report. Gartner, Inc.

Gartner. (2024). *Dark data*.

Gerdenitsch, C., Deinhard, L., Kern, B., Hold, P., & Egger-Lampl, S. (2021). Cognitive Assistance to Support Maintenance and Assembly Tasks: Results on Technology Acceptance of a Head-Mounted Device. In S. Ratchev (Ed.), *IFIP Advances in Information and Communication Technology. Smart Technologies for Precision Assembly* (Vol. 620, pp. 276–284). Springer International Publishing. https://doi.org/10.1007/978-3-030-72632-4_20

Geron, A. (2022). *Hands-on machine learning with scikit-learn, keras, and TensorFlow 3e* (3rd ed.). O'Reilly Media.

Gharatkar, S., Ingle, A., Naik, T., & Save, A. (2017). Review preprocessing using data cleaning and stemming technique. In *2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS)* (pp. 1–4). IEEE. https://doi.org/10.1109/ICIIECS.2017.8276011

Ghobakhloo, M., Iranmanesh, M., Foroughi, B., Babaee Tirkolaee, E., Asadi, S., & Amran, A. (2023). Industry 5.0 implications for inclusive sustainable manufacturing: An evidence-knowledge-based strategic roadmap. *Journal of Cleaner Production*, *417*, 138023. https://doi.org/10.1016/j.jclepro.2023.138023

Giustozzi, F., Saunier, J., & Zanni-Merk, C. (2018). Context Modeling for Industry 4.0: an Ontology-Based Proposal. *Procedia Computer Science*, *126*, 675–684. https://doi.org/10.1016/j.procs.2018.08.001

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. *Adaptive computation and machine learning*. MIT Press. https://books.google.co.in/books?id=Np9SDQAAQBAJ

Goyal, D., Vanraj, Pabla, B. S., & Dhami, S. S. (2017). Condition Monitoring Parameters for Fault Diagnosis of Fixed Axis Gearbox: A Review. *Archives of Computational Methods in Engineering*, *24*(3), 543–556. https://doi.org/10.1007/s11831-016-9176-1

GRLER, I. (2022). *SYSTEMS ENGINEERING: Cps erfolgreich entwickeln*. MORGAN KAUFMANN.

Gualtieri, L., Rauch, E., & Vidoni, R. (2021). Emerging research fields in safety and ergonomics in industrial collaborative robotics: A systematic literature review. *Robotics and Computer-Integrated Manufacturing*, *67*, 101998. https://doi.org/10.1016/j.rcim.2020.101998

Güler, M. G., & Geçici, E. (2020). A decision support system for scheduling the shifts of physicians during COVID-19 pandemic. *Computers & Industrial Engineering*, *150*, 106874. https://doi.org/10.1016/j.cie.2020.106874

Haarman, M., Mulders, M., & Vassiliadis, C. (2017). Predictive maintenance 4.0: predict the unpredictable. *PwC and Mainnovation*, *4*.

Hastie, T., Friedman, J., & Tibshirani, R. (2001). *The Elements of Statistical Learning*. Springer New York. https://doi.org/10.1007/978-0-387-21606-5

Henke, M. (Ed.). (2019). *Acatech Studie*. *Smart Maintenance - der Weg vom Status quo zur Zielvision*. Fraunhofer IML; acatech Deutsche Akademie der Technikwissenschaften.

Hernandez-de-Menendez, M., Morales-Menendez, R., Escobar, C. A., & McGovern, M. (2020). Competencies for Industry 4.0. *International Journal on Interactive Design and Manufacturing (IJIDeM)*, *14*(4), 1511–1524. https://doi.org/10.1007/s12008-020-00716-2

Hesser, D. F., & Markert, B. (2019). Tool wear monitoring of a retrofitted CNC milling machine using artificial neural networks. *Manufacturing Letters*, *19*, 1–4. https://doi.org/10.1016/j.mfglet.2018.11.001

Hevner, A., R, A., March, S., T, S., Park, Park, J., Ram, & Sudha (2004). Design Science in Information Systems Research. *Management Information Systems Quarterly*, *28*, 75-.

Hoang, D.-T., & Kang, H.-J. (2019). A survey on Deep Learning based bearing fault diagnosis. *Neurocomputing*, *335*, 327–335. https://doi.org/10.1016/j.neucom.2018.06.078

Hölbfer, S. A. (2014). *Modell zur Auswahl von Instandhaltungsstrategien in anlageintensiven Industriebetrieben*. BibTeX.

Hold, P., Erol, S., Reisinger, G [Gehard], & Sihn, W [Wilfried] (2017). Planning and Evaluation of Digital Assistance Systems. *Procedia Manufacturing*, *9*, 143–150. https://doi.org/10.1016/j.promfg.2017.04.024

Hossayni, H., Khan, I., Aazam, M., Taleghani-Isfahani, A., & Crespi, N. (2020). SemKoRe: Improving Machine Maintenance in Industrial IoT with Semantic Knowledge Graphs. *Applied Sciences*, *10*(18), 6325. https://doi.org/10.3390/app10186325

Hosseinioun, P., Kheirabadi, M., Kamel Tabbakh, S. R., & Ghaemi, R. (2020). A new energy-aware tasks scheduling approach in fog computing using hybrid meta-heuristic algorithm. *Journal of Parallel and Distributed Computing*, *143*, 88–96. https://doi.org/10.1016/j.jpdc.2020.04.008

Hotho, A., Nürnberger, A., & Paaß, G. (2005). A brief survey of text mining. In *Ldv Forum.* Symposium conducted at the meeting of Citeseer.

Huber, C., Passath, T., & Biedermann, H. (2021). Wissens- und lernorientierte Instandhaltung unter dem Aspekt zunehmender Digitalisierung. *HMD Praxis Der Wirtschaftsinformatik*, *58*(1), 197–211. https://doi.org/10.1365/s40702-020-00671-2

Huynh, K. T., Grall, A., & Berenguer, C. (2019). A Parametric Predictive Maintenance Decision-Making Framework Considering Improved System Health Prognosis Precision. *IEEE Transactions on Reliability*, *68*(1), 375–396. https://doi.org/10.1109/TR.2018.2829771

Iftikhar, S., Gill, S. S., Song, C., Xu, M., Aslanpour, M. S., Toosi, A. N., Du, J., Wu, H [Huaming], Ghosh, S., Chowdhury, D., Golec, M., Kumar, M., Abdelmoniem, A. M., Cuadrado, F., Varghese, B., Rana, O., Dustdar, S., & Uhlig, S. (2023). AI-based fog and edge computing: A systematic review, taxonomy and future directions. *Internet of Things*, *21*, 100674. https://doi.org/10.1016/j.iot.2022.100674

Indurkhya, N., & Damerau, F. J. (2010). *Handbook of natural language processing* (Second edition). *Chapman & Hall/CRC machine learning & pattern recognition series*. Taylor & Francis.

Iskandar, J., Moyne, J., Subrahmanyam, K., Hawkins, P., & Armacost, M. (2015). Predictive Maintenance in semiconductor manufacturing. In *2015 26th Annual SEMI Advanced Semiconductor Manufacturing Conference (ASMC)* (pp. 384–389). IEEE. https://doi.org/10.1109/ASMC.2015.7164425

Islam, M. R., Begum, S., & Ahmed, M. U. (2024). Artificial Intelligence in Predictive Maintenance: A Systematic Literature Review on Review Papers. In U. Kumar, R. Karim, D. Galar, & R. Kour (Eds.), *Lecture Notes in Mechanical Engineering. International Congress and Workshop on Industrial AI and eMaintenance 2023* (pp. 251–261). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-39619-9_18

Jafari, H. (2021). Nurse scheduling problem by considering total number of required nurses as well as nurses' preferences for working shifts: An algorithmic game-theoretic approach. *Scientia Iranica*, *0*(0), 0. https://doi.org/10.24200/sci.2021.56780.4906

Jalali, A., Heistracher, C., Schindler, A., Haslhofer, B., Nemeth, T., Glawar, R., Sihn, W [Wilfried], & Boer, P. de (2019). Predicting Time-to-Failure of Plasma Etching Equipment using Machine Learning. In *2019 IEEE International Conference on Prognostics and Health Management (ICPHM)* (pp. 1–8). IEEE. https://doi.org/10.1109/ICPHM.2019.8819404

Jan, Z., Ahamed, F., Mayer, W., Patel, N., Grossmann, G., Stumptner, M., & Kuusk, A. (2023). Artificial intelligence for industry 4.0: Systematic review of applications, challenges, and opportunities. *Expert Systems with Applications*, *216*, 119456. https://doi.org/10.1016/j.eswa.2022.119456

Jay Alammar. (2019). *The Illustrated GPT-2 (Visualizing Transformer Language Models) -- jalammar.github.io*.

Jeon, J., & Sohn, S. Y. (2015). Product failure pattern analysis from warranty data using association rule and Weibull regression analysis: A case study. *Reliability Engineering & System Safety*, *133*, 176–183. https://doi.org/10.1016/j.ress.2014.08.015

Joe Qin, S., Guo, S., Li, Z [Zheyu], Chiang, L. H., Castillo, I., Braun, B., & Wang, Z [Zhenyu] (2021). Integration of process knowledge and statistical learning for the Dow data challenge problem. *Computers & Chemical Engineering*, *153*, 107451. https://doi.org/10.1016/j.compchemeng.2021.107451

Jordan, P. W., Thomas, B., McClelland, I. L., & Weerdmeester, B. (1996). *Usability Evaluation In Industry*. CRC Press. https://doi.org/10.1201/9781498710411

Kang, Y., Cai, Z., Tan, C.-W., Huang, Q., & Liu, H [Hefu] (2020). Natural language processing (NLP) in management research: A literature review. *Journal of Management Analytics*, *7*(2), 139–172. https://doi.org/10.1080/23270012.2020.1756939

Khang, A., Rani, S., Gujrati, R., Uygun, H., & Gupta, S. (2023). *Designing Workforce Management Systems for Industry 4.0*. CRC Press. https://doi.org/10.1201/9781003357070

Kheirrouz, M., Melino, F., & Ancona, M. A. (2022). Fault detection and diagnosis methods for green hydrogen production: A review. *International Journal of Hydrogen Energy*, *47*(65), 27747–27774. https://doi.org/10.1016/j.ijhydene.2022.06.115

Kiangala, K. S., & Wang, Z [Zenghui] (2018). Initiating predictive maintenance for a conveyor motor in a bottling plant using industry 4.0 concepts. *The International Journal of Advanced Manufacturing Technology*, *97*(9-12), 3251–3271. https://doi.org/10.1007/s00170-018-2093-8

Kingma, D. P., & Welling, M. (2013, December 20). *Auto-Encoding Variational Bayes*. http://arxiv.org/pdf/1312.6114v11

Klees, M., & Wortmann, B. (2023). *Roadmap Predictive Maintenance*. https://publica.fraunhofer.de/handle/publica/441483

Kohl, L., & Ansari, F. (2023). A Knowledge Graph-based Learning Assistance Systems for Industrial Maintenance. *Procedia CIRP*(in Press).

Kohl, L., Ansari, F., & Sihn, W [Wilfried]. (2021). A Modular Federated Learning Architecture for Integration of AI-enhanced Assistance in Industrial Maintenance - A novel architecture for enhancing industrial maintenance management systems in the automotive and semiconductor industry. In W. Sihn & S. Schlund (Eds.), *Competence development and learning assistance systems for the data-driven future* (pp. 229–242). Goto Verlag. https://doi.org/10.30844/wgab_2021_14

Kohl, L., Eschenbacher, S., Besinger, P., & Ansari, F. (2024). Large Language Model-based Chatbot for Improving Human-Centricity in Maintenance Planning and Operations. *PHM Society European Conference*, *8*(1), 12. https://doi.org/10.36001/phme.2024.v8i1.4098

Kohl, L., Fuchs, B., Berndt, R., Valtiner, D., Ansari, F., & Schlund, S. (2021). Künstliche Intelligenz im Kompetenzmanagement. *Zeitschrift Für Wirtschaftlichen Fabrikbetrieb*, *116*(7-8), 534–537. https://doi.org/10.1515/zwf-2021-0100

Kohl, L., Fuchs, B., Valtiner, D., Ansari, F., & Schlund, S. (2021). Künstliche Intelligenz im Kompetenzmanagemen: Ein Fallbeispiel aus der Halbleiterindustrie. *Zeitschrift Für Wirtschaftlichen Fabrikbetrieb*(Jahrg. 116 (2021) 7-8).

Kohl, L., Stricker, P., Reisinger, J., & Ansari, F. (2024). digiTeachVR: Digitally-enhanced Teaching Platform for improving data science skills and virtual reality competences in cross-disciplinary engineering education. In *Proceedings of the Conference on Learning Factories (CLF)*.

Kovacs, K., Ansari, F., Geisert, C., Uhlmann, E., Glawar, R., & Sihn, W [Wilfried]. (2019). A Process Model for Enhancing Digital Assistance in Knowledge-Based Maintenance. In J. Beyerer, C. Kühnert, & O. Niggemann (Eds.), *Technologien für die intelligente Automation. Machine Learning for Cyber Physical Systems* (Vol. 9, pp. 87–96). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-58485-9_10

Krishnamurthi, R., Gopinathan, D., & Kumar, A. (2022). Using wavelet transformation for acoustic signal processing in heavy vehicle detection and classification. In *Autonomous and Connected Heavy Vehicle Technology* (pp. 199–209). Elsevier. https://doi.org/10.1016/B978-0-323-90592-3.00011-2

Kroetzsch, M., & Weikum, G. (2016). *Journal of Web Semantics: Special Issue on Knowledge Graphs*. http://www.websemanticsjournal.org/index.php/ps/

Kuter, U. (2009). Learning and Planning (Intelligent Systems). In R. A. Meyers (Ed.), *Encyclopedia of Complexity and Systems Science* (pp. 5188–5206). Springer New York. https://doi.org/10.1007/978-0-387-30440-3_308

Kwon, D., Hodkiewicz, M. R., Fan, J., Shibutani, T., & Pecht, M. G. (2016). IoT-Based Prognostics and Systems Health Management for Industrial Applications. *IEEE Access*, *4*, 3659–3670. https://doi.org/10.1109/ACCESS.2016.2587754

Ladj, A., Wang, Z [Zhiqiang], Meski, O., Belkadi, F., Ritou, M., & Da Cunha, C. (2021). A knowledge-based Digital Shadow for machining industry in a Digital Twin perspective. *Journal of Manufacturing Systems*, *58*, 168–179. https://doi.org/10.1016/j.jmsy.2020.07.018

Leaman, F., Vicuña, C. M., & Clausen, E. (2021). A Review of Gear Fault Diagnosis of Planetary Gearboxes Using Acoustic Emissions. *Acoustics Australia*, *49*(2), 265–272. https://doi.org/10.1007/s40857-021-00217-6

Lee, J., Davari, H., Singh, J., & Pandhare, V. (2018). Industrial Artificial Intelligence for industry 4.0-based manufacturing systems. *Manufacturing Letters*, *18*, 20–23. https://doi.org/10.1016/j.mfglet.2018.09.002

Lee, J., Lapira, E., Bagheri, B., & Kao, H. (2013). Recent advances and trends in predictive manufacturing systems in big data environment. *Manufacturing Letters*, *1*(1), 38–41. https://doi.org/10.1016/j.mfglet.2013.09.005

Lee, J., Ni, J., Djurdjanovic, D., Qiu, H., & Liao, H. (2006). Intelligent prognostics tools and e-maintenance. *Computers in Industry*, *57*(6), 476–489. https://doi.org/10.1016/j.compind.2006.02.014

Li, B., Wang, Y., Dai, G., & Wang, K. (2019). Framework and case study of cognitive maintenance in Industry 4.0. *Frontiers of Information Technology & Electronic Engineering*, *20*(11), 1493–1504. https://doi.org/10.1631/FITEE.1900193

Li, C [Chengxi], Zheng, P., Yin, Y., Pang, Y. M., & Huo, S. (2023). An AR-assisted Deep Reinforcement Learning-based approach towards mutual-cognitive safe human-robot interaction. *Robotics and Computer-Integrated Manufacturing*, *80*, 102471. https://doi.org/10.1016/j.rcim.2022.102471

Li, H., Xu, F., Liu, H [Hongyi], & Zhang, X [Xuefeng] (2015). Incipient fault information determination for rolling element bearing based on synchronous averaging reassigned wavelet scalogram. *Measurement*, *65*, 1–10. https://doi.org/10.1016/j.measurement.2014.12.032

Li, J [J.], Schaefer, D., & Milisavljevic-Syed, J. (2022). A Decision-Based Framework for Predictive Maintenance Technique Selection in Industry 4.0. *Procedia CIRP*, *107*, 77–82. https://doi.org/10.1016/j.procir.2022.04.013

Li, Q [Qun], Ban, X., & Wu, H [Haili] (2019). Design of Informationized Operation and Maintenance System for Long-distance Oil and Gas Pipelines. In A. Emrouznejad & Z. Xu (Eds.), *Proceedings of the 3rd International Conference on Computer Science and Application Engineering* (pp. 1–5). ACM. https://doi.org/10.1145/3331453.3360969

Li, Y [Yongbo], Feng, K., Liang, X., & Zuo, M. J. (2019). A fault diagnosis method for planetary gearboxes under non-stationary working conditions using improved Vold-Kalman filter and multi-scale sample entropy. *Journal of Sound and Vibration*, *439*, 271–286. https://doi.org/10.1016/j.jsv.2018.09.054

Li, Y [Yongbo], Wang, X [Xianzhi], Liu, Z [Zhenbao], Liang, X., & Si, S. (2018). The Entropy Algorithm and Its Variants in the Fault Diagnosis of Rotating Machinery: A Review. *IEEE Access*, *6*, 66723–66741. https://doi.org/10.1109/ACCESS.2018.2873782

Li, Z [Zhe], Wang, Y., & Wang, K. (2017). Intelligent predictive maintenance for fault diagnosis and prognosis in machine centers: Industry 4.0 scenario. *Advances in Manufacturing*, *5*(4), 377–387. https://doi.org/10.1007/s40436-017-0203-8

Liu, F., He, Y., Zhao, Y., Zhang, A., & Di Zhou (2019). Risk-Oriented Product Assembly System Health Modeling and Predictive Maintenance Strategy. *Sensors (Basel, Switzerland)*, *19*(9). https://doi.org/10.3390/s19092086

Liu, Q., Dong, M., Chen, F. F., Lv, W., & Ye, C. (2019). Single-machine-based joint optimization of predictive maintenance planning and production scheduling. *Robotics and Computer-Integrated Manufacturing*, *55*, 173–182. https://doi.org/10.1016/j.rcim.2018.09.007

Liu, R., Yang, B., Zio, E., & Chen, X. (2018). Artificial intelligence for fault diagnosis of rotating machinery: A review. *Mechanical Systems and Signal Processing*, *108*, 33–47. https://doi.org/10.1016/j.ymssp.2018.02.016

Liu, Z [Zhenbao], Jia, Z., Vong, C.-M., Han, J [Junwei], Yan, C., & Pecht, M. (2018). A Patent Analysis of Prognostics and Health Management (PHM) Innovations for Electrical Systems. *IEEE Access*, *6*, 18088–18107. https://doi.org/10.1109/ACCESS.2018.2818114

Lu, K., Jin, Y., Chen, Y [Yushu], Yang, Y., Hou, L., Zhang, Z [Zhiyong], Li, Z [Zhonggang], & Fu, C. (2019). Review for order reduction based on proper orthogonal decomposition and outlooks of applications in mechanical systems. *Mechanical Systems and Signal Processing*, *123*, 264–297. https://doi.org/10.1016/j.ymssp.2019.01.018

Luo, W., Hu, T., Ye, Y., Zhang, C [Chengrui], & Wei, Y. (2020). A hybrid predictive maintenance approach for CNC machine tool driven by Digital Twin. *Robotics and Computer-Integrated Manufacturing*, *65*, 101974. https://doi.org/10.1016/j.rcim.2020.101974

Lyu, P., Zhang, K [Kewei], Yu, W [Wenbing], Wang, B., & Liu, C [Chao] (2022). A novel RSG-based intelligent bearing fault diagnosis method for motors in high-noise industrial environment. *Advanced Engineering Informatics*, *52*, 101564. https://doi.org/10.1016/j.aei.2022.101564

Madreiter, T., Kohl, L., & Ansari, F. (2021). A Text Understandability Approach for Improving Reliability-Centered Maintenance in Manufacturing Enterprises. In A. Dolgui, A. Bernard, D. Lemoine, G. von Cieminski, & D. Romero (Eds.), *IFIP Advances in Information and Communication Technology. Advances in Production*

Management Systems. Artificial Intelligence for Sustainable and Resilient Production Systems (Vol. 630, pp. 161–170). Springer International Publishing. https://doi.org/10.1007/978-3-030-85874-2_17

Mark, B. G., Rauch, E., & Matt, D. T. (2021). Worker assistance systems in manufacturing: A review of the state of the art and future directions. *Journal of Manufacturing Systems*, *59*, 228–250. https://doi.org/10.1016/j.jmsy.2021.02.017

Martinez-Gil, J., Buchgeher, G., Gabauer, D., Freudenthaler, B., Filipiak, D., & Fensel, A. (2022). Root Cause Analysis in the Industrial Domain using Knowledge Graphs: A Case Study on Power Transformers. *Procedia Computer Science*, *200*, 944–953. https://doi.org/10.1016/j.procs.2022.01.292

Matyas, K. (2022). *Instandhaltungslogistik: Qualität und Produktivität steigern* (8., aktualisierte Auflage). *Praxisreihe Qualität*. Hanser, Carl.

McKinsey & Company (2022). Data Utilization in Manufacturing. *McKinsey Quarterly*.

Merkt, O. (2019). On the Use of Predictive Models for Improving the Quality of Industrial Maintenance: An Analytical Literature Review of Maintenance Strategies. In *Annals of Computer Science and Information Systems, Proceedings of the 2019 Federated Conference on Computer Science and Information Systems* (pp. 693–704). IEEE. https://doi.org/10.15439/2019F101

Mihaylov, E., & Tijdens, K. G. (2019). Measuring the Routine and Non-Routine Task Content of 427 Four-Digit ISCO-08 Occupations. *SSRN Electronic Journal.* Advance online publication. https://doi.org/10.2139/ssrn.3389681

Mikolov, T., Grave, E., Bojanowski, P., Puhrsch, C., & Joulin, A. (2017, December 26). *Advances in Pre-Training Distributed Word Representations*. http://arxiv.org/pdf/1712.09405v1

Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D., & Miller, K. J. (1990). Introduction to WordNet: An On-line Lexical Database *International Journal of Lexicography*, *3*(4), 235–244. https://doi.org/10.1093/ijl/3.4.235

Mitsotakis, D. (2023). *Computational Mathematics*. Chapman and Hall/CRC. https://doi.org/10.1201/9781003287292

Mobley, R. K. (2002). *An introduction to predictive maintenance*. Elsevier.

Moghaddam, K. (2020). A Multi-Objective Modeling Approach for Integrated Manufacturing and Preventive Maintenance Planning. *Operations and Supply Chain Management: An International Journal*, 83–99. https://doi.org/10.31387/oscm0440288

Molęda, M., Małysiak-Mrozek, B., Ding, W., Sunderam, V., & Mrozek, D. (2023). From Corrective to Predictive Maintenance-A Review of Maintenance Approaches for the Power Industry. *Sensors (Basel, Switzerland)*, *23*(13). https://doi.org/10.3390/s23135970

Monostori, L [L.], Kádár, B [B.], Bauernhansl, T., Kondoh, S., Kumara, S [S.], Reinhart, G [G.], Sauer, O [O.], Schuh, G [G.], Sihn, W [W.], & Ueda, K. (2016). Cyber-physical systems in manufacturing. *CIRP Annals*, *65*(2), 621–641. https://doi.org/10.1016/j.cirp.2016.06.005

Montero Jimenez, J. J., Schwartz, S., Vingerhoeds, R., Grabot, B., & Salaün, M. (2020). Towards multi-model approaches to predictive maintenance: A systematic literature survey on diagnostics and prognostics. *Journal of Manufacturing Systems*, *56*, 539–557. https://doi.org/10.1016/j.jmsy.2020.07.008

Mukherjee, A., Divya, A. B., Sivvani, M., & Pal, S. K. (2024). Cognitive intelligence in industrial robots and manufacturing. *Computers & Industrial Engineering*, *191*, 110106. https://doi.org/10.1016/j.cie.2024.110106

Müller, R., Ritz, F., Illium, S., & Linnhoff-Popien, C. (2020). Acoustic Anomaly Detection for Machine Sounds based on Image Transfer Learning. Advance online publication. https://doi.org/10.48550/arXiv.2006.03429

NAM. (2023). *DATA MASTERY: DATA MASTERY: Data Mastery: A Key to Industrial Competitiveness.*

Nassehi, A., Colledani, M., Kádár, B [Botond], & Lutters, E. (2022). Daydreaming factories. *CIRP Annals*, *71*(2), 671–692. https://doi.org/10.1016/j.cirp.2022.05.002

Nemeth, T., Ansari, F., & Sihn, W [Wilfried] (2019). A Maturity Assessment Procedure Model for Realizing Knowledge-Based Maintenance Strategies in Smart Manufacturing Enterprises. *Procedia Manufacturing*, *39*, 645–654. https://doi.org/10.1016/j.promfg.2020.01.439

Neo4j. (2024). *Neo4j Graph Database & Analytics – The Leader in Graph Databases -- neo4j.com*.

Neuhüttler, J., Christel, P., Le, T., Wenzel, G., & Ardilio, A. (2022). Potentials of a Metaverse for Smart Service Engineering. In *AHFE International, The Human Side of Service Engineering.* AHFE International. https://doi.org/10.54941/ahfe1002545

Nguyen, B. M., Thi Thanh Binh, H., The Anh, T., & Bao Son, D. (2019). Evolutionary Algorithms to Optimize Task Scheduling Problem for the IoT Based Bag-of-Tasks

Application in Cloud–Fog Computing Environment. *Applied Sciences*, *9*(9), 1730. https://doi.org/10.3390/app9091730

Nguyen, K. T., & Medjaher, K. (2019). A new dynamic predictive maintenance framework using deep learning for failure prognostics. *Reliability Engineering & System Safety*, *188*, 251–262. https://doi.org/10.1016/j.ress.2019.03.018

Nunes, P., Santos, J., & Rocha, E. (2023). Challenges in predictive maintenance – A review. *CIRP Journal of Manufacturing Science and Technology*, *40*, 53–67. https://doi.org/10.1016/j.cirpj.2022.11.004

O'Donovan, P., Leahy, K., Bruton, K., & O'Sullivan, D. T. J. (2015). Big data in manufacturing: a systematic mapping study. *Journal of Big Data*, *2*(1). https://doi.org/10.1186/s40537-015-0028-x

OECD. (2023). *OECD Social, Employment and Migration Working Papers.* https://doi.org/10.1787/903c19c9-en

OECD. (2024). *OECD Artificial Intelligence Papers.* https://doi.org/10.1787/8d900037-en

Pandiyan, V., Drissi-Daoudi, R., Shevchik, S., Masinelli, G., Logé, R., & Wasmer, K. (2020). Analysis of time, frequency and time-frequency domain features from acoustic emissions during Laser Powder-Bed fusion process. *Procedia CIRP*, *94*, 392–397. https://doi.org/10.1016/j.procir.2020.09.152

Pang, G., Shen, C [Chunhua], Cao, L., & van Hengel, A. den (2022). Deep Learning for Anomaly Detection. *ACM Computing Surveys*, *54*(2), 1–38. https://doi.org/10.1145/3439950

Paolanti, M., Romeo, L., Felicetti, A., Mancini, A., Frontoni, E., & Loncarski, J. (2018). Machine Learning approach for Predictive Maintenance in Industry 4.0. In *2018 14th IEEE/ASME International Conference on Mechatronic and Embedded Systems and Applications (MESA)* (pp. 1–6). IEEE. https://doi.org/10.1109/MESA.2018.8449150

Paranitharan, K. P., Velmurugan, K., Balaji, V., Venkumar, P., & Kumarasamy, R. (2024). SME 4.0: Health Monitoring of Maintenance Management Approaches in Smart Manufacturing. In V. K E K, S. Rajak, V. Kumar, R. S. Mor, & A. Assayed (Eds.), *Environmental Footprints and Eco-design of Products and Processes. Industry 4.0 Technologies: Sustainable Manufacturing Supply Chains* (pp. 99–112). Springer Nature Singapore. https://doi.org/10.1007/978-981-99-4894-9_7

Parente, M., Figueira, G., Amorim, P., & Marques, A. (2020). Production scheduling in the context of Industry 4.0: review and trends. *International Journal of Production Research*, *58*(17), 5401–5431. https://doi.org/10.1080/00207543.2020.1718794

Parks, K. M. A., Griffith, L. A., Armstrong, N. B., & Stevenson, R. A. (2020). Statistical Learning and Social Competency: The Mediating Role of Language. *Scientific Reports*, *10*(1), 3968. https://doi.org/10.1038/s41598-020-61047-6

Partovi, B., Bagheri, A., Kazarji, M. H., Pahl, C., & Barzegar, H. R. (2021). Virtual Machine Placement for Edge and Cloud Computing. In C. Zirpins, I. Paraskakis, V. Andrikopoulos, N. Kratzke, C. Pahl, N. El Ioini, A. S. Andreou, G. Feuerlicht, W. Lamersdorf, G. Ortiz, W.-J. van den Heuvel, J. Soldani, M. Villari, G. Casale, & P. Plebani (Eds.), *Communications in Computer and Information Science. Advances in Service-Oriented and Cloud Computing* (Vol. 1360, pp. 53–64). Springer International Publishing. https://doi.org/10.1007/978-3-030-71906-7_5

Passath, T., Huber, C., Kohl, L., Biedermann, H., & Ansari, F. (2021). A Knowledge-Based Digital Lifecycle-Oriented Asset Optimisation. *Tehnički Glasnik*, *15*(2), 226–334. https://doi.org/10.31803/tg-20210504111400

Paulheim, H. (2016). Knowledge graph refinement: A survey of approaches and evaluation methods. *Semantic Web*, *8*(3), 489–508. https://doi.org/10.3233/SW-160218

Pawellek, G. (2013). Planung der Instandhaltung. In G. Pawellek (Ed.), *Integrierte Instandhaltung und Ersatzteillogistik* (pp. 285–324). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-31383-7_7

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems*, *24*(3), 45–77. https://doi.org/10.2753/MIS0742-1222240302

Pennington, J., Socher, R., & Manning, C. (2014). Glove: Global Vectors for Word Representation. In Q. C. R. I. Alessandro Moschitti, G. Bo Pang, & U. o. A. Walter Daelemans (Eds.), *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 1532–1543). Association for Computational Linguistics. https://doi.org/10.3115/v1/D14-1162

Peres, R. S., Jia, X., Lee, J., Sun, K., Colombo, A. W., & Barata, J. (2020). Industrial Artificial Intelligence in Industry 4.0 - Systematic Review, Challenges and Outlook. *IEEE Access*, *8*, 220121–220139. https://doi.org/10.1109/ACCESS.2020.3042874

Peters, M. E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K [Kenton], & Zettlemoyer, L. (2018, February 15). *Deep contextualized word representations*. http://arxiv.org/pdf/1802.05365v2

Pipino, L. L., Lee, Y. W., & Wang, R. Y. (2002). Data quality assessment. *Communications of the ACM*, *45*(4), 211–218. https://doi.org/10.1145/505248.506010

Pivoto, D. G., Almeida, L. F. de, Da Rosa Righi, R., Rodrigues, J. J., Lugli, A. B., & Alberti, A. M. (2021). Cyber-physical systems architectures for industrial internet of things applications in Industry 4.0: A literature review. *Journal of Manufacturing Systems*, *58*, 176–192. https://doi.org/10.1016/j.jmsy.2020.11.017

Pohl, K., & Rupp, C. (2021). *BASISWISSEN REQUIREMENTS ENGINEERING;AUS-UNDWEITERBILDUNG NACH IREB-STANDARD ZUM CERTIFIED PROFESSIONAL FOR REQUIREMENTS ENGINEERING*. DPUNKT VERLAG.

Raffel, C., Shazeer, N., Roberts, A., Lee, K [Katherine], Narang, S., Matena, M., Zhou, Y., Li, W., & Liu, P. J. (2019, October 23). *Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer*. http://arxiv.org/pdf/1910.10683v4

Ragab, A., Ouali, M.-S., Yacout, S., & Osman, H. (2016). Remaining useful life prediction using prognostic methodology based on logical analysis of data and Kaplan–Meier estimation. *Journal of Intelligent Manufacturing*, *27*(5), 943–958. https://doi.org/10.1007/s10845-014-0926-3

Ranz, F., Komenda, T., Reisinger, G [Gerhard], Hold, P., Hummel, V., & Sihn, W [Wilfried] (2018). A Morphology of Human Robot Collaboration Systems for Industrial Assembly. *Procedia CIRP*, *72*, 99–104. https://doi.org/10.1016/j.procir.2018.03.011

Reinhold, J., Frank, M., Koldewey, C., Dumitrescu, R., & Gausemeier, J. (2019). Competence-based Planning of Value Networks for Smart Services. In

Rivas, A., Fraile, J. M., Chamoso, P., González-Briones, A., Sittón, I., & Corchado, J. M. (2020). A Predictive Maintenance Model Using Recurrent Neural Networks. In F. Martínez Álvarez, A. Troncoso Lora, J. A. Sáez Muñoz, H. Quintián, & E. Corchado (Eds.), *Advances in Intelligent Systems and Computing. 14th International Conference on Soft Computing Models in Industrial and Environmental Applications (SOCO 2019)* (Vol. 950, pp. 261–270). Springer International Publishing. https://doi.org/10.1007/978-3-030-20055-8_25

Rogers, A., & Augenstein, I. (2020, October 8). *What Can We Do to Improve Peer Review in NLP?* http://arxiv.org/pdf/2010.03863v1

Rosati, R., Romeo, L., Cecchini, G., Tonetto, F., Viti, P., Mancini, A., & Frontoni, E. (2023). From knowledge-based to big data analytic model: a novel IoT and machine learning based decision support system for predictive maintenance in Industry 4.0. *Journal of Intelligent Manufacturing*, *34*(1), 107–121. https://doi.org/10.1007/s10845-022-01960-x

Runji, J. M., Lee, Y.-J., & Chu, C.-H. (2023). Systematic Literature Review on Augmented Reality-Based Maintenance Applications in Manufacturing Centered on Operator Needs. *International Journal of Precision Engineering and Manufacturing-Green Technology*, *10*(2), 567–585. https://doi.org/10.1007/s40684-022-00444-w

Rusnock, C. F., & Borghetti, B. J. (2018). Workload profiles: A continuous measure of mental workload. *International Journal of Industrial Ergonomics*, *63*, 49–64. https://doi.org/10.1016/j.ergon.2016.09.003

Russell, S. J., Norvig, P., & Davis, E. (2010). *Artificial intelligence: A modern approach* (3rd ed.). *Prentice Hall series in artificial intelligence*. Prentice Hall.

Saini, K., Dhami, S. S., & Vanraj (2022). Predictive Monitoring of Incipient Faults in Rotating Machinery: A Systematic Review from Data Acquisition to Artificial Intelligence. *Archives of Computational Methods in Engineering*, *29*(6), 4005–4026. https://doi.org/10.1007/s11831-022-09727-6

Samtleben, S., & Rose, D. (2019). Die kleinen Helfer in der Produktion – ein Assistenzsystem wird konzipiert. In J. von Garrel (Ed.), *Digitalisierung der Produktionsarbeit* (pp. 197–214). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-27703-1_12

Scaife, A. D. (2024). Improve predictive maintenance through the application of artificial intelligence: A systematic review. *Results in Engineering*, *21*, 101645. https://doi.org/10.1016/j.rineng.2023.101645

Schank, R. C. (1972). Conceptual dependency: A theory of natural language understanding. *Cognitive Psychology*, *3*(4), 552–631. https://doi.org/10.1016/0010-0285(72)90022-9

Schlund, S., & Ansari, F. (2021). Fit for Future Work (Kompetenzen). In W. H. Güttel (Ed.), *AUSTRIAN MANAGEMENT REVIEW: v.8. AUSTRIAN MANAGEMENT REVIEW, Volume 8* (pp. 96–102). Rainer Hampp Verlag. https://doi.org/10.5771/9783957104014-96

Schlund, S., Mayrhofer, W., & Rupprecht, P. (2018). Möglichkeiten der Gestaltung individualisierbarer Montagearbeitsplätze vor dem Hintergrund aktueller technologischer Entwicklungen. *Zeitschrift Für Arbeitswissenschaft*, *72*(4), 276–286. https://doi.org/10.1007/s41449-018-0128-5

Schuh, G [Günther], Gartzen, T., & Wagner, J. (2015). Complexity-oriented ramp-up of assembly systems. *CIRP Journal of Manufacturing Science and Technology*, *10*, 1–15. https://doi.org/10.1016/j.cirpj.2015.05.007

Schulte, B., Fast, H., Flatt, H., Kleinhans, C., & Schulte, R. (2023). Low-Threshold Retrofit Strategy for CNC Machines: A New Process Data Acquisition Approach. In *2023 IEEE 21st International Conference on Industrial Informatics (INDIN)* (pp. 1–7). IEEE. https://doi.org/10.1109/INDIN51400.2023.10218263

Sexton, T., Brundage, M. P., Hoffman, M., & Morris, K. C. (2017). Hybrid datafication of maintenance logs from AI-assisted human tags. In *2017 IEEE International Conference on Big Data (Big Data)* (pp. 1769–1777). IEEE. https://doi.org/10.1109/BigData.2017.8258120

Sezer, E., Romero, D., Guedea, F., Macchi, M., & Emmanouilidis, C. (2018). An Industry 4.0-Enabled Low Cost Predictive Maintenance Approach for SMEs. In *2018 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC)* (pp. 1–8). IEEE. https://doi.org/10.1109/ICE.2018.8436307

Shaikh, A., Harreis, H., Machado, J., & Rowshankish, K. (2024). *Master data management: The key to getting more from your data.*

Shamsuzzoha, A., Toscano, C., Carneiro, L. M., Kumar, V [V.], & Helo, P. (2016). ICT-based solution approach for collaborative delivery of customised products. *Production Planning & Control*, *27*(4), 280–298. https://doi.org/10.1080/09537287.2015.1123322

Shcherbakov, M., & Sai, C. (2022). A Hybrid Deep Learning Framework for Intelligent Predictive Maintenance of Cyber-physical Systems. *ACM Transactions on Cyber-Physical Systems*, *6*(2), 1–22. https://doi.org/10.1145/3486252

Shen, C [Changqing], He, Q., Kong, F., & Tse, P. W. (2013). A fast and adaptive varying-scale morphological analysis method for rolling element bearing fault diagnosis. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, *227*(6), 1362–1370. https://doi.org/10.1177/0954406212460628

Shin, W., Han, J [Jeongyun], & Rhee, W. (2021). AI-assistance for predictive maintenance of renewable energy systems. *Energy*, *221*, 119775. https://doi.org/10.1016/j.energy.2021.119775

Siraskar, R., Kumar, S., Patil, S., Bongale, A., & Kotecha, K. (2023). Reinforcement learning for predictive maintenance: a systematic technical review. *Artificial Intelligence Review*, *56*(11), 12885–12947. https://doi.org/10.1007/s10462-023-10468-6

Stanford NLP. (2013). *The Stanford Natural Language Processing Group -- nlp.stanford.edu*.

Stock-Williams, C., & Swamy, S. K. (2019). Automated daily maintenance planning for offshore wind farms. *Renewable Energy*, *133*, 1393–1403. https://doi.org/10.1016/j.renene.2018.08.112

Strauss, P., Schmitz, M., Wostmann, R., & Deuse, J. (2018). Enabling of Predictive Maintenance in the Brownfield through Low-Cost Sensors, an IIoT-Architecture and Machine Learning. In *2018 IEEE International Conference on Big Data (Big Data)* (pp. 1474–1483). IEEE. https://doi.org/10.1109/BigData.2018.8622076

Taleb, I., Serhani, M. A., & Dssouli, R. (2018). Big Data Quality Assessment Model for Unstructured Data. In *2018 International Conference on Innovations in Information Technology (IIT)* (pp. 69–74). IEEE. https://doi.org/10.1109/INNOVATIONS.2018.8605945

Tang, W., Chen, Y [Yuejian], & Zuo, M. J. (2020). Health Index Development for a Planetary Gearbox. *Procedia Manufacturing*, *49*, 155–159. https://doi.org/10.1016/j.promfg.2020.07.012

Teern, A., Kelanti, M., Päivärinta, T., & Karaila, M. (2022). Knowledge graph construction and maintenance process: Design challenges for industrial maintenance support. In *BIR 2022 Workshops and Doctoral Consortium: joint proceedings of the BIR 2022 Workshops and Doctoral Consortium co-located with 21st International Conference on Perspectives in Business Informatics Research (BIR 2022) Rostock, Germany, September 20-23, 2022.* Symposium conducted at the meeting of RWTH Aachen University.

Teoh, Y. K., Gill, S. S., & Parlikad, A. K. (2023). IoT and Fog-Computing-Based Predictive Maintenance Model for Effective Asset Management in Industry 4.0 Using Machine Learning. *IEEE Internet of Things Journal*, *10*(3), 2087–2094. https://doi.org/10.1109/JIOT.2021.3050441

Terrissa, L. S., Meraghni, S., Bouzidi, Z., & Zerhouni, N. (2016). A new approach of PHM as a service in cloud computing. In *2016 4th IEEE International Colloquium on Information Science and Technology (CiSt)* (pp. 610–614). IEEE. https://doi.org/10.1109/CIST.2016.7804958

Thalhammer, A., & Rettinger, A. (2016). PageRank on Wikipedia: Towards General Importance Scores for Entities. In H. Sack, G. Rizzo, N. Steinmetz, D. Mladenić, S. Auer, & C. Lange (Eds.), *Lecture Notes in Computer Science. The Semantic Web* (Vol. 9989, pp. 227–240). Springer International Publishing. https://doi.org/10.1007/978-3-319-47602-5_41

Thomas, D., & Weiss, B. (2021). Maintenance Costs and Advanced Maintenance Techniques in Manufacturing Machinery: Survey and Analysis. *International Journal of Prognostics and Health Management*, *12*(1). https://doi.org/10.36001/ijphm.2021.v12i1.2883

Tolio, T. A. M., Monostori, L [László], Váncza, J., & Sauer, O [Olaf] (2023). Platform-based manufacturing. *CIRP Annals*, *72*(2), 697–723. https://doi.org/10.1016/j.cirp.2023.04.091

Toyoglu, H., Lin, A., Knapp, J., van Wyck, J., Rose, S., & Pentecoste, A. (2023). *Charting AI's Successful Course in Predictive Maintenance.*

Traini, E., Bruno, G., D'Antonio, G., & Lombardi, F. (2019). Machine Learning Framework for Predictive Maintenance in Milling. *IFAC-PapersOnLine*, *52*(13), 177–182. https://doi.org/10.1016/j.ifacol.2019.11.172

Treviño-Elizondo, B. L., & García-Reyes, H. (2023). What does Industry 4.0 mean to Industrial Engineering Education? *Procedia Computer Science*, *217*, 876–885. https://doi.org/10.1016/j.procs.2022.12.284

Tunstall, L., Werra, L. v., & Wolf, T. (2022). *Natural Language Processing with Transformers, Revised Edition*. O'Reilly Media, Inc.

U.S. Department of Labor. (2024). *The O*NET® Content Model at O*NET Resource Center -- onetcenter.org*.

Uhde, A., Schlicker, N., Wallach, D. P., & Hassenzahl, M. (2020). Fairness and Decision-making in Collaborative Shift Scheduling Systems. In R. Bernhaupt, F. '. Mueller, D. Verweij, J. Andres, J. McGrenere, A. Cockburn, I. Avellino, A. Goguey, P. Bjørn, S. Zhao, B. P. Samson, & R. Kocielnik (Eds.), *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1–13). ACM. https://doi.org/10.1145/3313831.3376656

Vajjala, S., Majumder, B., Gupta, A., & Surana, H. (2020). *Practical Natural Language Processing*. O'Reilly Media, Inc.

van Dinter, R., Tekinerdogan, B., & Catal, C. (2022). Predictive maintenance using digital twins: A systematic literature review. *Information and Software Technology*, *151*, 107008. https://doi.org/10.1016/j.infsof.2022.107008

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). *Attention Is All You Need.* https://doi.org/10.48550/arXiv.1706.03762

VDI (2014). *Operation Management and Process Monitoring in Power Generation and Utilities* (VDI Richtlinie). Verein Deutscher Ingenieure (VDI).

Vernon, D. (2021). Cognitive System. In K. Ikeuchi (Ed.), *Computer Vision* (pp. 157–164). Springer International Publishing. https://doi.org/10.1007/978-3-030-63416-2_82

Verspoor, K., & Cohen, K. B. (2013). Natural Language Processing. In W. Dubitzky, O. Wolkenhauer, K.-H. Cho, & H. Yokota (Eds.), *Encyclopedia of Systems Biology* (pp. 1495–1498). Springer New York. https://doi.org/10.1007/978-1-4419-9863-7_158

Wan, S., Li, D., Gao, J., & Li, J [Jing] (2019). A knowledge based machine tool maintenance planning system using case-based reasoning techniques. *Robotics and Computer-Integrated Manufacturing*, *58*, 80–96. https://doi.org/10.1016/j.rcim.2019.01.012

Wang, J., Yang, Z., Hu, X [Xiaowei], Li, L [Linjie], Lin, K., Gan, Z., Liu, Z [Zicheng], Liu, C [Ce], & Wang, L [Lijuan]. (2022). *GIT: A Generative Image-to-text Transformer for Vision and Language.* https://doi.org/10.48550/arXiv.2205.14100

Wang, X [Xiang], He, X., Cao, Y., Liu, M [Meng], & Chua, T.-S. (2019). KGAT. In A. Teredesai, V. Kumar, Y. Li, R. Rosales, E. Terzi, & G. Karypis (Eds.), *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (pp. 950–958). ACM. https://doi.org/10.1145/3292500.3330989

Wang, X [Xiaoqiao], Liu, M [Mingzhou], Liu, C [Conghu], Ling, L., & Zhang, X [Xi] (2023). Data-driven and Knowledge-based predictive maintenance method for industrial robots for the production stability of intelligent manufacturing. *Expert Systems with Applications*, *234*, 121136. https://doi.org/10.1016/j.eswa.2023.121136

Wang, X [Xiong], Ferreira, F. A., & Chang, C.-T. (2022). Multi-objective competency-based approach to project scheduling and staff assignment: Case study of an internal audit project. *Socio-Economic Planning Sciences*, *81*, 101182. https://doi.org/10.1016/j.seps.2021.101182

Wellsandt, S., Klein, K., Hribernik, K., Lewandowski, M., Bousdekis, A., Mentzas, G., & Thoben, K.-D. (2022). Hybrid-augmented intelligence in predictive maintenance with digital intelligent assistants. *Annual Reviews in Control*, *53*, 382–390. https://doi.org/10.1016/j.arcontrol.2022.04.001

Wiendahl, H.-H., Cieminski, G. von [G.], & Wiendahl, H.-P. (2005). Stumbling blocks of PPC: Towards the holistic configuration of PPC systems. *Production Planning & Control*, *16*(7), 634–651. https://doi.org/10.1080/09537280500249280

Wireman, T. (2014). *The Maintenance Strategy Series - 6 Volume Set*. Reliabilityweb.com. https://books.google.at/books?id=eVfNsgEACAAJ

Wong, P. K., Zhong, J.-H., Yang, Z.-X., & Vong, C. M. (2017). A new framework for intelligent simultaneous-fault diagnosis of rotating machinery using pairwise-coupled sparse Bayesian extreme learning committee machine. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, *231*(6), 1146–1161. https://doi.org/10.1177/0954406216632022

Woods, C., French, T., Hodkiewicz, M., & Bikaun, T. (2023). An ontology for maintenance procedure documentation. *Applied Ontology*, *18*(2), 169–206. https://doi.org/10.3233/AO-230279

Wu, D., Jennings, C., Terpenny, J., & Kumara, S [Soundar] (2016). Cloud-based machine learning for predictive analytics: Tool wear prediction in milling. In *2016 IEEE International Conference on Big Data (Big Data)* (pp. 2062–2069). IEEE. https://doi.org/10.1109/BigData.2016.7840831

Wu, D., Jennings, C., Terpenny, J., Kumara, S [Soundar], & Gao, R. X. (2018). Cloud-Based Parallel Machine Learning for Tool Wear Prediction. *Journal of Manufacturing Science and Engineering*, *140*(4), Article 041005. https://doi.org/10.1115/1.4038002

Xia, L., Zheng, P., Li, X., Gao, R., & Wang, L [Lihui] (2022). Toward cognitive predictive maintenance: A survey of graph-based approaches. *Journal of Manufacturing Systems*, *64*, 107–120. https://doi.org/10.1016/j.jmsy.2022.06.002

Xie, J [Jiawei], Huang, J [Jinsong], Zeng, C., Jiang, S.-H., & Podlich, N. (2020). Systematic Literature Review on Data-Driven Models for Predictive Maintenance of Railway Track: Implications in Geotechnical Engineering. *Geosciences*, *10*(11), 425. https://doi.org/10.3390/geosciences10110425

Xie, J [Jieyu], & Wan, J. (2023). Digital Twin Four-Dimension Fusion Modeling Method Design and Application to the Discrete Manufacturing Line. *Big Data and Cognitive Computing*, *7*(2), 89. https://doi.org/10.3390/bdcc7020089

Xie, T., Xu, Q., Jiang, C., Lu, S., & Wang, X [Xiangxiang] (2023). The fault frequency priors fusion deep learning framework with application to fault diagnosis of offshore wind turbines. *Renewable Energy*, *202*, 143–153. https://doi.org/10.1016/j.renene.2022.11.064

Xiong, J., Fink, O., Zhou, J., & Ma, Y. (2023). Controlled physics-informed data generation for deep learning-based remaining useful life prediction under unseen operation conditions. *Mechanical Systems and Signal Processing*, *197*(12), Article 110359, 110359. https://doi.org/10.1016/j.ymssp.2023.110359

Yan, J., Meng, Y., Lu, L., & Li, L [Lin] (2017). Industrial Big Data in an Industry 4.0 Environment: Challenges, Schemes, and Applications for Predictive Maintenance. *IEEE Access*, *5*, 23484–23491. https://doi.org/10.1109/ACCESS.2017.2765544

Yeung, K. (2020). Recommendation of the council on artificial intelligence (OECD). *International Legal Materials*, *59*(1), 27–34.

Yu, W [Wenjin], Dillon, T., Mostafa, F., Rahayu, W., & Liu, Y [Yuehua] (2020). A Global Manufacturing Big Data Ecosystem for Fault Detection in Predictive Maintenance. *IEEE Transactions on Industrial Informatics*, *16*(1), 183–192. https://doi.org/10.1109/tii.2019.2915846

Zäh, M., Vogl, W., Lau, C., Wiesbeck, M., & Ostgathe, M. (2007). Towards the Cognitive Factory. *2nd International Conference on Changeable, Agile, Reconfigurable and Virtual Production*.

Zenisek, J., Holzinger, F., & Affenzeller, M. (2019). Machine learning based concept drift detection for predictive maintenance. *Computers & Industrial Engineering*, *137*, 106031. https://doi.org/10.1016/j.cie.2019.106031

Zhai, S., Gehring, B., & Reinhart, G [Gunther] (2021). Enabling predictive maintenance integrated production scheduling by operation-specific health prognostics with generative deep learning. *Journal of Manufacturing Systems*, *61*, 830–855. https://doi.org/10.1016/j.jmsy.2021.02.006

Zhang, D., Stewart, E., Entezami, M., Roberts, C., & Yu, D. (2020). Intelligent acoustic-based fault diagnosis of roller bearings using a deep graph convolutional network. *Measurement*, *156*, 107585. https://doi.org/10.1016/j.measurement.2020.107585

Zhang, W., Yang, D., & Wang, H. (2019). Data-Driven Methods for Predictive Maintenance of Industrial Equipment: A Survey. *IEEE Systems Journal*, *13*(3), 2213–2227. https://doi.org/10.1109/JSYST.2019.2905565

Zhang, Z [Ziqiang], Zhou, L., Ao, J., Liu, S., Dai, L., Li, J [Jinyu], & Wei, F. (2022). *SpeechUT: Bridging Speech and Text with Hidden-Unit for Encoder-Decoder Based Speech-Text Pre-training*. https://doi.org/10.48550/arXiv.2210.03730

Zheng, X., Hu, X [Xiaodu], Arista, R., Lu, J., Sorvari, J., Lentes, J., Ubis, F., & Kiritsis, D. (2024). A semantic-driven tradespace framework to accelerate aircraft manufacturing system design. *Journal of Intelligent Manufacturing*, *35*(1), 175–198. https://doi.org/10.1007/s10845-022-02043-7

Zhou, C., Li, Q [Qian], Li, C [Chen], Yu, J., Liu, Y [Yixin], Wang, G., Zhang, K [Kai], Ji, C., Yan, Q., He, L., Peng, H., Li, J [Jianxin], Wu, J [Jia], Liu, Z [Ziwei], Xie, P., Xiong, C., Pei, J., Yu, P. S., & Sun, L. (2023, February 18). *A Comprehensive Survey on Pretrained Foundation Models: A History from BERT to ChatGPT*. http://arxiv.org/pdf/2302.09419v3

Zhou, Z., Li, F., Zhu, H., Xie, H., Abawajy, J. H., & Chowdhury, M. U. (2020). An improved genetic algorithm using greedy strategy toward task scheduling optimization in cloud environments. *Neural Computing and Applications*, *32*(6), 1531–1541. https://doi.org/10.1007/s00521-019-04119-7

Zonta, T., Da Costa, C. A., Da Rosa Righi, R., Lima, M. J. de, Da Trindade, E. S., & Li, G. P. (2020). Predictive maintenance in the Industry 4.0: A systematic literature review. *Computers & Industrial Engineering*, *150*, 106889.

11 Appendix II – Curriculum Vitae

Ing. Dipl.-Ing.

Linus Kohl, BSc

Education

Studies	Duration
Doctoral studies *TU Wien Mechanical Engineering-Management*	2019-now
Master of Science *TU Wien, Mechanical Engineering-Management*	2018-2019
Bachelor of Science *TU Wien, Mechanical Engineering-Management*	2011-2018

Experience

Position	Duration
Director of Digitalization *Voestalpine Krems GmbH*	2025-now
Head of Production Optimization and Maintenance Management *Fraunhofer Austria Research GmbH*	2023-2024
Researcher & Project Manager *Fraunhofer Austria Research GmbH*	2020-2023
Research Associate *TU Wien, Research Group of Production and Maintenance Management*	2019-2023
Big Data Engineer *TU Wien, TU.it*	2018-2019