DAI-DSS RESEARCH SPECIFICATION

D3.1

Editor Name	Sylwia Olbrych (RWTH Aachen)
Submission Date	June 21, 2023
Version	1.0
State	FINAL
Confidentially Level	PU

Co-funded by the Horizon Europe
Framework Programme of the European Union

EXECUTIVE SUMMARY

This report focuses on the deliverable "**D3.1 – DAI-DSS Research Specification**", part of the Horizon Europe project FAIRWork. The deliverable aims to describe the specific research factors in selected use cases of industrial partners FLEX and CRF. It presents a research strategies and factors catalogue that serves as a framework for conducting research within the Democratized AI-based Decision Support System (DAI-DSS). DAI-DSS research specifications are closely related to deliverables "D2.1 Specification of FAIRWork Use Case and DAI-DSS Prototype Report" and "D4.1 DAI-DSS Architecture and Initial Documentation and Test Report".

The first part of the report provides an **overview of the relevant literature** related to the research intended within the frame of this project. It covers the most significant research domains, such as the democratization of decision-making and digital shadows and twins for human experts. Additionally, it explores technical approaches like Artificial Intelligence (AI) and Multi-Agent System (MAS) crucial for improving Decision Support Systems (DSS). This section also presents the state-of-the-art crucial aspects of today's technology, particularly reliability and trustworthiness in AI. The output of this literature review leads to research questions in multiple domains addressed within the FAIRWork project.

The second part of the report focuses on the **research methodologies and strategies** employed to investigate the technical aspects of decision-making processes, human aspects, and digital human factors measurements. It presents research approaches for successfully implementing AI and MAS-based technologies into DSS. Methods such as data-driven modelling, prototyping, and testing are proposed within the AI and MAS domains. Additionally, the report outlines the use of sensors to capture critical information about humans' mental, affective, and motivational states, including implementation details of the Intelligent Sensor Box (ISB). Furthermore, a novel framework using Personas as Human Digital Twins for Decision Making in the context of Industry 5.0 is described.

The final part of the report presents the **key research factors** identified in the industrial use cases and potential AI services to address them. These research factors are categorized into two main perspectives: the human perspective and the technical perspective. The human perspective factors are derived from the research plan and are observed in given use cases. On the technical side, the requirements for modelling and testing new concepts using AI and MAS technologies primarily focus on data availability (process-relevant and expert knowledge) and the DSS architecture necessary to enable information flow and decision models related to the use cases.

The report also provides a strategy for **communication and dissemination** in the context of the research methodology of the FAIRWork project. The objective is to continuously disseminate project achievements, raise awareness about the project, and gather feedback to improve the created research artefacts.

PROJECT CONTEXT

Workpackage	WP3: Research on Method and Tools for DAI-DSS
Task	T3.1: Research on Democratization of Decision-Making using Multi Agent Systems T3.2: Research on Digital Shadows and Twins for Human Experts and Data Driven Algorithms T3.3: Research on AI-Based Decision-Making for AI, Robots and Human Experts T3.4: Research on Reliable and Trustworthy AI
Dependencies	WP2, WP4, WP5

Contributors and Reviewers

Contributors	Reviewers
Herwig Zeiner, Lucas Paletta, Michael Schneeberger (JRS) Gustavo Vieira (MORE) Sylwia Olbrych, Alexander Nasuta, Johanna Werz, Noushin Gheibi, Stefan Böschen (RWTH Aachen) Magdalena Dienstl, Marlene Mayer (BOC) Christian Muck (OMILAB)	Wilfrid Utz (OMILAB) Marco Kemmerling (RWTH Aachen) Werner Bailer (JRS)

Approved by: Robert Woitsch [BOC], as FAIRWork coordinator

Version History

Version	Date	Authors	Sections Affected
1.0	June 20, 2023	Sylwia Olbrych et al.	All

Copyright Statement – Restricted Content

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

1.1 Purpose of the Document

The goal of this document is to map systemic and individual research factors in selected use cases presented in deliverable "D2.1 Specification of FAIRWork Use Case and DAI-DSS Prototype Report" to confirm the appropriateness of Multi-Agent System (MAS) and Artificial Intelligence (AI) in Decision Support Systems (DSS). The purpose of gathering these research specifications is to provide a precise depiction of the requirements necessary for the investigation of the technical factors inherent in the given use cases, including the application of AI and MAS in a Democratic AI-based Decision Support System (DAI-DSS), as well as the examination of human aspects, such as the reliability and trustworthiness of AI.

To receive deeper insight into the project's scope and DAI – DSS architecture, please refer to deliverable "D2.1 Specification of FAIRWork Use Case and DAI-DSS Prototype Report" and deliverable "D4.1 DAI-DSS Architecture and Initial Documentation and Test Report".

1.2 Document Structure

The document is structured as follows. The next section presents an overview of the literature relevant to the research conducted in this project. The literature review includes several blocks that discuss topics such as the democratization of decision-making using MAS, digital shadows and twins for human experts, AI-enriched DSS, and reliable and trustworthy AI.

The research methodologies section outlines the research plan as well as the research methodology that the authors employed to investigate the human and technical factors in decision-making and the use of AI and MAS to enhance it.

The report then moves on to discuss the key research factors in FLEX and CRF use cases. The authors outline the different decision support scenarios, including resource mapping, solution configuration, and selection, and discuss how AI and MAS can be used to enhance decision-making in each of these scenarios. The section also discusses the human perspective in use cases and how it can be incorporated into decision-making processes.

Following, the authors present how they plan to communicate and disseminate their research findings.

Finally, the report concludes with a summary section, where the authors summarize the key points debated in the report and emphasize the importance of incorporating the human perspective into decision-making processes and the need for reliable and trustworthy AI.

2 OVERVIEW ON RELEVANT LITERATURE

The following section provides an overview of the literature related to the research conducted within the frame of this project and the proposed research questions. This report specifically focuses on four research directions:

- democratization of decision-making using Multi-Agent Systems,
- digital shadows and twins for human experts and data-driven algorithms,
- AI-enriched Decision Support Systems,
- reliable and trustworthy AI.

Analysing literature provides valuable insights into the latest AI and MAS advancements in a manufacturing environment and highlights best practices and current trends. Furthermore, this review also emphasizes how these technologies impact decision democratization and human trust and stress levels while applying them. The acquired knowledge allows identifying gaps in the current research, which can be addressed through further research efforts.

2.1 Democratization of Decision-Making Using Multi-Agent Systems

The question of democratization of decision-making contains several aspects and therefore has to be split into two parts. As a first step, the term and task of democratization have to be integrated into a larger picture. While the term *democracy* stems from a political and sociological domain, the sociology of technology approaches provide the opportunity to transfer the term into socio-technical settings. This transfer consists of explaining the political quality of the settings consisting of political ability, political will and political ought. It also consists of the notion that the quality of democracy consists of a legitimation process that needs to be examined in these socio-technical settings. Finally, “doing democracy” as a concept of democratic formation of socio-technical arrangements provides the ground for human-centred research in FAIRWork. These sociological considerations are explained in more detail in the following Section 2.1.1.

As a next step in Section 2.1.2, the Multi-Agent Systems (MAS) term is introduced. Stemming from Computer Science and Artificial Intelligence, they provide the opportunity to bring together agents that interact towards a common goal. While from a technological viewpoint, MAS is characterized by autonomy, distribution and negotiation; democratical concepts have not yet been applied to these systems. This way, the section investigates the suitability, challenges, gains, and initial approaches to applying democratic notions to MAS.

2.1.1 Democratization of Decision-Making in Socio-Technical Settings

In view of the various stakeholders - people and technologies - who are relevant to decision-making and therefore need to be represented, FAIRWork has set itself the task of developing a democratic tool in which all interests are taken into account and balanced as well as possible. On the one hand, this results in the challenge of granting people autonomy and the final decision. This requirement results not least from ethical guidelines according to which machines cannot be accountable and must not be above the decision autonomy of humans (Felzmann, Villaronga, Lutz, & Tamo-Larrieux, 2019[1]; A. J. B.-B. Hleg, 2019[2]). At the same time, the demand for democratic processes has multiple dimensions, as Lincoln put it in his famous Gettysburg Address: "the government of the people, by the people, and for the people". Issues of one's own concern demand one's own voice and appropriately established procedures for making it heard. The legitimacy of a procedure as democratic is measured according to whether and how procedures can bring the expectations and claim positions of those subject to the decision to bear - according to the legitimacy of one's own concerns (Schmidt, 2013[3]). Concerning the questions of democratic theory, two main strands of research are relevant to the FAIRWork project. The first is research on questions of discursive governance, in which discourses of the political and their respective qualities are distinguished (Böschen & Sigwart, 2020[4]). The second is research that focuses on the mobilization of civil society actors and citizens, addressing questions of institutionalization (Sørensen & Torfing, 2007[5]).

The political quality of discourses can be broken down into three dimensions. The decisive factor here is; In addition to questions of competent governance based on (technical) expertise, the fundamental democratic concerns of collective decision-making and moral self-understanding are always negotiated at the same time. These primary concerns correspond to three different semantics of public speaking. Using such a heuristic can also grasp challenges relevant to technology development. Firstly, democratic discourses have an implementation function, according to which democratic politics is primarily about enabling efficient political decision-making and overall societal control, i.e. generating and successfully applying effective forms of governance as a means of realising the political self-determination of society (*political ability*). The semantics of *political ability* ultimately focuses on real problems, which may always include norms. However, these are addressed primarily in the sense of coordination rules in processes of problem-solving (Trein, Thomann, & Maggetti, 2019[6]). Normatively, the semantics of *political ability* is therefore inscribed with an orientation towards the "values" of neutrality and objectivity in the context of "good governance". It also contains the ability to determine and understand those objective problems that are to be taken into account independently of the normative targets of collective will formation and moral convictions in the sense of competent and responsible problem-solving.

The understanding of such normative specifications is an essential component of democratic discourses. Therefore, secondly, they represent the attempt to enable a process of collective will formation through exchange and conflict, but also the bringing together of different opinions, interests and ideas about the goals of political shaping (*political will*). The semantics of *political* normatively express the basic democratic idea of collective self-determination and, with it, an understanding of power according to which political power is to be understood not only in the sense of "constituted" but also in the sense of "constituting power" and thus as a bottom-up phenomenon (Kalyvas, 2005[7]). It is true that the "collective body" or the public sphere as a whole always consists of a multitude of often contradictory positions, and processes of collective will formation take place erratically and conflictual. Despite - or precisely because of - this pluralistic-agonal form, the semantics of *political will* nevertheless represents a relevant element of democratic politics, which is articulated in concrete forms of "practical enactment in public life" (Wenman, 2013[8]; White & Ypi, 2017[9]) and has real effects on the agenda of democratic processes.

However, the democratic discourse on the normative primary coordinates of the political process is not exclusively about organizing collective will-forming processes. It is also – thirdly – about the attempt to generate categories of moral orientation and to articulate basic principles of public-political morality, which determine, for example, fundamental rights worthy of protection as indispensable prerequisites for the legitimate use of political power (*political ought*). This moral character of public discourses is expressed above all in the fact that the question of the fundamental restrictions to which political action should be subject is always negotiated in them. The semantics of political ought ultimately reflects the genuinely liberal idea that every political government, even one that invokes the democratic principle of popular sovereignty, i.e., the requirements of effective governance and the processes of collective will formation, require moral restraint. In contrast to the voluntaristic semantics of collective will formation, *political ought* is therefore articulated in the semantics of fundamental moral duties, to the fulfilment of which every democratic community should be normatively bound.

With regard to the quest for institutionalization, one sees in contemporary political theory acknowledging that there are a vast amount of practices of experimentally articulating and testing “representative claims” (Saward, 2006[10]). They are supposed to tell us how the people's will may become performative; that is, they bring a specific version of the will of the people (and thus a specific version of democracy) into existence (*The Constructivist Turn in Political Representation*, 2019[11]). At the same time, for their specific ways of reducing and translating the actual diversity of the people into a unitary representation of their collective will, all of them are selective, partial and biased (Brown, 2009[12]; Chilvers & Kearnes, 2015[13]; Gomart & Hajer, 2003[14]). It is acknowledged that the question cannot be to find the best way to neutrally represent what may be imagined as an independently existing, essential collective will of the people. The continuing quest for democracy must instead be concerned with the specific qualities of different practices of performatively representing the will of the people (Hayat, 2019[15]; Laurent, 2011[16]). As far as they are able to articulate representations which are accepted by those whom they comprise, such representational practices constitute the very collective subjectivity, the social group with a shared will, identity, value, need, and interest that they describe. Thus, there is not one true will of the people but many possible ones. The quality of democracy cannot be neutrally mirroring towards collective will that is already there. However, it must be in articulating possible collective wills in a way that they, in combination, reflect the potential collective concerns and collective positions on an issue that are latent in a historical situation (Chilvers, Pallett & Hargreaves, 2018[17]).

In selected technological domains, practices of "doing democracy" across several sites of research and innovation are reported (Felt & Fochler, 2010[18]; Konopásek, Soneryd, & Svačina, 2018[19]; Lezaun & Soneryd, 2007[20]; Soneryd & Amelung, 2016[21]; Voß, Schritt, & Sayman, 2022[22]). Thereby, metaphors, narratives, concepts, formats, methods, devices, embodied skills, and material tools for representing the subject as well as the collective are performed and constituting a space for configuring actors, discourses and institutions. These insights are decisive for the humans-centered research in FAIRWork. It is about the democratic formation of socio-technical arrangements. With Multi-Agent Systems, a new type of agent is coming into play, thereby re-ordering established socio-technical arrangements so far. Moreover, these systems structure the representation of humans in a highly specified way. Thus, the goal of social science research in FAIRWork must also be to establish the legitimacy of people in relation to machines and thus secure the democratic process on the part of the people involved.

Following this, there are specific research questions to explore the socio-technical structure of these interaction settings and their potential of "doing democracy":

- What does successful AI-enriched decision-making look like for human stakeholders?
- What is the influence of the socio-technical context in democratic decision-making?
- What are the boundary conditions for practising specific affordances related to AI – and what are the limits of such an approach?
- What are the dynamics and forms of constituting a "We" for participatory technology development with the ambition of democratic decision-making tools?
- How can we understand the practices of working and deciding at production-lines in order to identify the relevant factors for Multi-Agent Systems (MAS)?
- How can MAS be designed to ensure a democratic decision-making process in industry processes under consideration?
- How can MAS contribute to enhancing worker participation in decision-making in the industry?
- What are the conditions of legitimacy for establishing a MAS in the context of the industry for the purpose of algorithmic decision-making support of workforce-allocation?

2.1.2 Democratization in Multi-Agent Systems

The concept of Multi-Agent Systems comes from the field of Distributed Artificial Intelligence (DAI) (Bond & Gasser, 1988[23]). It can be defined as systems composed of agents abstracting the behaviour of actors in a socio-technical systems that interact towards a defined goal. They interact with each other aiming to cooperatively reach an objective that goes beyond the scope of a single part of the system. They coordinate their structure without the need for a central component in order to solve problems that require the ability to communicate between the parties involved.

These intelligent agents, capable of providing elaborate recommendations based on their capacity to embed intelligent algorithms, exhibit functional characteristics for overcoming challenges in the industry. Among these characteristics, some of the most interesting for industrial settings are:

- Autonomy and proactiveness: agents can decide on what action to take independently from other agents based on their perception of the environment to fulfil the system's objectives. They can act proactively to achieve a goal that depends on their deliberated action.
- Distribution: the way multi-agent systems are modelled offers advantages to the system being worked on, such as the scalability that comes from the modular nature of the agents and the easy reconfiguration of the system through a reorganization of its entities according to the needs of the system at any given time.
- Negotiation: agents can communicate between them and exchange data to achieve their goals aligned with the system's objectives. They can determine the best outcomes for the system when interacting with neighbour agents and negotiating with them.

These inherent characteristics of MAS make them desired for industrial scenarios where adaptation and reconfigurability are required. When modelled in the multi-agent perspective, the entities of an industrial system obtain a series of advantages regarding their dynamic in the system (Wooldridge, 2009a[24]). Being able to communicate with other entities and express their state to their peers makes it possible to achieve goals that do not depend on single entities. These entities represent a pivotal point without which the system cannot function. Designing a system in distributed manner means making the system more resilient, able to adapt to disruptions and to react faster. Not depending on a singular entity responsible for managing an entire system that could even choke its performance brings great value to the system. Following the value of these characteristics, Multi-Agent Systems provide significant advantages when applied to recommendation systems.

Among recommendation systems that use MAS, there are applications in the manufacturing and energy industry, among others. In (Dostatni et al., 2016[25]), the process of designing a device is modelled using MAS. The system gives recommendations on how a product should be designed to facilitate its recyclability. Different agents representing the product's features define what parameters can be changed and tuned to achieve the system's objective (improve recyclability). In the energy field, MAS is recognized to be used in various applications, such as increasing energy profit, securing monitoring substations and preventing failures (Xie & Liu, 2017[26]).

The employment of MAS in environments involving humans as part of the entities involved in the processes raises some discussions. Among them are the principles inherent in this type of system, taking into account social factors. Human beings have a range of social dynamics that may or may not be appropriate for all occasions. When interests are at stake, the chance of some conflict is always significant, especially when the system scales. Therefore, it is important to define certain principles that ensure fairness and proper conditions for interaction in the environment. One important principle is that of democracy. The democratization of decision-making processes marks a significant transition in how decisions are made in many settings where humans and machines interact. It represents a shift from centralized, hierarchical decision-making models to distributed models that involve the due participation of stakeholders to accomplish fairness.

The application of MAS in industrial decision-making scenarios, such as in manufacturing, can enhance decision-making efficiency, distributing sensing and control; inclusiveness, enabling representation of stakeholders involved in the process; promoting fairness, preventing dominance of any group of entities over another due to the lack of involvement of essential parties in the decision-making processes (Jong et al., 2008[27]); and transparency, the agents can provide a clear interface for human-machine interaction regarding their goals, actions, reasoning and even projections about future (Chen et al., 2018[28]). Through MAS, operators can actively participate in decision-making, expressing their concerns, opinions and ideas, bringing often neglected issues into the decision equation. This form of human participation not only enriches the process but also can lead to solutions that are more socially acceptable and sustainable. The ability of agents to operate autonomously within a predefined set of rules ensures that decisions are made more democratically or without undue influence. Furthermore, the traceability and visibility of agent interactions in MAS give transparency to the process, enabling stakeholders to understand how decisions are made and promoting accountability.

By leveraging MAS, decision-making parties can be disseminated across multiple systems in different relevant aspects, enabling a more inclusive and representative process that reflects the interests of various stakeholders. It facilitates a bottom-up approach, where decisions are made based on input from every relevant process participant, not only from what a supervisor can observe from their perspective. The collective intelligence gathered provides knowledge fundamental to implement a democratic system (Rădulescu et al., 2019[29]).

The personnel responsible for taking the decisions can act autocratically, not taking into consideration third-party interests. One of the factors that defines whether a decision is democratic is the verification that all entities involved in the activity in question participate in the decision-making process. Among these aspects of participation are the interests of the entity itself and the data that can be added to the process. The participation of each stakeholder is of utmost importance for a fairer and more adequate decision-making process for all involved. Neglecting the input that stakeholders can offer means limiting the system's overall view of the role and impact of these entities on the system (Zhang et al., 2008[30]). MAS facilitate an approach where decisions are not only made by managers and supervisors but are instead derived from the collective intelligence gathered from every relevant member. This participatory approach, supported by the implementation of MAS, can lead to more informed and balanced decisions. Potential benefits range from improved worker satisfaction and productivity to more effective decision-making.

The democratization of decision-making through MAS presents some challenges, however. The successful deployment of MAS requires careful consideration of some factors from the perspective of an agent-based approach. The design of the system, the distribution of decision-making power among agents, and the representation of stakeholders can all impact the democratic nature of the process. The design and implementation of MAS should be guided by democratic principles to ensure that it genuinely supports democratized decision-making.

In the democratization context, MAS offers a promising infrastructure for implementing democratic decision-making mechanisms. Its communication protocols are important in enabling the decentralization of information exchange. These systems, which consist of autonomous entities working collaboratively to achieve goals beyond individual parts of a whole structure, take into account precisely the general objective of the desired solution. They bring about new possibilities for collective decision-making, transforming traditional decision dynamics and creating opportunities for participatory engagement of parties usually left out (Krupitzer et al., 2020[31]). The representation of human individuals and machines as agents and their interactions is a factor that adds highly to active human participation in decision-making processes.

An interesting and fruitful application of MAS in decision-making occurs in task allocation processes. This type of activity requires well-timed management of current and upcoming tasks. Multi-agents are considerably efficient in allocating tasks thanks to the characteristics mentioned above. Task allocation is an optimization activity that assigns tasks to the best capable agents considering the current state of the agents, their capabilities and the overall system’s objective.

One of the challenges is handling information around the system so the agents are aware of their peers' intentions to distribute the tasks appropriately. The agent assigned as a job allocator is responsible for finding the most capable agent for the task. It follows a procedure where the job allocator asks for schedules for a capability list for the task. Following, one agent is selected to delegate the task and details of the task are sent to the agent. The scheduling problem presents a challenge that MAS can assess efficiently and offer desired outcomes considering deadline restrictions and each entity involved in the tasks (Kim & Matson, 2016[32]).

One of the factors that promote the effectiveness of employing MAS in democratizing decision-making in the industry is the introduction of learning methods to agents. Machine Learning techniques make Multi-Agent Systems capable of responding proactively to unexpected events and in an enhanced manner as agents are trained with new experiences over time. This continuous learning ability of agents adds to the system's capability to adapt and evolve, thereby improving the efficiency and robustness of decision-making processes over time. The iterative nature of these learning processes promotes a collective intelligence within the system, facilitating dynamic and inclusive decision-making, a relevant characteristic of a democratic system. Recent advances in AI technologies, notably collaborative robotics, enable new forms of cooperative work between humans and machines, extending beyond fully automated processes. MAS, as a branch of AI, comprise various interacting agents. These agents, which can be social (human) or artificial (machine), have diverse attributes, including differences in cognitive capabilities and knowledge about their environment. MAS can be employed to share these different levels of perception and cognition between their agents and consequently improve the overall performance of a production system through its integration in such aspects. The cooperation and collaboration among agents in MAS are vital for solutions in the industry field. The integration of these parties promotes democratic characteristics through decentralized decision-making (Nixdorf et al., 2022[33]).

Integrating Machine Learning (ML) and MAS has shown potential in managing complex and unpredictable issues, like in the Oil and Gas Industry. The mutable nature of this type of industry requires solutions that can effectively deal with adaptive and flexible scenarios featured in these environments, which MAS and ML can facilitate through learning and adjustment. One prominent ML approach in agent learning is Reinforcement Learning (RL) which can be applied to maintenance planning. Despite recent advancements in AI and deep reinforcement learning renewing interest in MAS, its industrial application remains limited due to factors like lack of standardization and technology maturity. ML models can be assigned to individual agents to improve prediction, prevention, and data handling, reinforcing the scalability feature of MAS (Hanga et al., 2019[34]). Enabling MAS and ML integration proves to be a step towards the democratization of decision-making.

Applying MAS with RL in smart manufacturing systems can dynamically respond to consumer-driven demand for personalized production. In such an environment characterized by high variability and uncertainty, the traditional mass production decision-making paradigm might be insufficient for addressing industry challenges related to supply chain bottlenecks and demand volatility. MAS and RL offer the ability to make autonomous decisions based on local information, adapt to changes, and provide novel reactions to uncertain situations. RL is utilized to create an intelligent system that learns from its environment and cooperates with other agents through negotiation and agreement, thus overcoming the limitations of rule-based negotiation protocols. This integration of MAS and RL allows for autonomous distributed decision-making, environment learning, and job prioritization in sequence-dependent environments. Agents embedded with RL algorithms can outperform conventional dispatching rules in scheduling for a personalized production system (Kim et al., 2020[35]). Consequently, MAS combined with RL provides a powerful tool to tackle the challenges of personalized manufacturing systems, paving the way for future research into how such algorithms can also provide added value when humans participate in the decision-making processes (Zhang et al., 2020[36]).

Following, the research questions are presented:

- How can MAS be designed to ensure a democratic decision-making process in the industry?
- How can MAS contribute to enhancing worker participation in decision-making in the industry?
- What are the necessary interaction mechanisms to achieve fair work balance using Multi-Agent Systems considering the human-in-the-loop?
- How to transmit outcome and decision confidence from MAS to humans?
- How to model the human information that enables the agent to concretely and definitely, behave on behalf of the human?

- How to develop the interactions between agents and humans?
- How to distribute and balance artificial intelligence algorithms in Multi-Agent Systems?

Answering the previous questions can push research in the Multi-Agent Systems domain applied to decision support systems and draw new perspectives on how principles such as fairness can be effectively introduced into production systems that have strong participation from both humans and machines.

2.2 Digital Shadows and Twins for Human Experts and Data-Driven Algorithms

Releasing workers from routine tasks and enabling them to focus on creative, value-adding activities within a sociotechnical system is critical to the successful implementation of Industry 4.0 (Kagermann et al., 2013[37]; Winkelhaus et al., 2021[38]). Using new technologies to assist workers and to adapt production processes to their needs is also the key vision of Industry 5.0 (Breque et al., 2021[39]; Dixson-Decl`eve et al., 2022[40]). To ensure that workplaces and processes match the requirements of workers, it is necessary to analyse how workers interact with their work environment.

The following sections first provide a sketch of the state-of-the-art on data-driven algorithms, sensor-based Digital Shadows of human activity, in particular, the mental processes of the worker, referring to wearable biosensors and eye tracking technologies in the production environments. In the sequel, the human factors in decision-making and the health-specific and economic impact are outlined. Finally, the state-of-the-art of the R&D on Human Digital Twins is presented with an outlook on the FAIRWork research in this direction.

2.2.1 Data-Driven Approaches

Production optimisation, including planning and scheduling, is a challenging task. The literature dealing with different variants of the problem is extensive. Broadly speaking, there are two main streams of work: The first is concerned with well-established scheduling problems and generalises them in the sense that they are not only applicable to a specific industry. The second stream works on less generic scheduling approaches for real cases in the industry by enriching the standard models with all the necessary realistic aspects, such as process overlaps or sequence-dependent set-up times. Furthermore, in terms of the size of the problem they can handle, the different approaches have different limitations. With the advent of Industry 4.0 or Industry 5.0, there has been a significant increase in data collection activities. The collected information is being used to construct more extensive and complex models. Therefore, there is a need to identify and highlight suitable approaches to meet the planning challenges in the era of Industry 4.0 and Industry 5.0 and to address the relevant practical problems.

We give an overview of the scheduling problems in the real world. These involve the characteristics of the resources used to perform a given task, the order in which tasks are processed, complex task relationships, and constraints on their time-aware aspects. The flexible job shop scheduling problem (FJSSP)[41] has been proposed to capture this more realistic aspect. It extends the standard job shop scheduling problem by the possibility of selecting the machining machine from a list of suitable machines—for example, Kress et al., (2019)[42].

The scheduling unit usually receives information from the internal Enterprise Resource Planning (ERP) and Manufacturing Execution System (MES). They make available information about the size of the batches and the due dates. It also provides information about specific time windows that must be met. For example, the earliest and latest start and finish times must be met, in addition to the sequence of operations to be performed. An overall due date for a job is usually associated with a cost of delay that has to be paid if the deadline is not met. Sometimes, a cost is also associated with being earlier than the deadline. Examples of these types of time windows can be found in Zhang et al., (2019)[43] or Liu et al., (2019)[44].

Furthermore, we may also consider the human aspects in scheduling problems for better integration of humans into the production process[45]. A worker is more difficult to analyse and model. A different person may have different skills, different learning rates, and different costs[46] [47].

2.2.2 Human Digital Shadows

Analytics from Wearable Biosensor

The industrial 4.0 paradigm has also led to the spread of wearable devices in work contexts (Di Pasquale et al., 2022) [48]. These devices were first developed in the medical field and then disseminated for tracking daily life and physical well-being data for personal use, presenting different functionalities suitable for industrial environments, especially for monitoring employees' psychological and physiological factors and promoting the worker's health and safety. Traditional measurements of human-technology compatibility use, for example, psycho-physiological indicators (e.g., heart rate, electromyography, or perceived human exertion) or indirect indicators (e.g., injury rates, economic losses, or operational effectiveness). Most of the devices identified in the literature for applications in production systems are positioned on the upper limbs, most of them on the wrist. These devices are mainly smartwatches, fitness trackers, and wristbands equipped with several types of sensors to monitor some human physiological and biomechanical signals (automatic recording of training times, heart rate monitoring, step count, or calories burned). Some examples of devices used are Empatica E4 or Fitbit wristbands (Akbar et al., 2019[49]; Heikkilä et al., 2018[50]; Leone et al., 2020[51]; Montesinos et al., 2019[52]; Sanchez et al., 2018[53]) and Samsung Gear S3 smartwatch (Heikkilä et al., 2018).

Analytics from Eye Tracking Technology

Eye Tracking (ET) can provide moment-by-moment insights into the cognitive state of the subject during task execution in a non-intrusive way. ET captures eye motion and gaze in response to a stimulus object and can improve our understanding of how humans behave and make decisions within complex systems. It reflects the user's visual attention, helps to quantify precisely where, how, and in which order the gaze is being directed, and provides cues to gather information on a person's intentions and current mental state (Pfeiffer et al., 2020[54]). Recent research has used ET to assess and improve human performance in various areas, for example, by studying a subject's situational awareness (SA) in aviation or road traffic (Peißl et al., 2018[55]), by investigating a subject's learning and mental state during training (Rosch & Vogel-Walcutt, 2013[56]). In manufacturing and logistics, ET has been used to detect the information extraction efficiency in a production workspace (Stork et al., 2007[57]), to predict a subject's intention and the subsequent steps of a task (Bovo et al., 2020[58]), to indicate mental workload under varying conditions (Straeter, 2020[59]), or to generate input for robot control in collaborative work scenarios (Paletta et al., 2019[60]). ET can help in identifying worker skills (Haslgrübler et al., 2019[61]) and inefficiencies in operational processes (Tuncer et al., 2020[62]), which generates valuable insights into options for improving human and process performance. A review of opportunities for using eye-tracking technology in manufacturing and logistics was outlined by (Zheng et al., 2022[63]).

2.2.3 Human Factors and Decision-Making

Cognitive fatigue is a ubiquitous human condition resulting from a sustained cognitive engagement that taxes our mental resources. Demanding work schedules lead many people to experience cognitive fatigue on a daily basis and have resulted in high burnout rates (Demerouti et al., 2001[64]; Carod-Artal & Vázquez-Cabrera, 2013[65]). Studies examining effects of such fatigue find that persistent mental resource burdens result in diminished motivation, increased distractibility, changes in information processing and poorer mood (e.g., Boksem et al., 2006[66]. Moreover, fatigued participants are more likely to fail to detect errors and less likely to take remedial action, and are more willing to take chances in everyday decision-making (Hockey et al., 2000[67]). Such general deficits can easily lead to diminished performance and health, such as progressive impairment of treatment decisions by doctors (Linder et al., 2014[68]).

Studies of human decision-making have demonstrated that stress exacerbates risk-taking. Since all decisions involve some element of risk, stress has a critical impact on decision quality (Porcelli & Delgado, 2017[69]). Decisions are found to improve with stress up to an optimal threshold beyond which deterioration is observed. Naturalistic decision-making research (Klein, 2008[70]) shows that experts may use intuitive decision-making rather than structured approaches in situations with higher time pressure, higher stakes, or increased ambiguities. They may follow a recognition-primed decision that fits their experience and arrive at a course of action without weighing alternatives.

Stress directly affects human decision-making (Galvan and Rahdar, 2013[71]): it can lead to many undesirable consequences, including a restriction or narrowing of attention, increased distraction, an increase in reaction time and deficits in the person's working memory (Driskell et al., 1999[72]). Studies of human decision-making demonstrate that stress exacerbates risk-taking and impacts decision quality. Since most managerial decisions involve some element of stress, decision aids such as DSS have been proposed to mitigate its effects. Existing research has mainly addressed two key stressors, time pressure and information overload. For a holistic understanding of decision-making under stress and to improve decision support, an extended set of stressors and psychological experiences underlying stressful decisions is examined. In terms of a class of stressors called 'Decision Stressors' there are four Decision Stressors that affect decision quality: information overload, time pressure, complexity and uncertainty.

- Stress and Risk-Taking. Decision-makers' likelihood to engage in risk varies greatly based on multiple decision-inherent features, including uncertainty, framing of a decision as a potential gain or loss (Porcelli & Delgado, 2017), and valuations of outcome valence, magnitude, and probability of receipt. As such, risk-taking decisions rely partly on stress-susceptible valuation/learning processes.
- Information overload is "a gap between the volume of information and the tools we have to assimilate". Information used in decision-making is to reduce or eliminate uncertainty. Excessive information affects problem processing and tasking, which affects decision-making. Humans' decision-making becomes inhibited because human brains can only hold a limited amount of information.
- Decision fatigue is when a sizable amount of decision-making leads to a decline in decision-making skills. People who make decisions in an extended period of time begin to lose the mental energy needed to analyse all possible solutions. Impulsive decision-making and decision avoidance are two possible paths that extend from decision fatigue. Impulse decisions are made more often when a person is tired of analysis situations or solutions; the solution they make is to act and not think.

FAIRWork implements DSS to mitigate the negative effects of stress, applying AI-driven decision-making. However, via Human Factors based analytics of the innovative decision-making process and considering the integration of stress-coping strategies above, it will lay a scientifically valid basis to design stress-reducing DSS for better, i.e., fair and more efficient decision-making.

For the purpose of Human Factors based analytics for decision-making, the digital shadow of human operators requires a so-called "Intelligent Sensor Box" (ISB). The meaning of an ISB is to represent a human operator by means of a set of parameters that are pivotal for any higher-level decision-making regime. Ideally, the human operator's digital shadow should be capable of quantifying variables that characterise its physical, cognitive, affective, social and motivational behaviour in interaction with its workplace environment. For this purpose, sensors would be attached within a body area sensor network via wearables, other sensors would apply surveillance tasks via remote measurement technology, and additional sensors would represent the workers' context within the workplace environment via human-machine and human-environment interactions. These data would be transmitted as well to a globally accessible data lake that is related to the DAI-DSS Knowledge Base after having applied an anonymisation procedure. Within this Data Lake, global key performance indicators (KPIs) and functions are stored to be input to the AI-based enrichment module that will provide machine learning for further insight and optimisation methods, both of which would enrich the global DAI-DSS environment.

2.2.4 Human Digital Twins

Recently, the term digital twin has been extended to humans using the term "human digital twin" (Miller & Spatz, 2022[73])." This term has been applied in diverse fields, including medicine (Lutze 2020[74], Chakshu et al. 2021[75]), sports performance (Barricelli et al. 2020[76]), manufacturing ergonomics (Caputo et al. 2019[77]; Greco et al. 2020[78]; Sharotry et al. 2020[79]), and product design (Demirel et al. 2021[80]).

A clear example of a human digital twin from the literature pertains to manual material handling within manufacturing or warehousing applications (Greco et al. 2020; Sharotry et al. 2020). In this example, sensors are deployed with the human in their environment that monitors the human's kinematic motion as they perform work within the environment. Other information, such as objects being lifted or forces required to activate items within the environment, is also collected. This data populates simulation models which estimate the fatigue in various muscle groups of the operator. This fatigue estimates support assessments and potential changes in work or rest schedules, handling processes, or material handling tools to improve overall worker health, safety, and productivity. Thus, a model of a human attribute is applied within a closed-loop system which fulfils the criteria for a digital twin (Boschert and Rosen 2016[81]).

Sharotry and colleagues discuss a digital twin system in a manufacturing environment which consists of a human conducting material handling tasks (Sharotry et al. 2020). The system includes a data collection module which includes motion capture, biometric suites among other sensors, a data analysis and forecasting model, and a database which contains the data which is collected for analysis. In this system, performance and operator fatigue metrics are provided to the individuals in the environment to support improved manufacturing performance and to help reduce human injury. Furthermore, this model can be used to identify material handling steps which induce substantial fatigue, permitting these steps to be evaluated and redesigned. Within this example, the real-world twin includes the manufacturing environment, the human, and the data collection subsystems. The database provides the interchange component, and the data analysis and forecasting model provides the digital representation or the digital twin.

From the viewpoint of quality-of-service, caring for attention can play a major role in the gain or loss of productivity (Paletta, 2021[82]). Attention metrics transfer into measurements of concentration, cognitive load, and situation awareness but also to an early indication of fatigue. Quality of attention is directly related to the quality of decision making, and therefore the related quality-of-service - in terms of increased fatigue, for example - indicates less safe decisions due to reduced attention spans, decreased reaction time as well as accuracy. It is well known that undetected and, therefore, frequently occurring fatigue means more team misunderstanding, a decrease of mood and motivation and- due to related physical ailments - consequently, short-term and long-term absence from work and increased medical investments. Consequently, quality of attention relates to avoiding loss of productivity due to distractions, errors, inability to concentrate and lack of motivation (Paletta et al., 2021[83]).

2.2.5 FAIRWork Research towards Human Digital Twins for Production

In the FAIRWork project, we apply only unobtrusive wearable and textile sensors to monitor and analyse the Human Factors of the worker and decision-maker. The raw data received from the biosensors are then analysed with Artificial Intelligence-based software in order to provide validated estimates of psychological constructs of various kinds. We are interested in those constructs that would provide an estimate of the quality of decision-making. The constructs refer to cognitive variables such as concentration, inhibition functionality, fatigue, and affective parameters: mood, arousal, and motivational variables and their impact on decision-making via changes in attention, working memory, and cognitive control.

Our work aims at the challenging long-term objective to enable non-obtrusive long-term analytics of workers' mental, physical, emotional and motivational stress at the manufacturing site. Observation-based analysis cannot fully interpret the psychophysical processes that reflect psychosomatic stress over long time periods. In order to estimate the actual cognitive load from the worker's interaction with the environment, the use of eye-tracking glasses is considered that continuously determine eye movements over time. In previous work, we developed a methodology to estimate concentration and mental workload from eye-tracking glasses-based data in a factory-like lab environment (Paletta et al., 2019[84]).

With above mentioned global research objectives in mind, the concrete research goals are identified as follows,

- Research on intelligent mapping of environment and machine-relevant data to digital Human Factors estimators in terms of services that integrate the local context with the objective to quantify the contextual impact on Human Factors, such as fatigue.
- Investigate appropriate decision support components and services from validated digital human factors estimators that are based on biosensor data. This mapping from human actors to digital phenotypes is termed the Digital Shadow of the real person (see Figure 1). The objective is here to validate high quality mental and physiological fatigue in terms of a digital biomarker.
- Research on risk levels for well-being of the worker that will be communicated to further components, such as the DAI-DSS Knowledge Base.
- Research and development of specific optimisation variants by using human profiles and other use case-relevant resources (for application-specific local services). This would include the consideration of quality levels for well-being / economic measures of human-machine interaction, quality levels for being capable to keep cycle time, and so on.

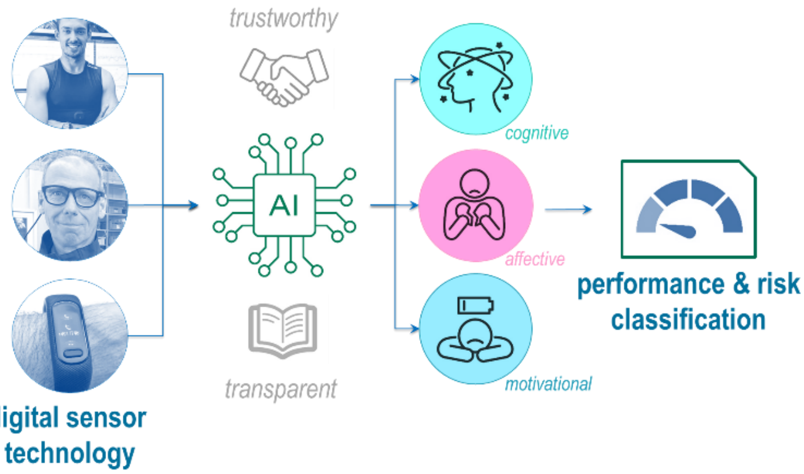


Figure 1: Digital shadows of workers and decision makers based on digital sensor technology mapping to digital human factors and risk levels for decision support

2.3 AI-Enriched Decision Support Systems

Humans cannot always cope with evaluating complex and ill-structured decision-making problems. Such unstructured, highly data-loaded challenges often lack clarity and have no single best answer. This circumstance leaves the potential for optimisation by supporting human skills with AI technologies. The aim of using AI in a decision-making process is to accelerate the speed and improve the accuracy and consistency of decision-making. AI also allows analysing big-volume data, one of the most critical points in today's digitalised manufacturing and business processes. There are various approaches to create a model or optimisation algorithm. However, for our purposes, the knowledge-based DSS proposed by Felsberger et al. (2016)[85] is the most suitable. This approach has its origins in intelligent decision support systems (IDSS) from Nemati et al. (2022)[86] and Negnevitsky (2005)[87].

The literature research presented in this section aims to identify suitable examples of AI applications within DSS systems in manufacturing environments. The objective was to comprehensively understand applied methodologies and connect them to FAIRWork use cases. Following, two primary branches of AI applications were identified:

1. Classical Decision Support Systems Enriched with AI: A hybrid approach enhances the classical decision support system (DSS) by incorporating AI methodologies.
2. AI-based Decision Support Systems: A pure AI approach based entirely on AI methodologies.

2.3.1 Classical Decision Support Systems Enriched with AI

The AI methodologies can be integrated into DSS and create a hybrid with traditional methods like rule-based decision tree, Analytical Hierarchy Process (AHP) or Computer Aided Process Planning (CAPP). Examples of classical DSS incorporating AI methodologies were selected from the literature and presented in Table 1.

Table 1: Selected applications of classical Decision Support System with incorporated AI methodology

Classical DSS	AI methodology	Application	Author
Knowledge-Based DSS	Evolutionary Algorithm	Flexible Manufacturing Systems	Kapanoglu, 2004
Computer Aided Process Planning (CAPP)	Graphplan Algorithm	Process Planning	Marchetta, 2007
Computer Aided Process Planning (CAPP)	Evolutionary Algorithm	Process Planning	Kumar, 2017
Analytical Hierarchy Process (AHP)	Fuzzy Logic	Robot Selection	Kapoor et al., 2005 Chu el al., 2003
Rule-based Algorithm	Fuzzy Logic	Robot Manipulation Tasks	Son, 2016
Multi-Attribute Utility Theory (MAUT)	Fuzzy Logic	Cost Estimation	Zhao et al., 2006

Knowledge-Based DSS with Evolutionary Algorithm (EA)

Knowledge-based DSSs require an extensive knowledge base and expertise for a limited problem variety. These systems provide decision tools for specific cases to the user, but they rely on expert knowledge and experience to function effectively. In contrast, the evolutionary algorithm-based DSS discussed in this paper does not require any acquisition of expertise. The Flexible Manufacturing Systems (FMS) engineer is not even required to be an expert on FMSs. Instead, the system utilizes an Evolutionary Algorithm (EA) with a memory of past successful experiments as the solution engine, which can maintain high performance and quality even without an expert decision-maker. Managing FMS is problematic because it requires optimizing machine loading, part scheduling, truck dispatching, and aiming for good solution quality in a short time. Due to their vast and discrete solution spaces (NP-Hard), these problems are challenging to solve as such an algorithm is unlikely to exist. Furthermore, FMS challenges are multi-criteria problems influenced by the decision-makers judgment. FMS managers must address these issues hierarchically or concurrently under time limitations, with the system's efficiency linked to response time.

To solve these issues, Kapanoglu et al. (2004)[88] suggested that DSS uses an EA as the solution engine, with a memory of past experiments identified as "good". EAs use natural selection theory to find the optimal solutions. The system's memory-based reasoning technique allows it to learn from previous events, which improves performance over time. This solution also addresses numerous FMS issues simultaneously, improving FMS management efficiency and effectiveness.

Since it solves numerous problems at the same time, the suggested DSS is versatile and applicable to diverse FMS kinds. However, depending on the FMS features, its performance may vary. Complex or specially configured FMSs that necessitate specialist optimization approaches may not be appropriate. The availability and quality of data impact the system's efficacy during training and testing. Significant computational resources and experience for implementation and maintenance are important considerations for its utilization. The system's reliance on previous experiences may hinder its ability to adjust to new or alternate circumstances or unforeseen situations.

Computer Aided Process Planning (CAPP) with Graphplan

Marchetta et al. (2007)[89] propose the incorporation of the Graphplan algorithm incorporated into generative Computer Aided Process Planning (CAPP). The CAPP uses expressive declarative languages for modelling domains and has generative capabilities, which reduces the knowledge representation needed. The AI planning model is a reasoning system that represents knowledge about goals and actions for achieving them, along with their preconditions and effects. It combines some advantages of special-purpose planners and knowledge-based systems, making it easier and cheaper to adapt and implement CAPP systems in different machining industries.

The Graphplan algorithm is a planning AI algorithm used in the experiments presented in this paper. It is a forward-chaining algorithm that works by constructing a planning graph, which is a Directed Acyclic Graph (DAG) that represents the state space of the planning problem. The nodes of the graph represent sets of propositions that are true at different levels of the plan, and the edges represent causal links between these propositions. The Graphplan algorithm works by first constructing an initial layer of nodes representing the initial state of the problem and then iteratively adding layers to the graph until a goal layer is reached. The algorithm then backtracks through the graph to construct a plan from the goal layer back to the initial layer. One advantage of Graphplan is that it can handle problems with large state spaces more efficiently than other planning algorithms. However, it has some limitations, such as not being able to handle problems with continuous variables or numeric constraints.

Computer Aided Process Planning (CAPP) with Evolutionary Algorithm (EA)

Evolutionary techniques like Genetic Algorithms (GA), Particle Swarm Optimization (PSO), and Ant Colony Optimization (ACO) are crucial in the development of AI systems in manufacturing. Kumar (2017)[90] mentioned that these methods are used to improve the efficiency of manufacturing processes, such as process planning, scheduling, and layout design. Onwubolu & Clerc (2004)[91] , for example, used PSO to minimize the operation route in a Computer Numerical Controlled (CNC) drilling process, and Solimanpur et al. (2004)[92] examined the topic of inter-cell architecture and developed a mathematical model for material flow in cellular manufacturing. The Quadratic Assignment Problem (QAP) was solved using ACO, and the results were compared to the facility layout algorithm. The parameters of machine learning algorithms employed in AI systems can also be optimized via evolutionary approaches. This can help to increase the accuracy of these algorithms' predictions, resulting in greater decision-making support for manufacturers.

The population of potential solutions in GA can be represented as chromosomes, which are strings of binary digits that encode the solution's parameters. The fitness function assesses how well each chromosome performs on a particular challenge, and the chromosomes with the highest fitness values are chosen for reproduction. Two or more selected chromosomes are combined by crossover and mutation processes to develop new solutions. Crossover is the process of swapping sections of two parent chromosomes to produce kids, whereas mutation is the process of randomly changing some bits in a chromosome. The fitness function is then used to evaluate the new candidate solutions, and the process is continued over numerous generations until a good solution is identified or a stopping requirement is satisfied. GA has been used successfully in various manufacturing applications, including process planning, scheduling, layout design, and cutting parameter optimization.

Analytic Hierarchy Process (AHP) with Fuzzy Logic

The Analytic Hierarchy Process (AHP) is a traditional multi-criteria decision-making approach. It involves breaking down a complex decision problem into a hierarchy of smaller, more manageable sub-problems and then using pairwise comparisons to compare the relative importance of each sub-problem. Kapoor et al. (2005)[93] propose a methodology that modifies the AHP by introducing fuzzy linguistic variables in place of integers. This solution enables the introduction of fuzziness into decision-making, which can better depict the imprecise and unpredictable nature of real-world problems. The methodology produces a fuzzy interface for any algorithm by using fuzzy logic to measure the values of input and output variables. This is accomplished using a trapezoidal distribution across the normalized 'Universe of Discourse'. Fuzzification assigns linguistic variables to numerical values of membership functions and develops appropriate decision rules. Finally, defuzzification reduces fuzzy output into crisp values, yielding a fuzzy score as a result. Consequently, adding fuzzy logic into AHP increases its ability to deal with complicated decision issues involving inaccurate or uncertain data.

Rule-based Algorithm with Fuzzy Logic

Son (2016)[94] proposes, in his study, an intelligent rule-based sequence-planning algorithm with integrated fuzzy optimization for robot manipulation tasks. The suggested algorithm is a rule-based method that uses stated rules to generate a feasible path. The degree of uncertainty associated with path planning task execution is employed as an optimization criterion. This approach uses fuzzy logic to optimize robot movements in partially dynamic settings. According to the author, traditional optimisation approaches may not be effective for robot manipulation tasks in dynamic environments since they require exact and complete information that may not be available in real-world scenarios. The suggested program can manage uncertain or imprecise information and make more informed decisions about robot motions.

Fuzzy Multi-Attribute Utility Theory

To determine the price of composite products, Zhao et al. (2006)[95] propose the Fuzzy Multi-Attribute Utility Theory (FMAUT), which combines the Multi-Attribute Utility Theory (MAUT) with fuzzy logic. The MAUT is a decision-making technique that assesses options based on numerous characteristics or criteria. Each attribute level is given a utility value, which reflects the level of preference for that level. The option whose total utility value is highest is regarded as the most optimal option. However, expert evaluation of the connections between cost and design feature levels is often ambiguous and confuses the experts' community. Because the utility value associated with a certain level is not a set and unchanging quantity, and the expert preference is uncertain, membership functions should be used to express it. Consequently, fuzzy logic can be applied in this scenario.

Fuzzy sets and membership functions can be used to describe hazy or imprecise concepts. An element may belong to a set to a degree ranging from 0 to 1 in a fuzzy set, which has degrees of membership. A membership function maps an element's level of membership in a fuzzy set. Instead of a utility value, as in the MAUT, a membership function with fuzzy true values compared to the utility values is assigned to each feature level in the FMAUT. This enables the cost assessment process to consider expert judgments about correlations between cost and design feature levels, even though they are frequently hazy.

Early in the design process, when manufacturing knowledge is typically lacking or ambiguous, the FMAUT can be used. With this cost model, cost engineers with little expertise can quickly determine the cost from the cost model by selecting a feature level for each characteristic following a particular design. When working with complicated systems that entail numerous qualities or criteria with hazy or imprecise ideas, FMAUT can incorporate expert opinions into decision-making procedures.

2.3.2 AI-based Decision Support Systems

Contrary to the previous subchapter, the focus in this section shifts towards DSSs, which solely utilize AI methodologies. The summary of selected applications of AI-based DSS is provided in Table 2.

Table 2: Selected applications of AI-based Decision Support Systems

AI methodology	Application	Author
Fuzzy TOPSIS	Robot Selection	Chu et al., 2003
Meta-heuristic method (GA) & Artificial Neural Network (ANN)	Intelligent Feature Selection	Mohammadhossein et al., 2020
Fuzzy Wavelet Neural Network (FWNN)	Supplier Selection	Rong et al., 2012
Random Forest (RF)	Tool Wear Prediction	Wu et al., 2017
Reinforcement Learning (RL)	Production Scheduling	Waubert de Puiseau et al., 2022 Samsonov et al., 2021 Tassel et al., 2021
Reinforcement Learning (RL) Graph Neural Network (GNN)	Production Scheduling	Zhang et al., 2020

Fuzzy TOPSIS

Chu et al. (2003)[96] propose a Fuzzy TOPSIS method for robot selection that uses fuzzy numbers to represent subjective criteria and dimensionless indices to ensure compatibility with objective criteria. The method facilitates decision-making by providing a systematic strategy for analyzing and rating different robots based on multiple criteria. The suggested solution enables decision-makers to consistently and transparently assess the importance of various criteria and the performance of various robots under those criteria. This can assist decision-makers in making more informed selections regarding which robot to choose for a specific application.

Meta-heuristic method (GA) and an Artificial Neural Network (ANN)

Smart manufacturing refers to optimization techniques implemented in production operations using advanced analytics approaches. Ghahramani et al. (2020)[97] suggests a dynamic feature selection model based on an integrated algorithm, including a meta-heuristic method (GA) and an Artificial Neural Network (ANN). This model is designed to determine the optimal number of features and their relevant cost, which are used to create a predictive model for controlling manufacturing processes. Traditional approaches, such as PCA-based algorithms, may not sufficiently capture nonlinear relationships and complicated patterns in manufacturing data.

The GA selects the most relevant characteristics from a vast amount of data to extract patterns. These significant characteristics in the GA for data manufacturing are chosen based on their fitness score, which is defined by how effectively they contribute to obtaining the intended goal. This approach is repeated until a suitable collection of characteristics that can predict outcomes and identify trends in manufacturing data is determined.

Fuzzy Wavelet Neural Network (FWNN)

The Fuzzy Wavelet Neural Network (FWNN) strategy for supplier selection, introduced by Guo et al. (2012)[98], involves two stages. In the first stage, a fuzzy system converts fuzzy input parameters into crisp parameters. This level makes extensive use of excellent knowledge representation competency. In the second stage, an ANN is used to reduce the negative effects of supplier selection with non-linear and dynamic properties. The simulation was created to test the algorithm's performance. FWNN's advantages over other approaches include its ability to analyze suppliers efficiently and solve shortages of traditional methods. It also blends fuzzy theory and ANN, making it more successful in dealing with complex and uncertain data in service-oriented manufacturing networks. The FWNN supplier selection technique is intended exclusively for service-oriented manufacturing networks. It is a flexible service network based on capacity needs, and it takes into account the particular characteristics of service-oriented manufacturing networks, such as supplier uncertainty and complexity.

Random Forest (RF)

Wu et al. (2017)[99] use Random Forest (RF) algorithm in smart manufacturing systems to forecast tool wear. RF algorithm can anticipate tool wear by using cutting force, vibration, and acoustic emission signals as input variables. RF is a data-driven machine learning system that generates predictions using an ensemble of decision trees. It builds each decision tree using a randomly sampled portion of the training data and the most vital characteristics that potentially predict the outcome. During testing, the input variables are routed through each decision tree, with the average result from all trees serving as the final forecast. The input signals for tool wear prediction in smart manufacturing systems are cutting force, vibration, and acoustic emission data. The output of the predictive model is the estimate of tool wear over time. RFs have the potential to handle a large number of input variables and avoid overfitting the model. The experimental results have shown that the predictive model trained by RFs is very accurate.

Reinforcement Learning (RL)

DSSs have been widely used to assist decision-makers in various industries. With the emergence of AI, the DSS can now be enhanced to solve more complex problems, such as resource mapping and solution configuration problems in manufacturing. Resource mapping problems are commonly encountered in manufacturing industries, including flow shop and job shop scheduling problems. Due to its NP-hardness, the job shop scheduling problem is considered one of the most challenging problems in this domain.

In a job shop, multiple jobs have to be processed on different machines, each with its processing time and requirements. The goal is to minimize the overall completion time or the makespan while ensuring that each job is completed without violating the constraints. The complexity of the job shop scheduling problem makes it challenging to find optimal solutions manually. However, AI techniques, such as Reinforcement Learning (RL), can be used to develop intelligent decision support systems that can learn from data and experience to solve such problems.

Literature suggests that AI-based approaches were researched with job shop scheduling problems to improve scheduling performance. These approaches range from traditional optimization algorithms to advanced machine learning algorithms like RL – a promising technique that can learn an optimal policy by interacting with a so-called environment iteratively. The makespan, or the time required to complete all tasks, is frequently selected as an optimization metric in job shop scheduling. A schedule's makespan can typically only be stated for the entire schedule, not for a partial one. Therefore, the intuitive formulation of a makespan-oriented reward function is sparse, giving non-zero rewards only when the agent has fully resolved a problem. In RL, dense rewards are known to be more favourable. Consequently, a dense reward function is preferred. Reward shaping is the process of designing the reward function to encourage RL algorithms to utilize domain knowledge to solve a task more effectively. Waubert de Puiseau et al. (2022)[100] examined reliability-focused RL production scheduling techniques. Samsonov et al., (2022)[101] describe a variety of JSP approaches with various optimization objectives, including, among others, tardiness minimization, machine utilization, and makespan minimization. Samsonov et al. (2021)[102] assess as well several strategies with regard to state representation, the construction of the action space, the RL algorithm, and the particular use case.

For the JSP, both sparse and dense reward functions have been considered. A sparse reward function was published by Samsonov et al. (2021), Zhang et al. (2020)[103], and Tassel et al. (2021)[104] have proposed dense reward functions. Each strategy integrates its reward mechanism in a particular environment. Between the various techniques, there are substantial differences in the modelling approach, observation space, and action space. It is challenging to determine how much the reward function and other modelling aspects contributed to the method's effectiveness.

Furthermore, Samsonov et al. (2021) developed a discrete time simulation for the JSP. Each task is handled by a virtual machine. When decisions need to be made about allocating tasks to machines, the agent receives requests for input. The suggested action space considers a set of possible jobs with a fixed-sized normalized processing time. In this manner, the size of the action space is indifferent to the magnitude of the problem. The presented reward function is sparse. Only the final step of an episode provides a reward; all other phases yield zero. Because the return is qualitatively inversely correlated with the makespan, improvements close to the optimal solution are rewarded particularly high.

Similar to Samsonov et al. (2021), Tassel et al. suggest formalizing the JSP in a time-discrete manner. The state representation, action space, and reward function are distinct. A choice is made for the assignment of open jobs (or tasks, respectively) to available machines at each time step. The reward system depends on how much the machine is utilized. The number of idle moments created on machines as well as the area a task occupies on the Gantt chart are taken into account when calculating rewards. Tassel et al. optimize the scheduled area rather than the makespan directly. The makespan greatly depends on the waiting durations for specific tasks. The Gantt chart displays waiting times as free areas or holes. Consequently, fewer holes or a larger planned area result in shorter makespans. The scheduled area can be evaluated in each time-step, which yields a dense reward function.

Instead of using a discrete-time simulation, Zhang et al., (make use of the JSP's formulation as a disjunctive graph. Additionally, Zhang et al. use a graph neural network (GNN) and disjunctive graphs to transform the state into a certain fixed size. At each scheduling step, the policy considers data from the embedded graph. The increment in the critical path during scheduling constitutes the foundation for the reward function. The algorithm may resolve instances of various sizes using the GNN. Zhang et al. studied its ability to learn a generalized solution process for JSP instances after training the agent with a large number of computer-generated JSP examples.

Currently, intelligent systems for decision-making use neural networks[105], genetic algorithms[106], decision trees or tables[107], fuzzy logic, as well as other accessible AI models or optimisation algorithms. The research area of the implementation of AI in the DSS is wide-ranging. Therefore the first exploration steps focus on AI-supported decision-making at the process level as well as at the manufacturing operation (e.g. job shop planning) and control level. The initial research provided a broad overview of the current state of the art and thus led to defining the following research questions:

- How can the decision-making process and architecture in the manufacturing environment be accelerated through the use of AI methodologies?
- How can existing AI methodologies enrich classical decision support systems in order to be applied in complex and dynamic manufacturing processes?
- How do different optimization metrics or constraints affect a schedule in a manufacturing context?
- How and to what extent can AI techniques be utilized for optimisation in industrial scheduling?

2.4 Reliable and Trustworthy AI

The available literature in organizational behaviour and psychology indicates that individuals' inclination to adopt new technology is influenced by their perceptions and beliefs about that technology[108] [109]. In this context, the Technology Acceptance Model (TAM) and its related frameworks have been widely used to understand the behavioural aspects of technology adoption and acceptance. Initially, the TAM focused on perceived usefulness and ease of use (Davis, 1989[110]; Venkatesh & Davis, 2000[111]). However, subsequent research has expanded upon this model by considering additional factors, moderators, and outcomes within the broader domain of technology acceptance (Venkatesh et al., 2016[112]). Trust has emerged as a significant driver of technology acceptance among these additional factors (Felzmann eta l., 2019[113]).

Trust is typically conceptualized using dimensions such as ability, benevolence, and integrity (Mayer et al., 1995[114]). In the context of technology, Söllner et al. (2012)[115] propose differentiating trusting beliefs based on dimensions like performance, process, and purpose. What is more, research suggests that people tend to have less inherent trust in AI compared to trust in humans. This phenomenon, termed "algorithm aversion", indicates that individuals are hesitant to trust algorithmic forecasters once they have seen them err even when they – in the long run – outperform human forecasters[116]. In subsequent studies, Dietvorst et al. (2018)[117] found that people were more likely to rely on algorithmic forecasts if they could modify a fraction of the AI's predictions, compared to not being able to change the AI's predictions. Additionally, trust in machine learning (ML) has been observed to develop more slowly and decline more rapidly than trust in humans (Dzindolet et al., 2003[118]). This could be attributed in part to the widespread media coverage of AI failures, such as instances where an Amazon recruiting tool shows a bias against women, autonomous cars end up in crashes, or the transformation of a Microsoft chatbot into a white supremacist [119] [120] [121].

These AI failures emerged due to the increasing complexity of AI solutions, which are so complex that their inner processes have become black boxes. That means even the developers themselves cannot reconstruct these black-box processes and decisions. Realizing this, developers and AI experts have increased their work on interpretability and explainability. Subsumed under the umbrella term transparency, the terms interpretability and explainability or explainable AI (XAI) stand for research from a solely technological viewpoint (Arrieta et al., 2020[122]; Rai, 2020[123]; Murdoch et al., 2019[124]). These approaches in computer science are based on the assumption that "by building systems that are transparent, interpretable, or explainable, users are better able to understand and thus trust the intelligent agent" (Miller, 2018[125]). From a technical perspective, explainability or interpretability offers three advantages (Arrieta et al., 2020):

- They help to ensure impartiality in decision-making, i.e. to detect and, consequently, correct bias in the training dataset.
- Facilitate the provision of robustness by highlighting potential adversarial perturbations that could change the prediction.
- Can act as insurance that only meaningful variables infer the output, i.e., guaranteeing that an underlying truthful causality exists in the model reasoning

While technological solutions were on the rise, social scientists realized that the solutions presented under the term of explainable AI were not understandable by the often lay end users. That is, researchers such as Paéz (2019)[126] and Miller (2018) requested to turn towards end users. Páez called for research on understandability which he called more important than detailed information about AI processes but a holistic investigation on how to make AI really understandable to increase trust. Such requests have since led to a rising number of research on transparency for end users.

At the same time, the AI failures, together with the widely unregulated market of digital services, have led the EU to work on legislation for regulation. It was the EU as the very first governmental organ to install a privacy act (GDPR) in 2016, and since then, other worldwide legislation has followed. Besides regulating the storage and access of private data, the act contains a paragraph on transparency[127]. It says that users have to have the right to know how platforms or support systems make decisions, i.e. why specific results are produced and shown. In the following years, the European Commission installed a high-level expert group on artificial intelligence (AI HLEG). These experts had the assignment to counsel on AI and respective legislation to eventually "shaping Europe's digital future", their slogan[128]. In their first document, they stated that "those [legal regimes] regarding transparency, traceability and human oversight are not specifically covered under current legislation" (AI HLEG, 2019, p. 9). This is why they published a Guideline for Trustworthy AI. Based on basic rights, it explains that for AI to be trustworthy – a goal of AI in general – it needs to be:

(1) lawful - respecting all applicable laws and regulations,

(2) ethical - respecting ethical principles and values,

(3) robust - both from a technical perspective while taking into account its social environment (AI HLEG, 2019[129]).

Based on these three principles, the AI HLEG developed seven Key requirements that AI has to fulfil to be considered trustworthy (AI HLEG, 2019):

1. Human agency and oversight
2. Technical Robustness and safety
3. Privacy and data governance
4. Transparency
5. Diversity, non-discrimination and fairness
6. Societal and environmental well-being
7. Accountability

Based on the documents and these guidelines, the expert group has worked on and published "Policy and Investment Recommendations for Trustworthy AI", which contained 33 recommendations for legislation concerning AI[130] It was presented at the first European AI Assembly in 2019. Further work followed in 2020.

Again, in these documents and AI requirements, transparency plays a central role. In this understanding, transparent AI means that data, systems and AI business models should be transparent – which they understand as a process adaptive to all stakeholders. Finally, transparency contains the right to know that users are interacting with an AI system and to be informed about its limitations and capabilities.

At the moment, the EU is working on the Digital Service Act, which is thought to regulate the whole market of digital services. While very large platforms and service providers (with more than 45 Mio users, i.e. more than 10% of the European population) already have to adhere to regulations, the legislation is to be fulfilled by all online services from February 2024 onwards[131]. The act is worked on and discussed as this document is published. That is, its implications for decision support systems as developed in FAIRWork need to be seen and considered. Besides defining and regulating highly risky services in a very strict manner, it will contain transparency requirements that all services have to adhere to.

Not only the EU but also researchers such as Zerilli et al. (2022)[132] mark transparency of AI systems to be one important antecedent to build trust among human users. Trust in AI systems is generally built through several key factors, and the dynamics of trust can vary depending on whether humans are operating the AI system or simply being affected by its results (Jobin et al., 2019[133]). When humans have direct control over the AI system and actively operate it, trust is usually assumed to build through factors such as:

a. System Performance and Reliability: Humans develop trust when the AI system consistently performs well and reliably produces accurate results. Frequent errors or inconsistencies can erode trust (e.g., Hoff & Bashir, 2015[134]; Ross et al., 2008[135]; Zerilli et al., 2022)
b. Explainability and Transparency: Humans need to understand the underlying processes and decision-making mechanisms of the AI system. Transparent AI systems that provide explanations for their recommendations or decisions foster trust and confidence (e.g., Glikson & Woolley, 2020[136]; Miller et al., 2014[137]; Hoff & Bashir, 2015)
c. Control and Agency: Allowing humans to maintain a level of control and agency over the AI system's actions builds trust. When humans have the ability to intervene, override, or provide feedback to the AI system, they feel more in control and are more likely to trust its outcomes (e.g., Dietvorst et al., 2018)

In scenarios where humans are not directly operating the AI system but are influenced by its results, trust is generally assumed to establish similarly:

a. Accuracy and Consistency: Humans tend to trust AI systems that consistently produce accurate results. If AI consistently provides reliable outcomes, humans are more likely to trust its recommendations.
b. Transparency and Explainability: Even if humans are not directly operating the AI system, they still value transparency and explanations for AI-generated decisions. Understanding how the AI arrived at its recommendations or conclusions helps individuals trust and comprehend the outcomes.

In a literature review on trust in automation, Hoff and Bashir (2015) differentiate between external (i.e. system and context) factors and internal (i.e. user) factors that impact a user's trust in a system. External factors include system type, system complexity, the task for which a system is used, the potential risks and benefits of using automation, the organizational setting of an interaction, the framing of a task, and the operator's workload. The internal factors that can impact trust include self-confidence, subject matter expertise, mood, and attentional capacity.

Trust functions as an important prerequisite of technology acceptance, adoption, and use in general (Venkatesh et al., 2016[138]) and for AI in particular (Siau & Wang, 2018[139]). At the same time, these assumptions miss out the fact that trust has a multidimensional character. This leads to the fact that "there is no simple correlation between explanation and trust, and that an adequate analysis of trust requires taking into account contextual factors that can foster or hinder it." (Páez, 2019, S. 9).

The ambivalent results of AI-user-interaction studies researching trust as a dependent variable support this multidimensional, contextual character of trust. While some researchers suggest that transparent systems "can facilitate appropriate trust [...] and improve human-automation task performance" (Hoff & Bashirt, 2015, p. 423), others suggest that transparency is negatively associated with trust (Schmidt et al., 2020)[140]. If users are asked before interacting with a system whether they prefer a transparent or a non-transparent system, they choose the transparent system. After use, however, the picture is less clear. In Springer's (2019) studies[141], users received very detailed feedback regarding individual words they typed. This level of detail was sometimes perceived as too high, and sometimes it led to assumptions about the way the system worked, which, although actually goal-directed, was perceived as less than optimal, e.g., because it contradicted the users' expectations. It turned out that users considered explanations useful only when the system was perceived to be faulty or not in line with their expectations. When asked after use, half preferred the non-transparent system (Springer, 2019). Schmidt et al. (2020) gave users explanations about the expected accuracy of a system as well as highlighted results as an explanation for the system's function. Again, users' usage of the system was lower with information and lowest when they received both. The studies show that transparency changes – rather than facilitates – the understanding of a system, sometimes with a negative effect. Kizilcec concluded in a study that compared expected outcomes vs unexpected outcomes in transparent vs non-transparent systems that "Designing for trust requires balanced interface transparency—not too little and not too much." (Kizilcec, 2016, p. 2390[142]).

Moreover, research shows that the relationship between trust and usage is not as close as some studies might indicate. Rather, Hoff and Bashir (2015) came to the conclusion that "situational factors play a leading role in determining the extent to which trust influences behaviour toward automation" (p. 418). In a model they developed, they list several factors that influence the strength of the relationship between trust and behaviour:

- The complexity of automation,
- The novelty of the situation,
- Operator's ability to compare automated performance to the manual,
- Operator's degree of decisional freedom.

The model indicates that, for instance, even if the user's trust is high, if the situation is very novel, they might not use an automated system. On the contrary, even if trust is low, they might use an automated system if they have to, i.e. when their decisional freedom is low.

In times when AI and its applications are on the rise – and dealt with in an increasing number of research projects – the matter of building reliable and trustworthy AI is an important one (AI HLEG, 2019; Larsson & Heintz, 2019[143]; Felzmann et al., 2019[144]). As described, practitioners and an increasing number of political institutions require transparency and control for AI systems (DSA, 2023; AI HLEG, 2019; Jobin et al., 2019[145]). However, whether transparency is meant to build trust or rather, the relationship between transparency and trust remains volatile. Research questions that arise comprise the following:

- What is the relation of transparency and trust, acceptance and usage in the application of production lines and Multi-Agent Systems, i.e. the FAIRWork project?
- Which system factors influence the relation of transparency and trust, acceptance and usage in more general terms?
- Which context factors influence the effect of transparency on trust in the use cases?
- How can context factors that influence the effect of transparency on trust be clustered in broader terms and for applications beyond FAIRWork?
- How can transparency be provided in an understandable way for all stakeholders in FAIRWork?
- How can AI systems be designed to enable human performance in the best possible way?

3 RESEARCH METHODOLOGIES

This section focuses on the research methodologies employed to investigate the technical aspects of decision-making processes, specifically the use of MAS and AI and the human factors involved in this process. The chapter also covers investigating digital human factors measurement technologies for decision support. The research methodologies used involve a range of techniques, such as laboratory experiments, questionnaires and modelling methods. Using these methodologies, the research team aims to provide a comprehensive understanding of the technical and human factors involved in decision-making, including the reliability and trustworthiness of AI.

3.1 Investigation of AI and MAS Aspects in Decision-Making

Research methods in data science, AI and MAS refer to systematic approaches and techniques used to investigate technological challenges, improve and implement existing models in real use cases, and develop new technologies. Existing methods often provide a structured framework for conducting research on collected data and drawing conclusions from results. Methods like data-driven modelling, prototyping and testing can be used within the research in mentioned domains.

Data-driven modelling is a process of building models and making predictions based on the available data. It involves utilizing statistical, machine learning, and deep learning techniques to extract patterns, relationships, and insights from the data. These models can be used to make informed decisions, optimize processes, or generate predictions based on real-world observations. They can also be used as an improvement of the existing traditional methods arriving at hybrid solutions.

Furthermore, engineers often create prototypes to validate designs and test their functionality. The exemplary prototype is usually a scaled-down version or a representative model of a product or system, followed by rigorous testing to assess performance, reliability, and durability. The studies in the domains mentioned above are mostly focused on the use cases, which involve in-depth analysis of specific technical projects, systems, or processes. Data scientists collect data from real-world situations to understand challenges, identify solutions, and draw conclusions that can be applied to similar scenarios. There are 4 experiment laboratories available in FAIRWork project:

- OMiLAB – Open Models Initiative Laboratory,
- Human Factors Lab (HFL) at Joanneum Research,
- Robotic Lab of Flex,
- CRF Lab.

It is important to note that the specific research methods employed in engineering research can vary depending on the nature of the problem, the available resources, and the research goals. Scientists commonly combine multiple methods to gain a comprehensive understanding and address complex engineering challenges effectively.

3.1.1 Research Strategy for Decision Support System enrichment with AI

Figure 2 illustrates the research strategy for enriching DSS with AI, one of the key components of the FAIRWork project. It is divided into three main phases: conceptual, examinational and summary.

The starting phase is focused on establishing the research foundations. This phase consists of a literature review, formulation of research questions, definition of research factors as well as possible on-site visits and preliminary demonstration of research directions. The research questions outline the specific aspects or problems that the research aims to address; hence the research factors provide an understanding of the variables, components, or elements that contribute to the phenomenon under study. An on-site visit may be conducted depending on the nature of the research. This visit allows researchers to gather first-hand information, observe a real industrial environment, and gain a deeper comprehension of the subject matter. In the conceptual phase, an initial demo is also provided. This step involves creating a prototype, model, or simulation that represents the intended solution or approach. The demo design serves as a proof of concept and provides a tangible representation of the research idea.

The examination is the second phase of the research strategy for DSS enrichment with AI. In this phase, the focus shifts towards data collection and analysis, experimentation with various AI models to address the research questions and selecting the most appropriate one. Data collection and understating is the key aspect of conducting data-driven modelling. Therefore, this matter is also underlined and extended in Chapter 4. Depending on available data, different AI methods and approaches are explored and evaluated. The aim is to select the most appropriate method that aligns with the research objectives and provides reliable results. This step goes along with AI model development that involves data preparation, model design, training and validation.

The final phase of the research plan involves summarizing the findings and drawing conclusions. The research findings and conclusions are documented and disseminated through the FAIRWork innovation shop, conferences, presentations, or academic journals. It serves as a record of the research process and outcomes, contributing to the respective research field (see Section 5).

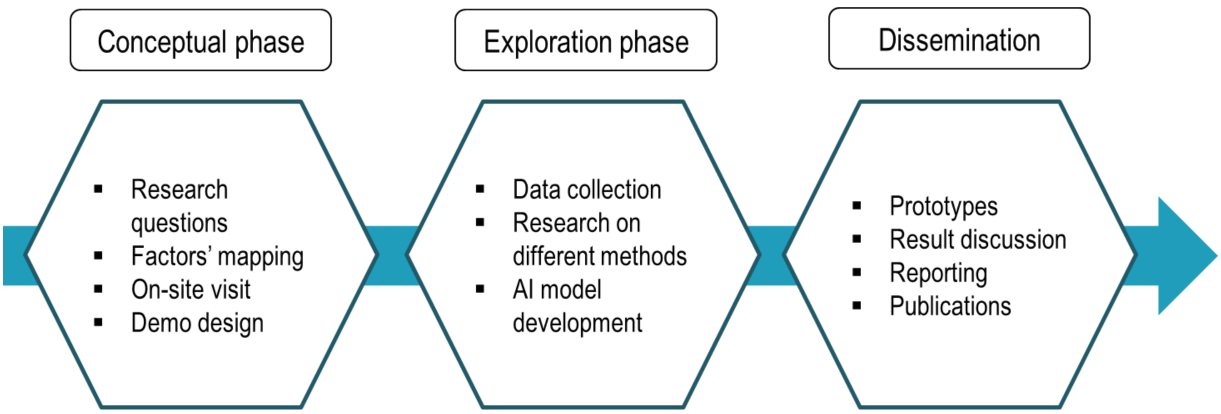


Figure 2: A research strategy for FAIRWork DSS enrichment with AI.

3.1.2 Research Strategy for Decision Support System Enrichment with MAS

Decision Support Systems (DSS) are rapidly evolving, with Multi-Agent Systems (MAS) paving the way for significant advancements. Integrating MAS into DSS can revolutionize how organizations approach complex decision-making tasks, enhancing both the efficiency and effectiveness of these systems. This research strategy aims to explore the enrichment of DSS with MAS, studying how these systems can collaborate and adapt within a DSS context, providing added value to tackle challenges in decision-making in complex environments. This strategy comprises the definition of how interactions between agents are modelled, their implications for decision-making, and the potential for achieving greater decentralization, robustness, and scalability in decision support.

The methodology of the work in the MAS field is presented as follows (see Figure 3):

1. Identify items for “agentification”: every asset, operator, production order or stakeholder involved in the process is properly modelled as an agent. Then, their behaviours, skills and communication rules are set. It is important to consider the role each agent will have within the system and how their interactions can be developed.
2. Goals identification: the system’s objective is segmented into goals that represent part of this objective. These goals are pursued by groups of agents that aim at reaching their established goals in order to achieve the overall system’s objective.
3. Static Modelling: The modelling of the system begins with UML diagrams that facilitate the visualization of the classes that are developed to represent the agents and the interaction scheme for representing the message exchange between agents. Static modelling is important for making the interactions clear and simple as possible. It is the initial part of making a complex system segmented into smaller and well-defined systems that are manageable and clear.
4. Dynamic Modelling: it comes into play to define each step the agent follows through their lives. Their functioning is represented between states and transitions in a Petri Net. It offers a graphical notation that allows visualization of concurrent actions and dynamics within the system with its well-defined semantics that leads to an unambiguous interpretation of the system.
5. Multi-Agent System development: in this step, the actual development of the agents is performed in code. Based on the previous steps, the current ideas are translated into a concrete software implementation. Each agent's functionalities, interactions, and decision-making processes are coded. The environment and communication channels are established, ensuring all agents can interact as designed.
6. Use case deployment and functional validation: the implemented system is deployed in a use case scenario where the functional validity of the system is verified. Each agent’s behaviour and interaction are monitored, and the overall system performance is tracked to ensure the MAS is aligned with the predefined system objective.

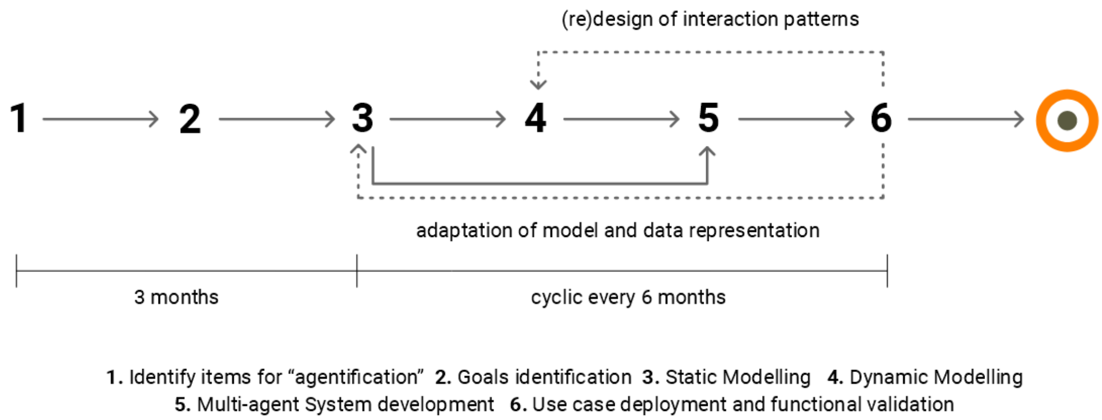


Figure 3: The methodology of research work for MAS.

From step 3 to step 6, the process can be iteratively adjusted, given new observations. It can be done based on the role of the agents in the system as the decision-making process itself becomes more complex with the time that would affect its modelling (steps 3 and 4) or by adjusting parameters in the last steps (steps 5 and 6) to achieve more optimal performance.

The MAS approach is triggered whenever the problem at hand is of a complex nature and needs the added value that MAS can provide through its autonomic, distributed control, and cooperative features appreciated in many of the challenges encountered by industry. The DAI-DSS Orchestrator is responsible for identifying the MAS-based approach as a possible solution to the problem and triggering the Production Process Simulation with Agents service or the Multi-Agent based Resource Allocation service according to the challenge to tackle.

Finally, we conclude with the MAS abilities that satisfy the requirements of the modelled problems, capable of offering flexible solutions to complex problems. An approach of distributed character and of cooperative scope between entities that represent the parts of fundamental interest for the solution of the problems. The conclusion is a system that offers services of extreme interest for applications in optimising resource allocation in production or agent-based simulation. FAIRWork is equipped with a robust mechanism that adapts to the demands of an industry that requires increasingly challenging technological solutions. A number of valuable attributes are added to a system that is able to offer solutions to issues that require high flexibility and impose a complex and highly scalable approach.

3.2 Investigation of Human Aspects in AI-supported Decision-Making

The case studies make it possible to deal with the manifold problems of a socio-technical "doing democracy". However, this also requires a design that allows for two things. First, achieving sufficient depth of focus in analysing socio-technical practices is crucial. Only such a view allows the representational needs of the workers on the production lines and the managers of the production lines to be adequately captured. If not, then the demand for the democratic design of the tool cannot be met. On the other hand, the identified needs for representation must be sufficiently generalisable. Against this background, the general strategy is to link a key case with cases for further refinement of the results found.

Furthermore, the lead case is raised in a three-stage procedure (see Figure 4). This procedure ensures that the expected depth of focus can be achieved. The first step with a one-site visit serves the purpose of onboarding and setting the framework for the empirical case study. The second step serves as a deepening, in which the socio-technical practices on the production line are recorded and examined as precisely as possible. Finally, the third step serves to translate the results back into the context of the use cases to ensure that the representation needs have been captured with sufficient precision.

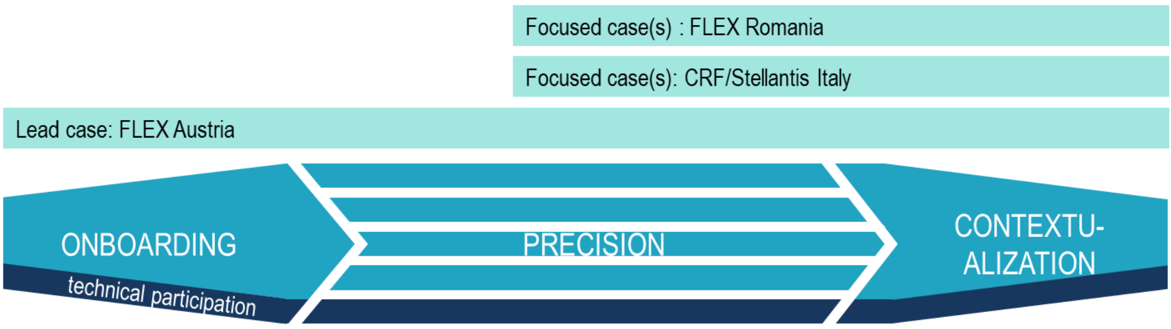


Figure 4: Extensive research plan for the investigation of human aspects in AI-guided decision-making in FAIRWork

At all steps, sociological, psychological and human-technological interaction approaches are used to investigate the practices, interactions, and application options. This initial phase takes place at one lead case, for which FLEX Kaernten was chosen due to its use case variety and accessibility. During an on-site visit, as a common starting point, observations and use case explorations set the ground for context understanding and embedding. Exploratory interviews with humans operating the system and being affected by it provide first insights into subjective standpoints and perspectives, but also preferences, fears, and assumptions. Having collected these primary data of situational as well as subjective human impressions, the selection of research settings and specific cases for further investigation in the second step can happen that set the foundation for the second phase of precision.

In this second phase, the research lines will – with close cooperation – run in parallel to deepen specific topics of interest. While on site-interviews and observations will analyze socio-technical practices to gain an understanding of processes, approaches for democratic implementations and their possible effects, other research will be conducted in experimental settings. On the one hand, these consist of human factor measurement technologies which are described in Section 3.3. On the other hand, transparency and trust experiments, both present and online, deepen the understanding of the relationship between transparency factors and trust, acceptance, and usage of AI-supported decision-making. Transparency aspects mean, in this context, global system explanations and single result instructions as well as democratic aspects, i.e. information about relations, accountabilities, and decision processes. Varying single aspects and measuring their impact on trust, acceptance, and legitimation provides the opportunity to derive insights from smaller experiments to the bigger picture of use cases. Figure 5 maps the concept of exploring AI-assisted trustworthy decision-making through transparency in an overview diagram.

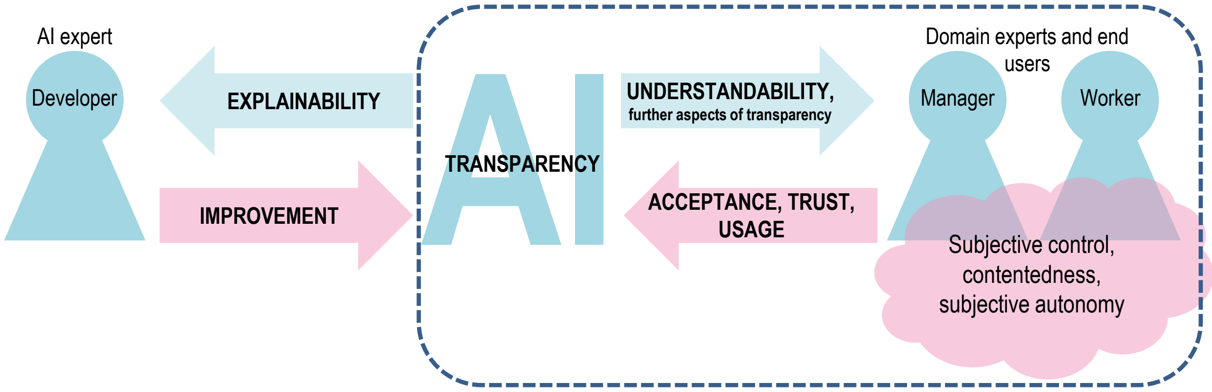


Figure 5: Overview concept of exploring AI-assisted trustworthy decision-making through transparency

Combining all research lines, this approach of the second phase serves the goal to gain a deeper understanding of the human perspective: how it is influenced by context, which relations come to play and which effects can be expected have to be considered and modulated when implementing democratic decision-making systems. While the lead case of FLEX Kaernten is deepened during this second phase, the other use cases from FLEX and CRF are added for both focusing the research of different streams and transferring the results to several use cases.

The comprehensive transfer of results back into the use cases will take place in the final phase, the one of contextualization. The results of the previous phases are integrated into practice in discussion rounds on-site with parties from different hierarchical levels. This procedure serves both to frame the results for the use case partners and to validate them for the researchers. As a second part of these discussion rounds, the implementation of planning is used – to get the results up and running for use case partners. This way, even possibly more abstract results can be turned into factual and manageable chunks for the industry. In the end, the application and implementation of the results are set on their way.

3.3 Investigation of Digital Human Factors Measurement Technologies for Decision Support

Digital technologies for the measurement of Human Factors are available on the basis of monitoring devices of the human state. These components are either wearable or environment-mounted sensors for the purpose of studies and standard measurements about the behaviour of human decision-makers in the industrial environment.

The following sections outline the sensors that enable capturing data with critical information about the mental, affective and motivational state of the human. After that, the focus shifts to the basic implementation details regarding the Intelligent Sensor Box (ISB) as a fundamental information processing unit in the industrial orchestrated DAI-DSS architecture. A fundamental processing part of the ISB is represented by the risk-level-oriented computing units that provide important data for the Decision Support System. Finally, we propose a novel framework to handle human profiles in the Industry 5.0 system by means of Personas in terms of Human Digital Twins for Decision Making.

3.3.1 FAIRWork Sensors for Human Factors Measurements

3.3.1.1 Sensors Mounted on Human

Non-obtrusive wearable sensors are the basis for measuring human factors during the execution of tasks at the workplace. The digital sensor network in FAIRWork consists of wearables that are specifically suited to measure psychophysiological variables, such as cognitive-emotional stress, physiological strain, and fatigue. In the following, we present wearable sensor systems that can be deployed within the scope of the project FAIRWork.

Cardiovascular Information Processing from the Biosensor Shirt

The main sensor equipment within this project is the QUS smart shirt solution. Based on this smart shirt product (see Figure 6), which is already used extensively in competitive sports, empirical values from many sports are at our disposal. Approaches that have a direct reference to the stress-driven production scenarios, as well as a near-real-time analysis of the measurement data of the different sensors, are development goals in this project. Based on the requirements from the application scenarios and the results of the sensor evaluation, additional sensor systems are selected as necessary extensions to the smart shirt sensor technology. Sensor extensions for the smart Shirt solution will be developed as part of the project development. This integration of additional sensor technology and the necessary communication modules enables better usability in industrial deployment scenarios, especially in combination with eye-tracking glasses and smartwatches or smart wristbands.

Due to its mobile application, the QUS biosensor shirt QUS (see Figure 6) is well suited for monitoring vital parameters during physically as well as cognitively demanding activities. The integrated sensors record the biosensor data and forward it to the onboard unit (OBU) via woven conductor paths. The OBU, in turn expands the signal set with acceleration data as well as – if needed - the current GPS position and transmits all data via Bluetooth Low Energy (BLE) to a nearby relay station such as a smartphone.

The OBU can be easily attached to and detached from the shirt with four snap buttons (see Figure 6c) and put into operation with a press on a button. The shirt has good wearing comfort and is available in different sizes for both women and men.

The following key facts are relevant to the functioning of technology in everyday activities:

- **Woven-in sensor technology**: the sensor technology is totally unobtrusive because it is woven into the textile and does not provide bulky parts. Therefore it is highly usable in long-term measurements.
- **On-board unit stores vital and geo data**: All sensors are integrated into the shirt package.
- **4GB data memory**: The storage unit is capable to archive long-term recordings.
- **Lightweight and washable**: in contrast to other textile technologies, QUS shirts can be washed many times before the sensors would decrease in precision.
- **Battery life** with GPS 5 hours, without GPS 24 hours: this technology enables long-term measurements at the workplace as well.

- **Bluetooth connectivity** (BLE 4.0): there is a continuous sensor stream in the vicinity of the server unit that enables real-time computation of human-in-the-loop functionality.

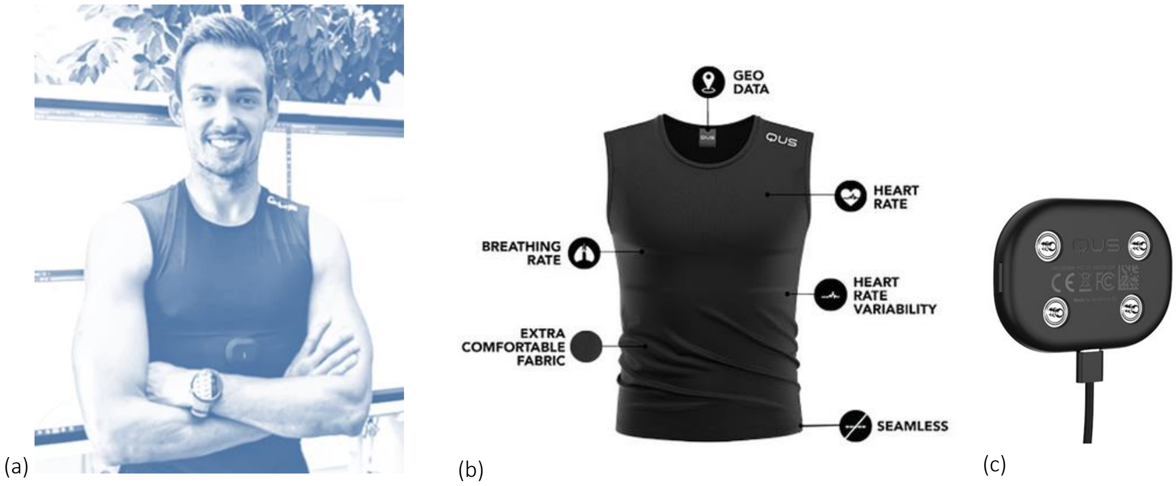


Figure 6: (a) QUS biosensor shirt (credits: JOANNEUM RESEARCH). (b) Signal overview (credits: sanSirro) and (c) on-board unit (OBU) backside with snaps for easy mounting (right; credits: sanSirro).

The QUS shirt records the following raw sensor data with the respective sampling rate being given in parentheses,

- Electrocardiographic data (ECG; 200 Hz) for psychophysiological strain computing.
- Respiration raw signal (100 Hz) for stress computing.
- Acceleration x/y/z (100 Hz) for pattern recognition from physical activities.
- Gyroscope (100 Hz) for continuous motion computations.
- GNSS (10 Hz) for precise localization of geo-data.

The OBU computes specific psycho-physiologically relevant parameters and general dynamics characteristics derived from the raw sensor data. This type of processing takes place directly on the OBU.

- Heart Rate (HR),
- Heart Rate Variability (HRV,
- Respiration Rate (RR),
- Distance,
- Speed,
- Dynamic load,
- Acceleration (GNSS).

Raw data and the derived parameters are finally transferred to the data lake of the Intelligent Sensor Box via web connection (REST interface). The data lake is a non-relational database based on mongoDB.

Psychophysiological Information Processing from the Biosensor Wrist Band

Wrist bands are equipped to provide biosensor-based data without being obtrusive. In FAIRWork we will apply the Garmin vívosmart 5 fitness wrist band. There is ample experience in using this wrist band and it provides a multitude of multisensory data about the human behaviour.

Figure 7: Wearable Garmin vívosmart 5 wrist fitness tracker (credits: NextPit).

The vívosmart 5 fitness tracker from Garmin (see Figure 7), which is worn on the wrist, records various biosignal data during physical activity. The device has a touchscreen and waterproof housing and, thus, can also be worn during activities such as showering or swimming. The band is available in three sizes (small, medium, and large). The following list provides more important features for everyday use, as follows,

- Weight: 24,5 g (small(medium), 26,5 g (large)
- Waterproof (until 5 ATM)
- Battery life 7 days (smartwatch mode)
- Connectivity: Bluetooth Smart and ANT+

The wearable device provides the following sensor data:

- Heart rate (HR),
- Beat-to-beat intervals (BBI),
- Heart rate variability (HRV),
- Blood oxygen saturation (SpO_2),
- Respiration rate,
- Number of steps,

as well as further parameters that are most probably not required in FAIRWork from the current perspective.

The software platform Fitrockr® is used to synchronise the wristband to get access to high-resolution psychophysiological raw data. Fitrockr also allows independent and secure hosting, survey and questionnaire responses, real-time data streams, health data reports and exports.

Oculomotor-based Mental Information Processing from the Eye Tracking Glasses

Eye-tracking glasses enable to follow the natural gaze and attention of their users to understand how persons behave in industrial settings. However, apart from following the gaze into the geometry of the environment, research has been concerned with the statistics of eye movement features, such as saccade lengths and fixation durations, during the free viewing of natural images. Bahill, Adler and Stark (1975) [146] examined the statistics of saccade length while participants walked outdoors wearing a mobile eye tracker, but there has been no subsequent effort to replicate these findings with modern eye trackers, with more statistical analyses and in a wider range of natural tasks. Bulling et al. (2011) [147] investigated eye movement analysis as a new sensing modality for activity recognition. Eye movement data were recorded using an electrooculography (EOG) system and recognised activity classes, such as reading a printed paper, taking handwritten notes, watching a video, etc.

Human-machine interaction in manufacturing has recently experienced an emergence of innovative technologies for intelligent assistance (Saucedo-Martínez et al., 2017 [148]). Human factors are crucial in Industry 4.0, enabling humans and machines to work side by side as collaborators. Human-related variables are essential for evaluating human interaction metrics (Steinfeld et al., 2006[149]). To work seamlessly and efficiently with their human counterparts, complex manufacturing work cells, such as for assembly, must similarly rely on measurements to predict the human worker's behaviour, cognitive and affective state, task-specific actions and intent to plan their actions. A typical application is an anticipatory control with human-in-the-loop architecture (Huang and Mutlu, 2016[150]) to enable robots to proactively perform task actions based on observed gaze patterns to anticipate the actions of their human partners according to their predictions. Paletta et al. (2019)[151] introduce several novel methodologies for the assessment of gaze-based human factors, which provides the potential to measure indicators of executive functions performance, such as attention selection and inhibition, as well as task-switching ability, in real time, illustrating the potential of interpretation from gaze-based human factors data to evaluate a complex human-robot collaborative system.

Essentially, high-end eye trackers direct near-infrared light towards the centre of the eyes (pupil), causing detectable reflections in both the pupil and the cornea (the outermost optical element of the eye). These reflections – the vector between the cornea and the pupil – are tracked by an infrared camera. In the case of eye-tracking glasses, near-infrared cameras are placed around the glasses' lenses, either on the frame or directly embedded in the lenses.

The Pupil Labs Neon technology (see Figure 8) has two eye cameras, one for each eye, running at 200 Hz. The cameras are optimally located near the nose to ensure maximum visibility of the eyes across all subjects and minimise occlusions. Access to all the raw, intermediate, and final data is always available - no special licenses are required. The scene camera has a wide field of view with 132 x 81 degrees, covering the full active field of vision. It captures full HD video with a resolution of 1600 x 1200 px. Gaze data is available at 200 Hz as 2D points in the scene camera video or as a 3D gaze direction originating in the scene camera. The integrated 9-DOF IMU allows tracking head movements at 220 Hz and estimating the absolute pose of the head, including its pitch, roll, and yaw. Fixation detection is automatically applied in the Pupil Cloud using an algorithm designed for head-mounted eye trackers. The algorithm compensates for head movements, providing higher robustness in dynamic settings compared to other algorithms.

(a) Credit: JOANNEUM RESEARCH DIGITAL / L. Paletta

(b) Credit: Pupil Labs GmbH

(c) Credit: Pupil Labs GmbH

Figure 8: Pupil Labs eye tracking glasses for the analytics of mental health in eye movement data.

3.3.1.2 Sensors Mounted at Workplace

Low-cost ambient biosensors are the basis for enabling digital living spaces at the workplace and, thus, are a necessary part of producing digital data in a sociotechnical and ecological system. Biosensors can be either wireless or wired and can be deployed throughout the infrastructure, such as buildings (Manic et al., 2016[152]; Tushar et al., 2018[153]; Jung & Jazizadeh, 2019[154]).

We are interested in supporting ideal indoor conditions (Erickson et al., 2011[155]) as suitable indoor conditions are essential in our living spaces where cognitive and physical activities take place (Zeiner et al., 2021[156]). These conditions are, for example, the composition of temperature, carbon dioxide, and humidity. These properties can be adjusted appropriately by humans, actively by controlling some devices, e.g. opening and closing doors, specific local heat and cold temperature workplace conditions, using ventilation and lights, or passively, e.g., by breathing. Typical recommended temperature ranges are between 20-23°C. This also depends on personal preferences. Based on common recommendations, carbon dioxide levels should be below 1000 ppm and humidity levels between 30 % and 60 % to ensure adequate indoor air quality (IAQ). The term IAQ is defined as the quality of the air inside buildings (Wolkoff, 2018[157]). This indoor air quality is an important factor for our health at the workplace (Tanir & Mete, 2022[158]). To set up an ideal indoor sensor, we need the right environmental biosensors, such as CO2 biosensors, light, temperature or humidity biosensors. Every person emits CO_2 when breathing, for example, and by monitoring changes in the detected concentration, occupancy can be inferred (Jiang et al., 2016[159]). However, air exchange, e.g. opening a window or weather conditions such as humidity, etc., have a decisive influence on the feasibility of this approach.

In Chen et al. (2018)[160], strategies are reported on combining more than one technology. If we combine sensor approaches, the chosen approach with data spaces is, therefore, particularly relevant and certainly has great potential in this application.

3.3.2 Intelligent Sensor Boxes Representing Cognitive, Affective and Motivational Human State

3.3.2.1 System Architecture

The FAIRWork Intelligent Sensor Boxes collect, process and analyse the incoming stream of digital data from the various sensor networks and provide information about all psychophysiological aspects of physiological and cognitive strain.

Figure 9 depicts the system architecture of the proposed Intelligent Sensor Box. The Local Wearable Sensor Network collects bio-signals from the wearables that are attached to the workers and the Local Workplace Sensor Network records. The gathered data streams are stored in the Local Database with the aid of the Local Data Receiver and Local Data Management components. The Local Decision Support System (DSS) generates higher abstractions of human factors and accesses Human Factors Intelligent Services for this purpose. The Human Factors Intelligent Services are also used by the user interfaces providing a worker which his/her human factors state. Additionally, the Intelligent Sensor Box has a Developer Dashboard to display various biosignals, human Factors parameters and the current system state. The higher-level DAI-DSS Knowledge Base (see Deliverable 4.1) can access all necessary data from the Intelligent Sensor Box via the dedicated Data Lake Streaming Incoming Interface. To ensure data protection, all provided data is anonymised by the Data Anonymization component.

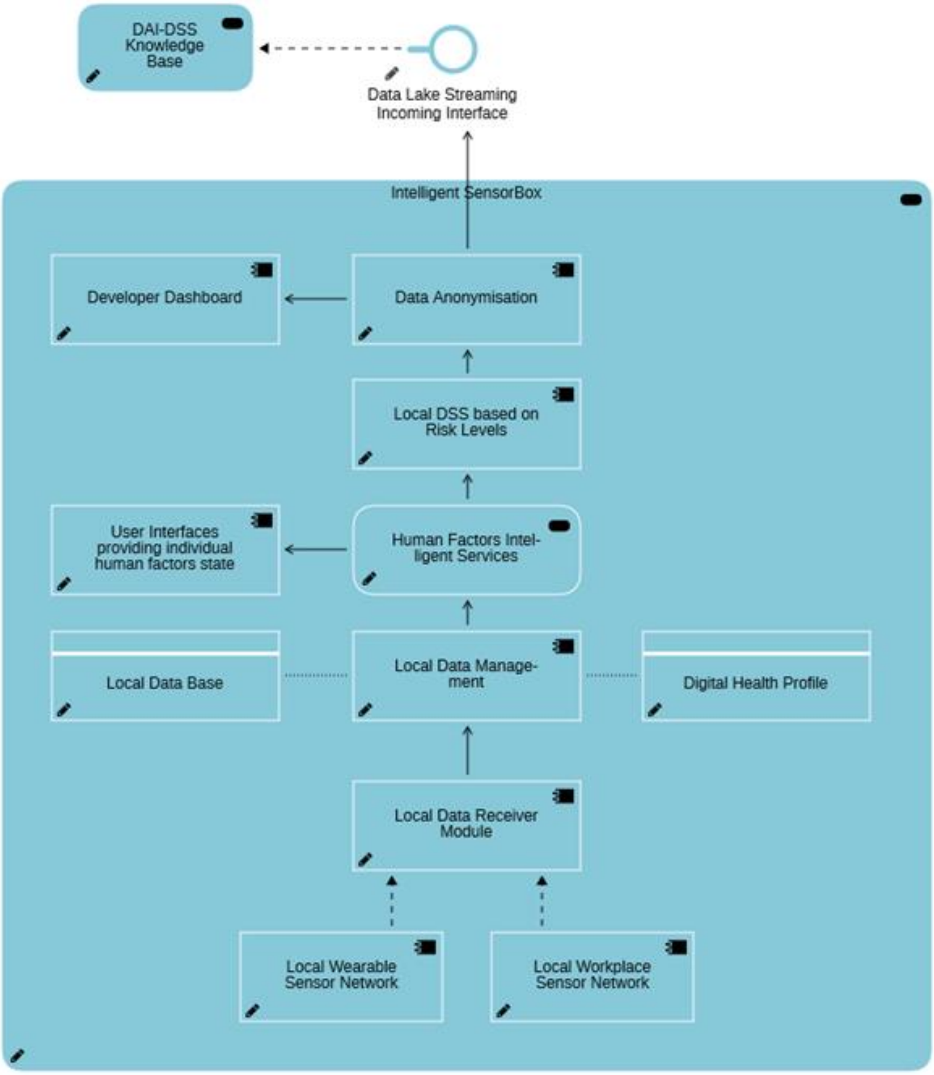


Figure 9: System Architecture of DAI-DSS Intelligent Sensor Box.

The various components of the Intelligent Sensor Box are described as follows,

Local wearable Sensor Network: Consists of various *human-mounted* sensors (via QUS biosensor shirt or wrist-worn biosensors) that collect biosignals like Heart Rate, Heart Rate Variability, Skin Temperature or Respiration Rate.

Local Workplace Sensor Network: A network that consists of various connected sensors that collect *environmental* data like temperature, humidity, barometric pressure or noise level.

Data Receiver Module: Provides an API to the sensor networks comprising listeners for biosignal and environmental data. The Module forwards the received data to the local database by accessing methods of the Local Data Management.

Local Data Management: Provides all necessary access methods for the Local Database.

Local Data Base: Non-relational data base on the basis of a mongoDB. The database is structured in session-related tables, storing all biosignals and environmental data captured from dedicated sensor networks.

Digital Health Profile: User Profile containing metadata like age, HRmax, HR0, etc., necessary for PSI* calculation and other scores determined by DSS.

Data Anonymisation: Ensures the anonymization of all captured human-centred data and the determined highly vulnerable psychophysiological data.

Developer Dashboard: (see Figure 10): The dashboard offers extensive analysis and debugging possibilities supporting model development and implementation. For example, already recorded data can be replayed at different speeds, and different scenarios can be run through. The dashboard can also be used in the final system as a live viewer or analysis monitor. The dashboard retrieves pseudonymous data from the local database using the internal REST interface or via a web socket-based interface for streaming data.

Local DSS: This core analysis module takes the collected biosignals and environmental data and computes dominantly ergonomic and psychophysiological constructs that determine the mental state and behaviours of the human and, from this, the sociotechnical systems within the production process.

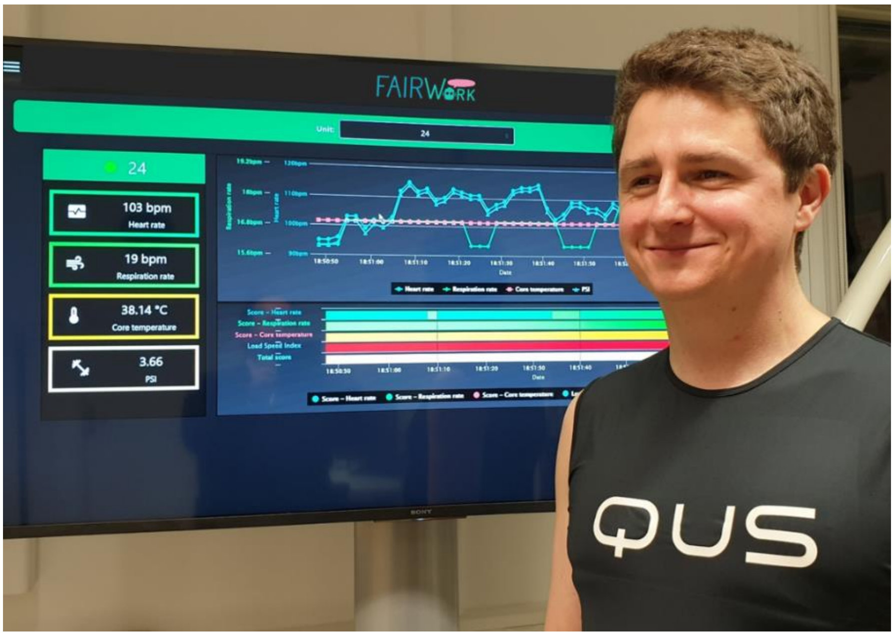


Figure 10: The Developer Dashboard of FAIRWork will enable to follow raw data but also metadata such as, physiological strain-based or cognitive-emotional stress-related scores. The objective is to enable real-time monitoring of the data as the stream of data from the biosensor short, the eye-tracking glasses, and the wrist-worn band would evolve.

3.3.3 Digital Human Factors and Risk Levels for Decision Support

The Local DSS of the Intelligent Sensor Box (see Section. 3.3.2.1) is represented by the "Human Factors Analytics DSS" (see Figure 12). It consists of three sub-systems:

- "**PS-DSS**" includes the physiological strain model and provides risk levels to prevent physiological overcharge and breakdown,
- "**SAW-DSS**" focuses on cognitive readiness with respect to the monitoring and reporting capability, i.e., the situation awareness of the decision-maker, and, finally,
- "**CR-DSS**" is applied to computer the score for cognitive-emotional stress and will feed forward to the risk level that is related to stress-based deficits of reasoning and planning, as represented in executive functions (Diamond, 2013[161]): "executive function" refers to a set of skills that underlie the capacity to plan ahead and meet goals, display self-control, follow multiple-step directions even when interrupted, and stay focused despite distractions, among others.

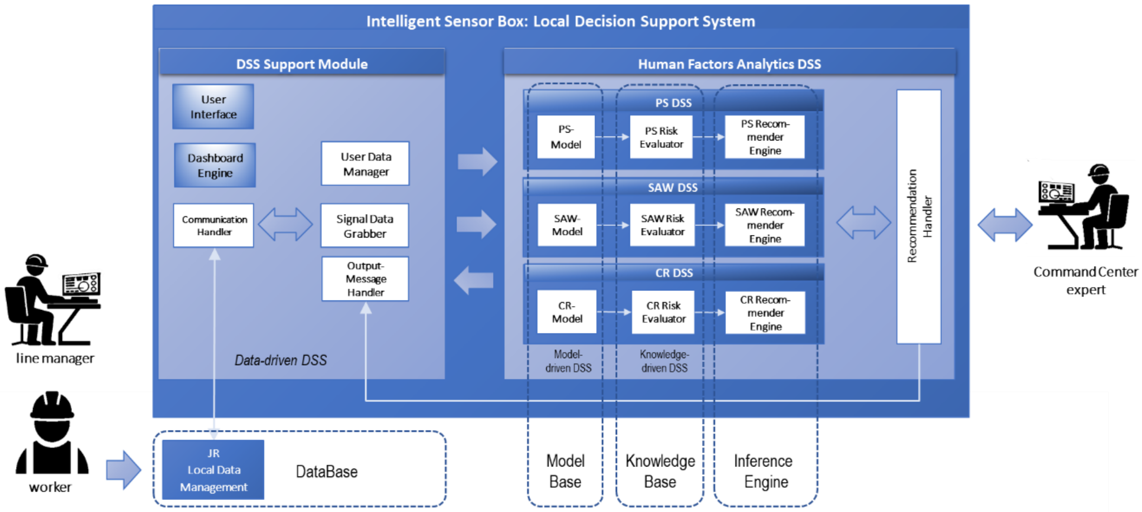


Figure 11: Schema of the Local Decision Support System of the Intelligent Sensor Box.

In the following, a detailed view of the software components of the PS-DSS is given. Figure 11 depicts the Local Decision Support System (DSS) of the Intelligent Sensor Box with the following main components, as follows,

- **DSS Support Module**: Provides all necessary auxiliary functions for communication with the Local Data Management (LDM). It consists of the following submodules:
  - **Communication Handler**: A software class that provides methods for reading of data from, writing data to and listening for newly added data in the Local Data Management (LDM). All data is transferred over an SSL/TLS secured connection.
  - **User Data Manager**: Handles user metadata like health profile data necessary for computing the physiological strain scores determined by DSS. The User Data Manager uses the Communication Manager for communication with LDM. There is a separate user table for each session (field trial, laboratory test, etc.).
  - **Signal Data Grabber**: Provides listeners with bio-signal data from a user recently received by the LDM and forwards data to decision modules like PS-DSS. The Data Grabber uses Communication Manager for communication with LDM.
  - **Output Message Handler**: Generates predefined DSS Output messages and DSS Recommendation messages in order to update corresponding tables in LDM. The Output Message Handler uses the functionality provided by the Communication Handler to update tables in the LDM.
- **Human Factors Decision Support System (HFA-DSS)**: Consists of DSS subsystems for physiological strain, situational awareness, and cognitive readiness and a recommendation handler that coordinates all recommendations. Each sub-system consists of a dedicated model, a risk evaluator, a recommender engine. The **PS-DSS subsystem,** e.g. consists of
  - **PS-Model**: Takes the recorded bio-signals and calculates the physiological strain score proposed by JR.
  - **Risk-Evaluator**: Defines risk zones and classifies the physical strain according to that zones and a predefined rule set proposed by JR.
  - **Recommender Engine**: This module determines rule-based recommendations neatly defined in advance.

- **Recommendation Handler**: Keeps track of all DSS subsystem recommendations and coordinates them.

Complex decision-making processes often involve significant risks. These risks can stem from a wide range of sources, including operational disruptions, supply chain uncertainties, machinery malfunctions, variability in material quality, and unpredictable downtime for maintenance. Understanding and effectively managing these risks is crucial for maintaining operational efficiency and ensuring long-term production sustainability. One promising approach to facilitating complex risk management processes is through the use of MAS, which offers advanced capabilities for distributed decision-making, robustness against system failures, dynamic adaptability to changing environments, and collaborative problem-solving.

From a Multi-Agent perspective, risk levels in the manufacturing industry can be viewed as a collective concern that requires coordinated efforts of multiple decision-making entities. Each agent, representing different stakeholders of the production process, can help evaluate and manage risks related to the tasks they are associated with. Through communication and coordination mechanisms, agents can assess risk, share their evaluations, negotiate over risk mitigation strategies, and collectively make decisions that balance risk and reward across the entire operation. MAS's inherent flexibility and adaptability make it well-suited for dealing with the dynamic and uncertain nature of risk in the industry.

Collaborative Filtering (CF) was one of the first techniques applied to decision support that is still applied today with improvements. The concept involves collecting data related to similar users and applying their preferences to assist decision-making. From this, trust can be estimated by the accuracy of predictions that have been made over some time (O'Donovan, 2005[162]), increasing the overall robustness of the system by making decisions that were statistically in line with what one has already made in the past. Trust can then be deduced from choosing similar entities that have already made decisions within the scope of those to be made. In Cheng et al. (2021)[163], MAS trustworthiness is quantified and monitored by a manager agent responsible for tracking and calculating the agent's trust levels based on the Subjective Logic approach. It provides a formal representation of reasoning on the trustworthiness of agents' uncertainty related to their actions and observations. Agents can develop beliefs on trustworthiness by deriving confidence and reputation by assessing past interactions with the agent in question and the agent's society past experiences with that agent (Ramchurn et al., 2004[164]). Trust can be modelled and used for influencing negotiations and improving the quality of agreements, hence diminishing risks associated with the parties involved. It is a factor that delineates agent interaction and ensures reliable collaboration, further reducing risk in complex decision-making scenarios.

MAS incorporate various trust and reputation models that utilize different metrics and approaches to calculate trustworthiness, detect fraud, and manage reputations. The following describes some exemplary models found in the literature. The ReGreT model uses direct experiences, third-party information, and social structures to assess trust, reputation, and credibility. The model postulates that trust is calculated exclusively from direct experiences and reputation and uses a credibility module to assess the veracity of witness information (Sabater et al., 2001[165]). Complementary to ReGreT, the Yu and Singh model store the result of direct interactions as Quality of Service (QoS). It deals with the problem of deception, focusing on distinguishing reliable witnesses from deceptive ones (Yu et al., 2003[166]). Following, to handle reputation management and fraud detection, Muller and Vercouter propose a model that applies normative language to formalize prohibited situations, relying on reputation formed from direct experiences and observations to identify lies (Muller et al., 2005[167]). Sierra & Debenham (2005)[168] propose another model that uses information-based trust calculations for agents in negotiation processes. The trust measure is computed as the entropy of the distribution, indicating the probability of an agent accepting a proposition.

Lastly, the Repage model supports agent architectures differentiating between image and reputation based on social cognitive theory. It calculates image and reputation by aggregating contracts, fulfillments, direct experiences, and third-party communications. This model shows that image and reputation, though both are social evaluations, serve distinct roles within a given context (Sabater-Mir et al., 2006[169]). Also, a mechanism that provides a way to ensure trust in the system through a biological paradigm approach is a possible route to address the problem of trust and risk. In this case, the agents' interaction is based on animals that use pheromones. These animals release chemical odours in order to signal to their peers their perceptions about the environment, especially of risks that may threaten their species. The agents propagate the risks of disruption to the control system in which they are located in order to signal to their peers the increased possibility of the system's state moving away from the desired working state (Pereira et al., 2013[170]). In summary, different approaches regarding trust in Multi-Agent Systems have been tested and applied over the years and offer possible paths for ensuring the management of risks regarding the stakeholders' negotiations.

3.3.4 Personas and Human Digital Twins for Decision Making

A manager, as well as a worker, generates Human Factors relevant data that are then converted into cognitive, affective and motivational features, and related to data that are relevant to performance optimisation, in the context of specific activities, be it tasks with a specific machine, or tasks during human-robot interaction.

These experiences are then clustered with an unsupervised strategy to develop evidence regarding data clusters and populations. They are called personas that provide specific characteristics – for example, old, experienced people that should not be stressed too much or young people that could provide maximum performance on the short term - in the overall distribution.

The Human Digital Twin finally is represented by a decision-making framework that is based on that persona information, and the results can finally retroact to these persons in a feedback loop (see Figure 12).

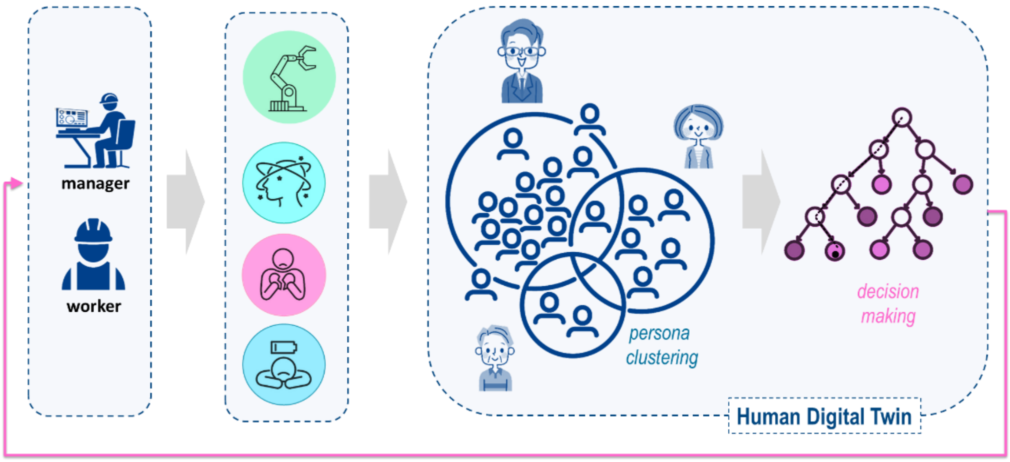


Figure 12: Sketch of the adaptive DAI-DSS persona framework (Paletta et al., 2023). Digital Human Factors data are captured and brought into context with the workplace environment. The accumulation of individual behaviour profiles is then clustered into characteristic schemata that represent certain typical "persona" behaviours. Decision-making structures on these representations are then applied to provide output data of the Persona Digital Twin and feed back to all individual actors through overall human-system interaction.

3.4 Investigation of Mapping of Decision Process Models with Decision Support Offerings

Based on design thinking workshops with use case partners, the use case scenarios and corresponding business processes have been defined (see Deliverable D2.1). Business processes give a broad overview of the challenges and current procedures in each use case scenario. However, they need to be further investigated to work out how these processes can be supported. Therefore, these business processes have been further detailed by conducting interviews with experts to get closer insights into the ongoing decision-making routines, as complex decision-making is the aspect that should be supported in the FAIRWork project. Models can be used to externalize and represent the knowledge of all stakeholders involved in the decision-making. It is important to identify the parts and aspects of the business process which are of interest for the project, more concretely in our case, to identify decision points in each process, which have a certain complexity and potential to be supported by services. The result of refining the business processes and highlighting decision-making aspects are decision models, which are further described in Section 4.1.4 of this document. One primary purpose of the decision-making models in FAIRWork is to identify challenges for the decision-makers within the model that can be supported by (AI) services.

Decisions can be based on different problems the responsible parties are trying to solve in the decision-making process. In the case of FAIRWork, these identified problems have been abstracted into conceptual challenges like resource mapping, solution configuration, and selection. In addition, the number and identity of decision-makers and stakeholders involved in the decision-making process differ in each use case. As we are dealing with decisions of various natures and complexity, the possibilities to support these decisions also vary. Decision support systems can be categorized into five different types, namely Model Driven DSSs, Data-driven DSSs, Communication driven DSSs, Document Driven DSSs, Knowledge-driven DSSs, and Relative DSSs, which highlights the facts that a decision can be supported in different ways (Felsberger et al., 2016; Psarommatis el al.,2022[171]). As an example, AI can support the accessing of data for a specific decision parameter; at the same time, it can enhance the parameters or tasks within the process. It might even be possible that the former sources of human input can be replaced through intelligent algorithms like image recognition or data analysis algorithms. Another way of supporting decision-makers is the possibility of running optimizations on top of the processes and thus producing scenarios that the responsible actors might consider. Considering the variety of support possibilities and different levels of complexity coming with decision-making, the decisions in the process may be taken by machines (fully automated), humans, or in a hybrid way.

Not only must the identification of support potentials of AI in the decision models be focused on, but it is also essential to consider which type of AI is more suitable for a specific decision environment—supposing that the decision models are very long and complex with many vague input parameters rarely known or impossible to retrieve from data sources or log files. Symbolic AI, like Fuzzy Logic, might be a better approach in that case. On the other hand, if log files are present in complex process settings, data-driven AI might provide better results than humans would be able to provide [172]. The decision support system includes an (AI) service catalogue with services based on different AI approaches (e.g., ML services, symbolic AI services, agent-based services, etc.). These services differ not only according to their AI approach or what aspect they try to support, but they also vary in form of their size, complexity or accessibility.

After the identification of business needs and requirements, which is, in our case, expressed through decision models, and, on the other side, the specification of capabilities and offerings by listing AI solution offerings, a mapping process is necessary to bring the needs and capabilities together. During this mapping process, the decision identified in the process and the available AI services in the catalogue should be matched to find a suitable support solution. The mapping process can be seen as a model-based alignment between the characteristics of the AI services with the use case-specific need. Characteristics of the services also include data requirement aspects, which need to be considered when searching for a suitable match. It needs to be checked if relevant data is available and if this data is public or confidential. The decision models represent the requested features and specify the requirements for each decision relevant to the specific process. If the information to specify all needed requirements cannot be directly derived from the decision model, annotations can be helpful to enrich the model semantically. The catalogue of application services for decision-making represents the available offerings and capabilities. The outcome of the mapping process between requirements and capabilities should be a preselection of possible support solutions, which can include individual services, a combination of services in the form of workflows, or MAS solutions.

Once the mapping is successful, the solutions need to be integrated into a legacy system. This step looks at how the services can be put into practice. During the implementation phase, the AI solutions need to be configurated and orchestrated to fit into the overall system. In the configuration phase, the input parameters of a service can be adjusted in order to meet the use case needs. This can also be used to specify data source requirements for the services. The configuration of a Multi-Agent System can include which agent types are used, which skills are assigned to each agent, or which behaviours each agent could have.

After a successful implementation, the task is not over, but the system needs continuous management. Business needs and support capabilities may change, as well as aspects of the legacy system (e.g. system architecture). We propose a model-based approach to support these management actions. Models have the capability to reduce complexity, make processes transparent and comprehensible and can be interpreted by humans and machines. It will be investigated if these characteristics could be beneficial for a management approach.

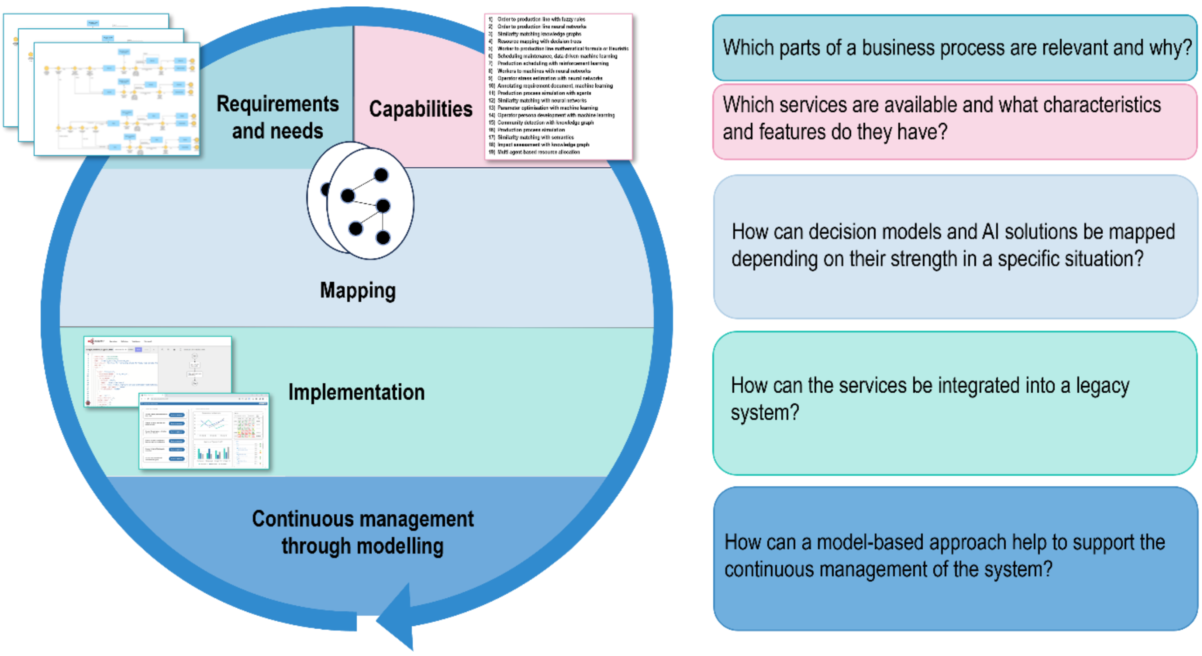


Figure 13: **Overview of model-based approach and corresponding research questions.**

In summary, the described approach consists of five conceptual building blocks connected to each other. In general, we have business needs and requirements, which are represented by decision models. On the other side, we have capabilities, which are present in the form of AI solutions. These two concepts need to be mapped to find offerings that fit the business requirements. In particular, this mapping phase involves the task of matching challenges in decision-making with suitable AI decision support solutions. After various matches are identified, which in our case are suitable AI offerings, they need to be integrated into a legacy system. This part can be seen as the implementation phase. The described process is not a one-time issue; ongoing management is required to maintain it. Here a model-based approach is investigated. Figure 13 gives a visual overview of the overall concept. The research and investigation of this approach and its conceptual building blocks are driven by different questions which should be examined during the project:

- Which parts of a business process are relevant and why?
- What are the capabilities of the available services, and what are their characteristics?
- How can decision models and AI solutions be mapped depending on their strength in a specific situation?
- How can suitable services and offerings be integrated into a legacy system?
- Is a model-based approach suitable to support the continuous management of a decision support system?

It is typical in the research area of information systems. A combination of research and development methods are used to address these questions [173]. In the following, the chosen approaches are presented in more detail.

Modelling can be seen as a method that creates simplified representations of reality. This can be done inductively if the model is based on observations or deductively if the model is created out of theories. Numerous variants can be distinguished regarding the types of models, but formal and conceptual (semi-formal) models are the most common in business informatics [174]. For our research focus, conceptual models are more relevant. Conceptual modelling allows one to examine a system from a conceptual perspective and to capture its most important structural, behavioural, or semantic features. This can be very beneficial when a current system should be examined and when creating a new one [175]. For conducting our research regarding the use and integration of AI algorithms and MAS in a DSS, modelling is used to externalize the knowledge of decision-makers and experts familiar with the use case scenario and depict the decision paths. Such models have the advantage that humans and machines can interpret them. So these models can further be used to support the mapping process between defined requirements and available AI solutions.

A prototype can be considered an executable model of the planned system currently being developed, which is created with minimum work, is simple to alter, and can be tested and assessed. When choosing prototyping as a method, the implementation process starts as early as possible, also if not all requirements are fixed yet [176]. This enables early feedback regarding the suitability of a solution approach and makes it possible to identify problems at an early stage. Rapid prototyping can be seen as a very explorative type of prototyping and is therefore used at an early stage to identify system requirements and evaluate solution approaches. At this point, the system's functionality is the primary concern. These characteristics make the method a good fit for investigating the integration into a legacy system.

3.5 Exploratory Studies within Laboratories

Experimental laboratories play an important role in research by enhancing the understanding of various technologies and shedding light on the decision-making processes behind them. The laboratory environments available in FAIRWork serve as platforms where experiments and innovative approaches are conducted to test new ideas. Researchers can develop novel approaches within laboratories to examine how decisions are formulated and represented. This allows for a more inclusive and transparent decision-making process, ensuring all stakeholders understand the decision path and the technology behind it.

3.5.1 OMiLAB Experiment Environment

OMiLAB uses and offers the *Digital Innovation Environment*[177], which uses model-based experiments to create and evaluate designs for complex systems. The experiment environment is based on three layers[178]:

- The Business Value Layer,
- Agile Digital Twin Layer and,
- Digital Experimentation Layer.

Depending on the concrete application case, the layers can be utilized to different degrees, meaning that some experiments rely more on one layer than others.

Each layer has its own focus, which is used within the experiments, and the layers combined represent the experiment. In the *Business Value Layer,* concrete scenarios are established which capture the important aspects of the system that should be created, for example, the problems, involved stakeholders or entities, constraints etc. The created scenarios are captured on a high abstraction level to facilitate understanding and communication of the problem and possible solutions with the stakeholders of the experiment with different expertise. Additionally, the scenario will provide a frame against which the experiments will be evaluated to analyse if the wanted solution is found.

For FAIRWork, one scenario could be how an information system can support the rearrangement of the production plan if workers are missing, including the humans, machines and organisational steps that must be checked and done to get a feasible alternative.

The conducting of an experiment needs an environment where it can be run, which is created in the *Digital Experimentation Layer* in the context of OMiLAB's experiment environment. Therefore, a prototype of the real or future system is created, which contains *Cyber-Physical Systems (CPS)* and/or purely digital components which represent the real system. The intention is that the important entities and their needed capabilities and characteristics are captured.

The *Agile Digital Twin Layer* consists of digital, conceptual models, which are the core of OMiLAB's experiment approach. They capture the information from the other two layers to enrich and combine this knowledge, getting a full view of the system for the experiment. The conceptual models, created in a modelling tool, will then be linked to the prototype of the *Digital Experimentation Layer* and directly executed. Therefore, the knowledge of the models must not be implemented in the prototype but can be used directly from the models. Through this, the knowledge from the experiment execution and the capabilities of the prototype environment are decoupled.

The experiments can vary in their size by focusing either on the system, which should be created or analysed, as a whole or parts of it. The design of an experiment corresponds to the questions and objects which should be answered and reached with it. The insights gathered from the execution of an experiment can influence the scenario, the models and/or the prototype and trigger improvements. Based on the decoupling of the experiment knowledge in the models and the prototype, both can be adapted independently and reused for other experiments.

To ease the creation and execution of model-based experiments, OMiLAB provides and uses free to use tools, which can be used out-of-the-box or adapted to concrete experiments. The ADOxx metamodeling platform (www.adoxx.org) is used for the creation of domain-specific modelling tools, which are used in all three layers, but are especially important for the middle layer. For creating and defining the scenario, the ADOxx-based Scene2Model[179] tool is offered by OMiLAB, which supports physical, creative workshops and an automated transformation of the physical artefacts in a digital model to establish the scenarios. For the models in the middle layer, dedicated modelling methods and tools should be used and created to capture the experiment knowledge.

The prototypes of the *Digital Experimentation Layer* must be created and tailored to the experiment but must be integrable with the models.

Within the research of the FAIRWork project, OMiLAB's model-based experimentation environment can be used to design, prototype and evaluate the AI services and approaches to tackle the complex decision problems which lay at the core of FAIRWork. The logic of the decision can be saved in an understandable model, which can be discussed with stakeholders, which are no experts in cocreate research topic for which the experiment was created. This is additional fitting, as conceptual modelling will also be used for the DAI-DSS system itself, which can lead to synergies between the conceptual models from the experiments and those created for the DAI-DSS.

3.5.2 Human Factors Lab for Digital Behavior Analysis

The Human Factors Lab (Figure n-n) at Joanneum Research, Austria, combines state-of-the-art human-centred measurement technologies with AI-enabled software for behaviour-based analytics and assessment of psychological constructs in digital systems.

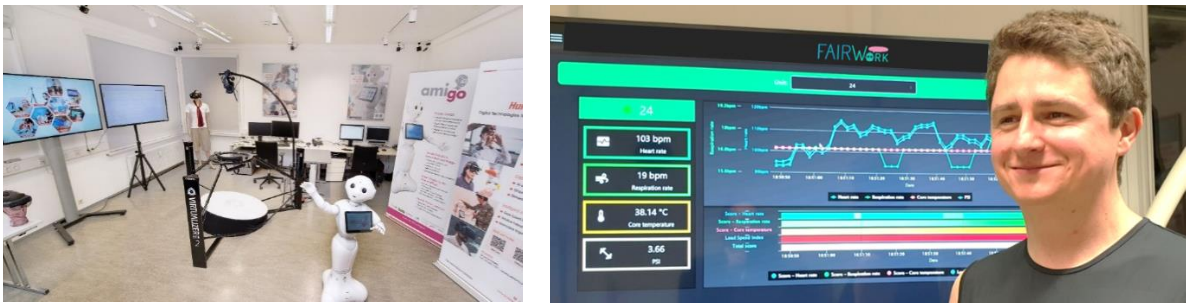
Dr Lucas Paletta is heading the Human Factors Lab and the research team "Cognitive Sensing and Interaction" that is associated with it, including experts on wearable biosensors, eye tracking analytics, VR/AR/XR immersive environments, AI-based interpretation of human-based sensor data, and the development of digital biomarkers for cognitive and mental assessment.

In the context of cognitive ergonomics at work, the Tools of the Lab provide metrics for the impact of stressors on the cognitive, affective, and motivational human state, in particular, by means of wearable technologies.

Lab infrastructure. The infrastructure is equipped to investigate psychologically relevant parameters from digital phenotypes, emphasising real-time analytics. It includes a wearable biosensor and Neuroergonomics (fNIRS, EEG) workplace, another workplace for immersive training environments with VR/AR/XR technologies, high-precision eye tracking in stationary, wearable and mobile conditions, high-precision motion capture devices, a VR locomotion platform for simulated orientation in artificial space, a precision treadmill to finetune physical strain, social robots to study social interaction, and software libraries for sensor-based, specifically, gaze-based analytics. The Human Factor Lab has access to various IoT software libraries for permanent data acquisition for sensors and gateways to set up wireless sensing networks.

The specific benefit of performing studies in the Human Factors Lab is to take advantage of a highly flexible technological and expert environment to apply rapid research prototyping. Using the VR/AR/XR facilities, it is possible to simulate arbitrary work conditions and enable natural movements in simulated environments. Multiple wearable, mobile and stationary sensor technologies empower to capture human behaviour in arbitrary conditions. The Human Factors Toolbox and Decision Support System analyse the human data and provide risk levels for various requirements on health status and cognitive readiness.

Targeted research. In the Human Factors Lab (see Figure 14), we will develop a novel approach to estimating physiological as well as cognitive strain. Cognitive-emotional workload will be studied related to task switching, multitasking and interruption as well as monotony effects, using wearable bio-sensor shirts, eye tracking glasses, smartwatches with biosensors, digital events and spatiotemporal patterns from human-machine interaction and in relation to typical stressors, such as time pressure, or environmental strain, such as insufficient air quality. Specific attention is dedicated to the development of the 'Human Digital Twin' (HDT), applying AI-enabled Human Factors measurement technology. Each instantiation of an HDT provides a vector of Human Factors state estimates – e.g., on stress, affective state, concentration, workload, situation awareness, fatigue, etc. – with the purpose of determining cost function parameters associated with typical (inter-)actions in the work environment.



(a) (b)

Figure 14: The Human Factors Lab and its role in FAIRWork. (a) The lab embeds various workspaces. (b) Using wearable sensor technologies, such as biosensor shirts, we will monitor psychophysiological parameters during an interaction.

3.5.3 Robotic Lab of Flex

Combining state-of-the-art equipment and infrastructure with Flex's broad range of topics and expertise, solutions for implementation in production are tested in the Robotic Lab of Flex. Within the on-site robotic laboratory, two key engineers are present for exchange and support. They play a role in facilitating communication and collaboration regarding the laboratory's operations and projects. In the laboratory, various robots and grippers are available for testing new ideas. The concept of this laboratory environment goes along with circular economy standards. These go beyond classic system integration and include innovative solutions for new approaches in production. The current focus is on implementing collaborative robotics solutions that are rapidly deployed and implemented in production.

The Robotic Lab of Flex brings the following benefits:

- Affordable Agile Automation,
- Support for the operator through,
- Automation of simple tasks,
- Focus on small and medium lot sizes,
- Quick configurability,
- Automation based on modules,
- A robot can use a mobile platform and landmark detection to perform multiple tasks.

3.5.4 CRF Lab

CRF (Centro Ricerche Fiat), headquartered in Orbassano (Turin) with other branch sites in Italy, was established in 1978. As a focal point for research activities of FCA (Fiat Chrysler Automobiles), CRF has the mission to:

- develop and transfer innovative powertrains, vehicle systems and features, materials, processes, and methodologies together with innovation expertise in order to improve the competitiveness of FCA products;
- represent FCA in European collaborative research programs, joining pre-competitive projects, and promoting networking actions;
- support FCA in the protection and enhancement of intellectual property.

Also, through cooperation with a pan-European network from industry and academia, CRF conducts collaborative research initiatives at the national and international levels in partnership with all the key public and private stakeholders concerned with sustainable mobility, targeting specifically the industrial exploitation of research.

4 KEY RESEARCH FACTORS IN FLEX AND CRF USE CASES

This section presents the key research factors relevant to the use cases addressed in our project. Through a series of discussions between research and industrial partners, the necessary resources to conduct research that focuses on the trustworthiness and democratization of intelligent technologies and the application of AI and MAS methodologies and creating DSS were identified. Defining the research factors is critical to creating a solid foundation for future research. The factors are divided into two blocks:

- technical perspective in use cases,
- human perspective in use cases.

From a technical perspective, four main research requirements were identified: the available and relevant data, DSS architecture, expert knowledge and decision models. These factors are important for modelling and testing new concepts using AI and MAS technologies. The human perspective factors were defined based on the research plan and can be observed in each scenario in the use cases. The goal is to integrate psychological and sociological perspectives. This requires close collaboration with use case partners to capture the viewpoints of individuals directly involved in decision-making as well as those indirectly affected by the decisions.

4.1 Technical Perspective in Use Cases

The exploration of the key factors from a technical perspective is essential to investigate selected used cases from FLEX and CRF effectively. This section focuses on the critical research factors contributing to the successful implementation of model, prototype and testing-oriented research. Based on FLEX and CRF use cases, the technical research factors, such as relevant data, DSS architecture, expert knowledge, and decision models, were identified (see Figure 15). These factors are essential for building accurate and reliable models and prototypes that utilize AI and MAS methodologies. Analyzing decision paths and exploring industry use cases provide valuable insights into decision-making processes within manufacturing environments, which are important in defining relevant data for further application and conducting research.

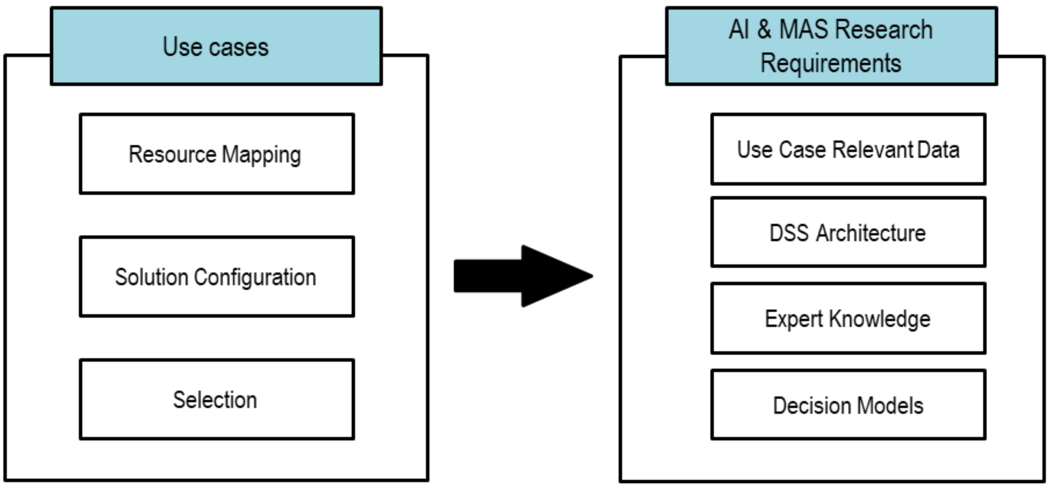


Figure 15: Key technical research factors in FLEX and CRF use cases.

4.1.1 Use Case Relevant Data

Accessibility to relevant data is of essential importance for an effective application of DSS enrichment strategies. The data needed for the execution of process improvement is not always easily accessible or available digitally. Digitizing aspects inherent to industrial production processes, such as tacit knowledge, is fundamental to satisfy the data requirement in many industrial cases. This involves converting implicit, experience-based knowledge into explicit, digitally accessible information. It enables not only the preservation and dissemination of critical skills and know-how but also facilitates AI and MAS-driven effectiveness and decision-making. The integration of such digitized knowledge into the DSS is crucial for reaching the desired objectives. The efficacy and accuracy of agents' operations, designed to interact, learn, and adapt, hinge significantly on the quality and accessibility of data they have at their disposal. Some techniques involving MAS with Machine Learning include strategies addressing data scarcity. These methods must be taken into account in real industrial scenarios. In conclusion, data is fundamentally necessary for MAS properly work in the desired condition.

4.1.2 Decision Support System Architecture

This section gives an overview of relevant concepts of the key components of DAI-DSS architecture, which applies AI services for complex decision-making. In FAIRWork, AI is used in all our scenarios to automate business processes or to make their processes more resource-efficient. Since humans are an important part of the overall decision process, trust in AI and human factors play an essential role, and therefore these aspects are explained in detail. In this part, we show essential techniques to support humans with AI. This is because humans cannot always comprehend complex and unstructured decision-making processes. This circumstance offers optimization potential by supporting human capabilities with AI technologies. The goal of using AI in a decision-making process is to speed up decision-making. Traceability is essential for establishing trust in the overall process.

The key component of the technical services architecture of the FAIRWork is presented in the next part.

Relevant core components for the complex decision-making services are the DAI-DSS AI Enrichment components, the DAI-DSS Orchestrator, the DAI-DSS Configurator and the DAI-DSS Knowledge Base. The DAI-DSS User Interface plays a significant part in the decision-making process because it should provide a clear picture of the alternatives and possibilities available to the decision-makers and provide a supportive environment.

4.1.2.1 AI Enrichment

The AI enrichment in DAI-DSS is a service which consists of various AI algorithms and models that contribute to decision-making. The main role of AI enrichment in the DAI-DSS environment is to provide a collection of data-driven models which mimic various resources (e.g. production lines, machines, humans) and consequently predict their behaviour in the dynamic industrial environment. The principle is to use explainable, traditional AI-algorithms as well as solutions that are more advanced for the representation of the resources as a service. Each of such models or optimisation algorithms contributes to the decision-making process bringing closer the decentralised and fair approach. In the DAI-DSS environment, AI Enrichment communicates with the DAI-DSS Knowledge Base via REST-API, which allows data to flow between the two services.

4.1.2.2 Orchestrator

The DAI-DSS Orchestrator is critical to the overall execution of the decision support system because it manages and coordinates the configured services and workflows needed to provide decision-making options for the end user. Services can be standalone or linked in the form of a workflow. While in the first case, the orchestrator is responsible for simply calling the individual service and retrieving the necessary data from the Knowledge Base, the second option is more complex due to the involvement of more services. The key part of the orchestrator is the Controller. This component allows for managing microservices and controlling their whole lifecycle. The orchestrator component depends on the DAI-DSS Configurator as it generates microservice instances based on configuration files provided via the Dataflow and Service Orchestration Interface. It also provides and receives data from the User Interface, as services can be triggered through user interaction, and results need to be displayed. In addition to that, it is connected to AI and non-AI service catalogues by the Service Interaction Proxy. If these services need input data, the orchestrator can access the Knowledge Base to retrieve it using the Data

4.1.2.3 Knowledge Base

The Knowledge Base is a central part. It acts as a centralized repository for the majority of data involved in a decision-making configuration. A decision-making process requires data from a wide range of different sources and produces results. The Knowledge Base is intended to collect all necessary data sets and streams, structure them, potentially merge them, store them, and then provide an interface layer which allows other DAI-DSS components to harvest said data. To enable AI algorithms to learn from previous decisions properly, and to enable processes to use earlier results, such decisions and results are stored in the Knowledge Base.

4.1.2.4 DAI-DSS Configurator

The DAI-DSS Configurator is an intermediate component which connects various model environments with the user interfaces and the orchestrator component of the DAI-DSS system and enables their model awareness. The configurator combines the different types of models and generates decision models out of it, containing the configuration information for the other components in the system.

4.1.3 Expert Knowledge

Although new technologies and AI leverage automation potentials in the manufacturing environment, the complexity of real-world use cases and the problem settings still require human intelligence and reasoning. Human experts working for FELX or CRF have acquired know-how and different problem-solving strategies for many years. Thus, expert input and experience are crucial preconditions for setting up a DAI-DSS, especially when unstructured processes characterized by uncertainty should be included. Complexity, uncertainties, and relations between many aspects shape most real-world use cases. Also, the size of the problem and its interconnectivity to other problem areas adds to the difficulty of conflicting goals. Therefore, a broader view and human reasoning are often required, which can be attributed to the human advantage. This advantage includes the identification of conflicts and choosing the option that can likely achieve the desired outcome. However, it means taking risks and potentially selecting the wrong option. One ability as part of the human advantage is the anticipatory behaviour done continuously and naturally by humans. It describes constant reflection, adjustment, and weighting of options to get closer to future goals[180] (Grieves, 2022). For this human aspect, all possible relevant pieces of information are reflected even with loose connections to the specific problem, such as macro-environmental aspects like customer behaviour or information and company or department particular preferences.

Until now, there is no such technology that can extract, simplify and illustrate the important parameters of a very specialized real-world problem setting similar to humans. Moreover, as there is explicit and implicit knowledge utilized for solving complex problem settings, it is impossible for externals to distinguish relevant and irrelevant information without this expertise. Also, FAIRWork depends on collaborating with a domain specialist to specify business processes and, in a second step, decision processes. Due to all the mentioned aspects, the requirement for expert knowledge for a DAI-DSS is highlighted as a critical research factor.

4.1.4 Decision Models

The decision-making process models are crucial for developing the DAI-DSS focused within the FAIRWork project. Based on the previously defined business processes of FLEX and CRF, the decision models aim to illustrate the issues and problem-solving mechanisms on a more detailed level, including all relevant decision parameters that must be considered to get to a solution. These relevant parameters are variables that must be known by the human expert or given by data sources. A single business process might involve multiple decision processes, while one single decision-making process again can include numerous sub-decisions.

To derive the relevant decision models from the business processes, three essential questions must be asked:

1. Who is the relevant decision-maker?
2. What should be the outcome of the specific decision-making process?
3. Which aspects and parameters must be considered to find a solution?

The first step was to examine the previously elaborated scenarios and business processes of the relevant actors. For the CRF "Workload Balance" scenario, the UTE Head or the Line Managers are perceived as relevant decision-makers. Quality personnel would be considered decision-makers in the "Quality Issues" scenario. For FLEX, the decision-makers in the "Automated Test Building" scenario are the automation engineers; for the "Machine Breakdown" scenario, it is the maintenance personnel.

The second step was to identify the outcomes or solutions that can emerge for each decision process, e.g., in "Workload Balance", the output is 1) allocate the worker or 2) do not allocate the worker to a line with specific geometry or in "Automated Test Building" the output options include: use a 1) collaborative, 2) cooperative or 3) industrial robot.

The third step is to analyze the path that leads to the different decision options. This path is characterized by many aspects or parameters that the responsible decision-makers consider to get to the corresponding output option. The parameters involved can be of many types, such as yes-no decisions or variables exceeding a specific value, i.e., the number of days until delivery.

In addition to the requested data from the use case partners, interviews and workshops with the domain experts contributed to approaching the modelling of the decisions graphically. For extracting their knowledge, multiple rounds of modelling and illustrating the processes and reviews are necessary until an acceptable decision process model is achieved. The CRF and FLEX decision processes relevant to the FAIRWork project are explained in more detail in Deliverable 5.1.

One main goal of the decision models is to provide a joint base for all types of readers, including domain experts, researchers, and IT specialists. The modelling aspect focuses on the easy and understandable preparation and illustration of information. The ADONIS NP tool[181] and the modelling language BPMN were used to model the decisions. Typical symbols used for the models include activities, exclusive decision gateways, and start and end points. One characteristic of decision models is the excessive use of exclusive decision items representing the decision parameters. These symbols express an either-or decision and can result in different output options. To get to an output option, minimum criteria or a sequence of simultaneously fulfilled decision aspects must be given. The decision parameters can come from different sources, such as sensors or legacy systems. It is also possible that human input is required due to a lack of digitized information or a too complex decision parameter.

The granularity of decision models is a central precondition to enable, on the one hand, a better understanding of the concrete problem setting. On the other hand, it eases the differentiation of available data and information as relevant or irrelevant. A third aspect that fosters the decision models' relevance is the support for easier identification of suitable (AI) services within the corresponding use cases. Additionally, it structures all relevant aspects of the decision into a sequence so that a workflow and an orientation for orchestrator technologies emerge. To sum up, the models are not only relevant for building the bridge from use case-specific issues to data but also for structuring decisions and workflows on which's basis an identification of suitable decision support technologies is promoted.

4.2 Human Perspective in Use Cases

The research on the use cases offers the unique opportunity to thoroughly analyze the respective socio-technical constellations which are framing the situations. Thereby, we aim to combine a psychological as well as a sociological perspective. For this purpose, there is a need for close cooperation with use case partners to gather the view of humans both directly involved or structuring the decision-making and also those indirectly touched by the decisions. With regard to this design, this can claim to be a co-design approach. This approach is methodologically inspired by mainly two strands of investigation, first practice sociology and the psychology of decision-making.

The actions of actors cannot be understood from a subject perspective alone; rather, the socio-technical arrangement as the context of such actions plays a central role. Therefore, the second perspective in the FAIRWork project, which focuses on people and human-technology interaction, takes a close look at precisely such socio-technical arrangements. These are investigated as practices in order to be able to analyze the interrelationship between individual action and structural action pattern schemes. To directly address this tension, this is of particular importance in highly technical contexts, such as the production lines studied. Such a study can be done by looking at socio-technical interaction situations within socio-technical arrangements. The concept of socio-technical arrangement allows studying meso/micro situations of social actors and technical actants, which is a determinant of a practice. Such arrangements operate in terms of a "structured and structuring structure" (Bourdieu 1982;[182] Giddens 1984[183]). This is, first of all, structured by the ensemble character, the relational orchestration of people as well as things, and embodies an institutional form. At the same time, however, it is not rigid; rather, it is obstinately appropriated by actors and thereby changed. Therefore, it makes sense to use an independent term for it because of its ensemble character. The concept of socio-technical arrangement allows one to depict the elements (from single technical objects to a multitude of human actors) that are put together in a selected action situation to form an ensemble. Thus, it is the joining of structured elements to a more or less complex structure that constitutes the socio-technical arrangement.

Human involvement in AI-supported decision-making has to be considered at many stages: When implementing rules or fuzzy rules, human knowledge needs to be abstracted. While this is a difficult task from a technological perspective, it is also a psychological question of how AI can be developed and implemented in a way that grants human acceptance and builds trust. In this regard, an important demand towards AI is transparency. Flyverbom emphasized that transparency is a “form of visibility management” (p. 110, 2016). Páez requires transparent AI to go beyond technical explainability and provide understandability to the end users (2019). Transparency applies both when developing AI as well as when executing and using AI for decision support. Decision trees or rule-based AI are often interpretable by humans and can thus (easier) be displayed in an understandable manner. Nevertheless, psychological studies will shed light on how their processes are set up and communicated in a way to support human acceptance. However, neural networks build black boxes that are neither interpretable for developers nor understandable for end users. The challenge is to develop explainable models – most likely in the form of post-hoc explanations – that allow for insight into the black boxes. Such insight is important on the one hand for developers to be able to improve the system and also for end users. The studies about the difference in transparency required for the user groups to adapt the system to their needs are planned.

Against this background, the following questions guide the empirical research:

- How are individual practices of decision-making constituted?
- What does successful AI-enriched decision-making look like for the involved human stakeholders in the respective use cases?
- What type of transparency do the different stakeholders require for which type of AI and which use case?

- How does transparency differ for the different AI stakeholders, and how can AI be designed to meet their requirements?
- How do different types of transparency foster trust, and how can the system be improved to increase acceptance and trust in these specific domains?
- What are the boundary conditions for practising specific affordances related to AI – and what are the limits of such an approach?
- What is the influence of the socio-technical context in democratic decision-making?

These questions will be answered empirically in two case studies, which will be conducted by the two industrial partners. These case studies will be conducted using the following empirical methods. On the one hand, the scenarios on which the respective work processes are based will be identified and mapped by means of document analysis. Based on this, interviews with employees will be conducted on the basis of two or three exemplary scenarios in order to find out the relevant expectations and forms of work performance. Finally, the sociological-technical arrangement of the production lines will be examined through participant observation. In the context of triangulation, the readings of the socio-technical interaction situation gained through the various empirical methods are brought into context and systematically evaluated. In this way, not only an empirically saturated analysis of practices is available. Rather, relevant factors for modelling the decision support tool can also be derived from this, mapped, and the most relevant factors for further use in the tool can be selected in a joint analysis with the other research and development partners.

The psychological studies will research the implementation of visibility and transparency in AI to grant the effective use and transfer of information. In FAIRWork, the goal of these studies is to use transparency as a means to increase acceptance and usage. First, qualitative analyses shed light on the perspective and requirements of different stakeholders by means of interviews. Based on these results, hypotheses about usage, and the effects of transparent AI systems in specific use cases and beyond can be formulated. Second, quantitative studies compare different ways of transparency and their effects on usage, acceptance and trust in the system, but also subjective control, contentedness with the working processes and subjective autonomy. These quantitative studies combine use-case independent AI research (e.g., Werz et al., 2020) and specific FAIRWork systems to be able to find answers for the systems at hand as well as more general solutions.

4.3 Research on DAI-DSS Services Catalogue in FLEX and CRF Use Cases

This section focuses on exploring the various aspects related to the utilization of the DAI-DSS services catalogue in FLEX and CRF use cases (see Table 3). By investigating the potential applications of these services in decision-making processes, this section aims to provide valuable insights into integrating the AI Service Catalogue into DAI-DSS functionalities within the FAIRWork project.

Table 3: Overview of the use cases with potential AI services as a solution

Use case	Suitable AI service
Resource Mapping	
FLEX Worker Allocation	Rule-Based Resource Allocation Service Pattern-Based Resource Allocation Service Multi-Agent based Resource Allocation Similarity Matching Knowledge Graph Resource Mapping with Decision Trees Community Detection with Knowledge Graph Similarity Matching with Semantics Optimization/Heuristic-Based Resource Allocation Service Operator Stress Estimation with Neural Networks Operator Persona Development with Machine Learning
CRF Workload Balance	Rule-Based Resource Allocation Service Pattern-Based Resource Allocation Service Reinforcement Learning Based Resource Allocation Service Multi-Agent based Resource Allocation Similarity Matching Knowledge Graph Resource Mapping with Decision Trees Community Detection with Knowledge Graph Similarity Matching with Semantics Optimization/Heuristic-Based Resource Allocation Service Operator Stress Estimation with Neural Networks Operator Persona Development with Machine Learning Order to Production Line with Fuzzy rules Order to Production Line with Neural Networks
Solution Configuration	
FLEX Automated Test Building	Rule-Based Resource Allocation Service Configuration Support Service Production Process Simulation with Agents Production Process Simulation
FLEX Machine Maintenance After Breakdown	Configuration Support Service Multi-Agent based Resource Allocation Production Process Simulation with Agents Production Process Simulation Impact Assessment with Knowledge Graphs Heuristic-based Delay Assessment Service
CRF Delay of material	Reinforcement Learning Based Resource Allocation Service Production Process Simulation with Agents Production Process Simulation Impact Assessment with Knowledge Graphs
CRF Quality Issue	Configuration Support Service
Selection	
"Selection" is applied to the results of the other decision support challenges.	

4.3.1 AI Services: Decision Support for “Resource Mapping”

Resource mapping often involves numerous criteria and constraints that must be considered. These criteria can include factors like operator workload, machine compatibility, and operator preferences. Incorporating such complex decision-making processes within the DSS requires a robust algorithm and methodologies. The “Resource Mapping” decision support challenges the following scenarios.

4.3.1.1 FLEX Worker Allocation

Service: Rule-Based Resource Allocation Service

The output of this AI solution is a list of potential candidates for replacing a missing colleague at a chosen machine. Based on predefined rules, each worker receives a score that would suggest their suitability for the given job. The rules can be defined based on the available information on workers' skills and preferences. This solution also allows the introduction of fuzziness for a better ability to deal with complicated decision issues related to human resources.

Likewise, evolutionary techniques like Genetic Algorithms (GA), Particle Swarm Optimization (PSO), and Ant Colony Optimization (ACO) could also be applied to this use case.

Service: Pattern-Based Resource Allocation Service

For this approach, the DAI-DSS Knowledge Base provides historical data on resource allocations. Assuming that the past allocations are optimal or at least good allocations, machine learning methods can be applied to generate allocations for newly emerging resource allocation problems. Technologies such as neural networks or decision tree learning can be employed to solve this task.

Service: Multi-Agent based Resource Allocation

MAS can provide a dynamic decision-making framework for the worker allocation process. When a worker misses their shift, the Multi-Agent system can analyze the pool of available workers, their respective training levels, compatibility with the tasks assigned to the absent worker and the priority in production plan execution in order to recommend rearrangement in worker allocation evaluating the suitability of an available worker taking over the missing worker's task. It also takes input from workers’ preferences and physical and mental conditions into account, following ethical guidelines to deliver fair and human-centred support in decisions. MAS offers a responsive solution ensuring productivity, minimises risks, and upholds quality standards in work allocation processes.

Service: Similarity Matching Knowledge Graph

The output of this solution is a suggestion of a person that can be assigned to a certain machine, e.g., to find a replacement for a missing colleague. The allocation should not be based on a one-to-one matching of people to possible machines but on the capabilities of the people in comparison to the requirements to fulfil the task at the machine. Therefore, the matching is based on the similarity between the capabilities of workers to the requirements of machines. For this approach, both the capabilities and the requirements for the tasks at a machine are described in a knowledge graph, which is then used to find provided capabilities that are similar to the task's requirements. The tasks on the machines and a list of workers must be provided as input.

Service: Resource Mapping with Decision Trees

The output of this solution is a list of workers who are allowed to work at a certain machine, e.g., as a replacement for a missing colleague. Like similarity matching with a knowledge graph, the requirements for working on a machine must be defined so that they can be matched to the capabilities of the work. For each type of task, a decision path is, in this case, described as a decision tree. Such a decision tree can support determining which worker can be allocated to the chosen machine.

Service: Community Detection with Knowledge Graph

The output of this solution is a list of workers similar to the provided one. Therefore, the workers and their characteristics are saved in a knowledge graph, and this structure is then used to find similar ones and defines them as one community. Such a community can then be applied to find workers with similar skills, which can be used to find a replacement for an absent worker.

Service: Similarity Matching with Semantics

The output of this solution is an estimate of how similar a defined capability of a worker is to the needed requirements of a task at a machine. This should be achieved by matching the labels in the description of the capabilities to the requirements so that they do not have to use the same label. This means the semantics of the labels is matched rather than the direct comparisons of strings. This solution can also be combined with other solutions where labels must be matched.

Service: Optimization/Heuristic-Based Resource Allocation Service

The output of this solution is the concrete mapping of workers to one task in the production line. This mapping is use case specific and based on the consideration of time constraints, constraints on the worker profiles (e.g. stress levels, etc.) and available time for making the decision. This approach yields efficient and timely decisions.

Service: Operator Stress Estimation with Neural Networks

Stress directly affects human decision-making and might lead to numerous undesirable consequences, including increased distraction and deficits in the person's decision-making capabilities. Therefore, the consideration of levels of cognitive-emotional stress is important to harmonize man-machine production processes. A neural network is defined to map specific vital parameters (such as heart rate, heart rate variability, breathing rate, eye movement features, etc.) to a score for cognitive-emotional stress and consequently to the communication of risk levels in the context of efficient decision-making. In this use case, the responsible decision-makers and workers are attributed with risk levels associated with certain challenging tasks, and this score can be integrated into the reasoning and optimisation process on the worker allocation.

Service: Operator Persona Development with Machine Learning

Personas represent a group or population of real human operators regarding their Human Factors features, cognitive, affective, social, and motivational parameters. Specifically, based on a set of human profiles, a clustering methodology, such as a Machine Learning (ML) driven stochastic process, determines a mathematical representation of characteristic human profiles that can be applied for holistic optimisation processes with the human-in-the-loop. In this use case, specific persona profiles would be recommended to optimally allocate workers to production processes and maximise economic and well-being-related objectives.

4.3.1.2 CRF Workload Balance

Service: Rule-Based Resource Allocation Service

The output of this AI solution is a list of potential candidates for replacing a missing colleague at a chosen machine. Based on predefined rules, each worker receives a score that would suggest their suitability for the given job. The rules can be defined based on the available information on workers' skills, age and preferences. In this use case, the rule-based approach can also be applied for allocating production of different items in the time and to the specific production line. This solution also allows the introduction of fuzziness for a better ability to deal with complicated decision issues related to human resources.

Likewise, evolutionary techniques like Genetic Algorithms (GA), Particle Swarm Optimization (PSO), and Ant Colony Optimization (ACO) could also be applied to this use case.

Service: Pattern-Based Resource Allocation Service

The only function of the Configuration Support Service is to provide a list of similar configurations to assist the user in a configuration task. This approach requires a vast amount of historical data of various workload setups based on which the most suitable one can be suggested for a new production demand as well as workers allocation to the production line. Technologies such as neural networks or decision tree learning can be employed to solve this task.

Service: Reinforcement Learning Based Resource Allocation Service

The use of AI in decision-making has become increasingly prevalent in recent years, particularly in the domain of resource allocation and scheduling. Resource mapping problems, such as flow shop and job shop scheduling, require allocating resources to tasks to minimize the completion time or maximize the utilization of resources. Several approaches to solving these problems include reinforcement learning, integer programming, and rule-based methods (heuristics).

Service: Multi-Agent based Resource Allocation

Similarly to the worker allocation use case, workload balance requires allocating workers to tasks, but this time is related to rearrangement regarding a change in production demand over time. Considering planning on a weekly basis, the agents can execute a resource allocation by accessing the production demand data and availability considering physiological parameters, capabilities and preferences of the workers. The agents, representing production orders, workers, robots and machines of the shop floor, interact with each other collaborating for a solution where an optimal allocation of workload given the circumstances is recommended to the line manager. MAS gives room for scalability and is capable of handling a high number of workers in large factories. It enhances efficiency in workload balance processes.

Service: Similarity Matching Knowledge Graph

The output of this solution is a suggestion of a person that can be assigned to a certain machine, e.g., to find a replacement for a missing colleague. The allocation should not be based on a one-to-one matching of people to possible machines but on the capabilities of the people in comparison to the requirements to fulfil the task at the machine. Therefore, the matching is based on the similarity between the capabilities of workers to the requirements of machines. For this approach, both the capabilities and the requirements for the tasks at a machine are described in a knowledge graph, which is then used to find provided capabilities that are similar to the task's requirements. The tasks on the machines and a list of workers must be provided as input.

Service: Resource Mapping with Decision Trees

The output of this solution is a list of workers who are allowed to work at a certain machine, e.g., as a replacement for a missing colleague. Like similarity matching with a knowledge graph, the requirements for working on a machine must be defined so that they can be matched to the capabilities of the work. For each type of task, a decision path is, in this case, described as a decision tree. Such a decision tree can support determining which worker can be allocated to the chosen machine.

Service: Community Detection with Knowledge Graph

The output of this solution is a list of workers similar to the provided one. Therefore, the workers and their characteristics are saved in a knowledge graph, and this structure is then used to find similar ones and defines them as one community. Such a community can then be applied to find workers with similar skills, which can be used to find a replacement for an absent worker.

Service: Similarity Matching with Semantics

The output of this solution is an estimate of how similar a defined capability of a worker is to the needed requirements of a task at a machine. This should be achieved by matching the labels in the description of the capabilities to the requirements so that they do not have to use the same label. This means the semantics of the labels is matched rather than the direct comparisons of strings. This solution can also be combined with other solutions where labels must be matched.

Service: Optimization/Heuristic-Based Resource Allocation Service

This AI service determines the effectiveness of job rotation to increase productivity and decrease both physical and mental fatigue. This rating is based on the available information on the current workload, the worker's experienced stress and their preferences.

Service: Operator Stress Estimation with Neural Networks

Stress directly affects human decision-making and might lead to numerous undesirable consequences, including increased distraction and deficits in the person's decision-making capabilities. Therefore, the consideration of levels of cognitive-emotional stress is important to harmonize man-machine production processes. A neural network is defined to map specific vital parameters (such as heart rate, heart rate variability, breathing rate, eye movement features, etc.) to a score for cognitive-emotional stress and consequently to the communication of risk levels in the context of efficient decision-making. In this use case, the responsible decision-makers and workers are attributed with risk levels associated with certain challenging tasks, and this score can be integrated into the reasoning and optimisation process on the distribution and balance of workload.

Service: Operator Persona Development with Machine Learning

Personas represent a group or population of real human operators regarding their Human Factors features, cognitive, affective, social, and motivational parameters. Specifically, based on a set of human profiles, a clustering methodology, such as a Machine Learning (ML) driven stochastic process, determines a mathematical representation of characteristic human profiles that can be applied for holistic optimisation processes with the human-in-the-loop. In this use case, specific persona profiles would be recommended to optimally allocate workers to production processes and maximise economic and well-being-related objectives.

Service: Order to Production Line with Fuzzy Rule

Experts that work with machines regularly have a deep understanding of how they operate and can frequently solve occurring problems intuitively. Fuzzy Logic allows for capturing and using unstructured, usually not mathematically definable, knowledge to strengthen decision-making. Therefore, this solution approach enables the inclusion of experience-based rules. Fuzzy Logic is particularly intriguing when support for very complicated decisions with numerous input factors should be offered since the concept of fuzziness allows one to handle these issues without requiring a lot of computing power. The purpose of the service is to incorporate informal, practical knowledge about the order assignments made by the machine operators. The service may evaluate and forecast whether a machine-to-order allocation is appropriate.

Service: Order to Production Line with Neural Networks

This service incorporates a neural network that learns and modifies weights based on a training set of historical allocation data. Since this solution approach is data-driven, the accuracy and quality of the outcomes largely depend on appropriate historical machine allocation activities. The neural network provides the capability to predict how orders will be distributed to a machine. The service is focused on recommending whether or not a particular sort of machine should be used to produce an order. Tasks involving operational production planning can benefit significantly from this information.

4.3.2 AI Services: Decision Support for "Solution Configuration"

Different machines' parameters and job shop features, like an order to machine, can be adjusted to find the most optimal setup for the system. The system can be configured differently by assigning specific values to its parameters. Each unique combination of parameter values represents a different solution configuration, which defines a particular variant or setup of the system. Based on the chosen configuration, these values determine how the system behaves, operates, or performs. The "Solution Configuration" decision support challenges the following scenarios.

4.3.2.1 FLEX Automated Test Building

Service: Rule-Based Resource Allocation Service

In this use case, a hybrid of AHP and Fuzzy logic could be applied to calculate the rating, e.g. safety and the efficiency of the proposed setup for the robot. In this approach, the AHP prioritises safety criteria by breaking down complex safety aspects into manageable components, assigning weights to each component based on its importance, and aligning with specific safety guidelines, ensuring high certainty in the safety rating. On the other hand, fuzzy logic is used to calculate efficiency, which can be difficult to quantify due to interdependent variables with different degrees of importance. Fuzzy logic is well-suited for dealing with unstructured or imprecise data, capturing the inherent uncertainty and variability of efficiency data and generating more accurate and nuanced results. As a result, a decision maker receives a rating – a score that suggests how well suited for the worker and production proposed design.

Service: Configuration Support Service

The only function of the Configuration Support Service is to provide a list of similar configurations to assist the user in a configuration task. This approach requires a vast amount of historical data of various designs based on which the most suitable design for a new idea could be suggested.

Service: Production Process Simulation with Agents

This service can address an Automated Test Building process by simulating interactions between the agents that represent these parties (engineer, operator, robots) using order specifics data to define restrictions in this approach. MAS can provide a recommendation between cooperative and collaborative modes for the robot operation based on the requirements of the production plan and eventually optimize the test time when workers' safety matters don't restrict further development of a possible solution. An Automated Test Building simulation on an agent-based approach can assist the manager in foreseeing an optimized process configuration and deterministically define where improvements can be made regarding cycle time.

Service: Production Process Simulation

This solution offers the capability to simulate the defined production processes. Simulations can be used to get dynamic information about a process, such as cycle time or waiting time. For the same problem, simulations with different configurations (e.g., using different working stations) can be executed and then compared. This comparison then supports the decision for the best alternative. Here it must be defined what the best solution is and which parameters are analyzed for this decision.

4.3.2.2 FLEX Machine Maintenance after Breakdown

Service: Configuration Support Service

The anomaly detection for the time series data is based on the time series prediction models like Long Short-Term Memory (LSTM). As an input to such a solution, a great deal of historical data relevant to the failure process parameters is required. History of machine failures along with the time when they happened, is also required as it serves as labels for anomaly detection. The benefits of anomaly detection include early detection of machine malfunctions, enabling timely actions to prevent further damage or minimize downtime. It can also assist in identifying potential issues that might go unnoticed through manual monitoring, allowing for prompt investigation and resolution.

Service: Multi-Agent based Resource Allocation

Machine breakdown is one of the most common unwanted events in medium and large industries. Whether due to a human problem in operation or technical problems that arise over the lifetime of the apparatus, among other types of malfunctions that occur due to specific issues, many reasons can stop machines during critical production periods. In order to make the production system resilient, MAS can reconfigure the shop floor by rearranging machines to contemplate the criticality of the production plan by taking an action that seeks a readjustment due to the downtime of a machine. Based on data which defines the shop floor map very well - containing available machines needed for the Multi-Agent System to have the flexibility and provide added value to the solution – this service improves production resilience. It offers a solution to disruption in production due to machine breakdown employing a physical restructuring. Also, MAS can assist in checking the reason of failure by accessing a failure database and communicating possible solutions to the error between operator and machine based on previous actions taken in resolution.

Service: Production Process Simulation with Agents

Simulation with agents can exploit diverse possibilities of machine arrangement beforehand and consider many possible formations based on established restrictions and data described in the previous service. Simulation allows the exploration of environmental reorganization innovatively and adapting production to new perspectives on operational efficiencies. Also, regarding maintenance, simulation can experiment with repair strategies based on historical data and even evaluate the benefits of predictive maintenance practices using real-time data from the machinery.

Service: Production Process Simulation

This solution offers the capability to simulate the defined production processes. Simulations can be used to get dynamic information about a process. This can be variables like cycle time or waiting time. For the same problem, simulations with different configurations (e.g., using different working stations) can be executed and then compared. This comparison then supports the decision for the best alternative. Here it must be defined what the best solution is and which parameters are analyzed for this decision.

This solution offers the capability to simulate the defined production processes. Simulations can be used to get dynamic information about a process, such as cycle time or waiting time. For the same problem, simulations with different configurations (e.g., using different working stations) can be executed and then compared. This comparison then supports the decision for the best alternative. Here it must be defined what the best solution is and which parameters are analyzed for this decision.

Service: Impact Assessment with Knowledge Graphs

This solution provides an analysis of how certain events influence the configuration of a production line. Therefore, the configuration is saved in a knowledge graph, and this structure is then used to analyze the impact. This can be used to identify the components of a production line configuration, which are influenced by certain events, e.g., the braking down of a machine or the delay of materials.

Service: Heuristic-Based Delay Assessment Service

Determines the temporal severity of a malfunction after a machine failure has occurred by making use of the historical data in the DAI-DSS Knowledge Base. This information can be used to assess the potential monetary impact on the production process, potential breaks for workers or a delay in completion.

4.3.2.3 CRF Delay of Material

Service: Reinforcement Learning Based Resource Allocation Service

A material delivery and availability delay can be treated as a resource allocation problem. It is, therefore, strongly related to ordering raw materials from the warehouse and keeping the warehouse stocks on a level to meet production demand. For this purpose, reinforcement learning could be applied similarly to the order to a machine.

Service: Production Process Simulation with Agents

In the delay of material scenario, agents can be deployed to coordinate future production plan by interacting present execution parameters with future production plans with the objective of supporting decisions on future material incoming to avoid material shortage. In case a material shortage is predicted, MAS simulation can provide a path to reorganizing machine and worker arrangement in the shop floor in order to fulfil priority orders.

Service: Production Process Simulation

This solution offers the capability to simulate the defined production processes. Simulations can be used to get dynamic information about a process. This can be variables like cycle time or waiting time. For the same problem, simulations with different configurations (e.g., using different working stations) can be executed and then compared. This comparison then supports the decision for the best alternative. Here it must be defined what the best solution is and which parameters are analyzed for this decision.

Service: Impact Assessment with Knowledge Graphs

This solution provides an analysis of how certain events influence the configuration of a production line. Therefore, the configuration is saved in a knowledge graph, and this structure is then used to analyze the impact. This can be used to identify the components of a production line configuration, which are influenced by certain events, e.g., the braking down of a machine or the delay of materials.

4.3.2.4 CRF Quality Issue

Service: Configuration Support Service

In this use case, ML methods can be applied to predict the sub-product or end-product quality based on input data, such as process parameters and quality measurement results. The focus can be on a certain production step, which is important to reach the intended quality, but also on the few process steps, complete process chain or hybrid configurations between knowledge- and data-driven methods.[184] Additionally, an optimizer can find the optimum process parameters to improve the quality and provides recommendations for adjustment.

4.3.3 AI Services: Decision Support for "Selection"

The "Resource Mapping" and "Solution Configuration" do not have to be completely automated tasks but could need further human input to create the solution candidates. The "Selection" challenge aims to provide a ranked list of possible solutions as output based on a set of input solutions. This ranking serves as a pre-selection for the DSS, which is then presented to the human decision-makers. The goal is not for the DSS to make the final decision on behalf of humans but to include them in the decision-making process.

In order to achieve the pre-selection, contextual information about the decision is required. This includes knowledge of the desired goal, which could be minimizing costs, reducing execution time, maximizing worker satisfaction, or a combination of multiple factors. The specific goal depends on the decision problem at hand and must be defined accordingly.

5 COMMUNICATION AND DISSEMINATION

This chapter discusses how information created during FAIRWork's research will be distributed to interested stakeholders. The focus is set on introducing the communication and dissemination channels and strategies which will be used to inform interested stakeholders about the progress and results of the FAIRWork project. As an outlook, we will also discuss how created information and artefacts are made available to interested people.

The channels of webinars, website and social media are used to communicate results and ongoing progress in the FAIRWork project to a broad audience. These channels are used to reach a vast number of people and inform them about the project to raise interest. Therefore, here we cannot presuppose specialized prior knowledge like in scientific communities. Social media activities, which are mainly focused on LinkedIn, will be used to regularly communicate important FAIRWork events (e.g., webinars) and results (e.g., publications) to a broad audience.

The website will be regularly updated to contain information about finished artefacts to make them accessible to interested parties. In the research context, this will be information about publications which were made in the context of FAIRWork. General information like the title, the authors and a short introduction will be provided with a link where the paper can be found.

Webinars will be held monthly by different partners communicating their current results and work. As the webinars are used for communication, they will be tailored to a broad audience and focus on informing and motivating. The webinars will be recorded and made publicly available.

For disseminating scientific results, specifically in the context of WP3, the FAIRWork project aims to publish in scientific conferences and journals. A set of possible conferences and journals has been identified initially and will be continuously updated during the project. These candidates are based on the scientific fields which fit the project's topics. Therefore, conferences and journals in the areas of AI, MAS, decision-making, information systems and conceptual modelling were explored. Table 4 contains the initial list of conferences and journals which were found until now.

Table 4: Conference and journal candidates for dissemination of project results

Title	Abbreviation	Type
Autonomous Agents and Multi-Agent Systems		Journal
Business and Information Systems Engineering	BISE	Journal
Business Process Management	BPM	Conference
Computers in Industry		Journal
Conference on Advanced Information Systems Engineering	CAiSE	Conference
European Conference on Artificial Intelligence	ECAI	Conference
Hawaii International Conference on System Sciences	HICSS	Conference
Human Factors: The Journal of the Human Factors and Ergonomics Society	Human Factors	Journal
IEEE International Requirements Engineering Conference	RE	Conference

Information Sciences		Journal
International Conference on AI for People	CAIP	Conference
International Conference on Computational Collective Intelligence	ICCCI	Conference
International Conference on Control, Decision and Information Technologies	CoDIT	Conference
International Conference on Decision Support System Technology	ICDSST	Conference
International Conference on Emerging Technologies and Factory Automation	ETFA	Conference
International Conference on Enterprise Information Systems	ICEIS	Conference
International Conference on Industrial Technology	ICIT	Conference
International Conference on Knowledge Science, Engineering and Management	KSEM 23	Conference
International Conference on Perspectives in Business Informatics Research	BIR	Conference
International Joint Conference on Artificial Intelligence Organisation	IJCAI	Conference
Internationale Tagung Wirtschaftsinformatik	WI	Conference
Journal of Artificial Intelligence	AIJ	Journal
Journal of Intelligent Information Systems	JIIS	Journal
Journal of Intelligent Manufacturing		Journal
KES Intelligent Decision Technologies	KES IDT	Conference
Modeling Decisions for Artificial Intelligence	MDAI	Conference
Open Research Europe	Open Research Europe	Open Access Publishing Platform
Practice of Enterprise Modelling	PoEM	Conference
Working Conference on Virtual Enterprises	PRO-VE	Conferences

Additionally to the above communication and dissemination channels, we will deploy an innovation shop to provide exploitation assets as information and artefacts to interested stakeholders. Here diverse artefacts with different readiness levels which are created within FAIRWork can be uploaded and made publicly accessible to a broad audience. This includes finished not only software or publications but also artefacts with a lower readiness level, like experiments, datasets, described methodologies or prototypes, which stakeholders can consume independently from other artefacts. The published innovation items can also be used to communicate information within the FAIRWork project. This work is performed in WP8.

6 SUMMARY AND CONCLUSIONS

This report consists of three main parts. The first part provides an overview of relevant literature, covering research directions such as the democratization of decision-making, digital shadows and twins, reliable and trustworthy Artificial Intelligence, and Artificial Intelligence and Multi-Agent Systems for improving decision support systems in manufacturing. The second part focuses on research methodologies, including data-driven modelling, prototyping, and testing. It also shows the application of sensors to capture critical information about humans' mental, affective, and motivational states. Furthermore, it introduces a novel framework using Personas for Human Digital Twins in decision-making. The final part identifies key research factors in industrial use cases, categorizes them into human and technical perspectives and provides potential AI services to address those use cases.

Identifying key research factors within industrial use cases further strengthens the practical implications of future studies. By analysing these factors from both human and technical perspectives, the report offers valuable insights that can guide developers and practitioners in optimizing their decision-support systems. This comprehensive understanding of the challenges and requirements in real-world scenarios eases the development of tailored solutions that address demanding manufacturing needs.

Ultimately, the collective efforts of examining literature, employing research methodologies, identifying key research factors, and implementing an effective communication strategy contribute to the broader goal of advancing decision-making processes and facilitating the successful adoption of Artificial Intelligence and Multi-Agent Systems technologies in Decision Support Systems.

7 REFERENCES

[1] Hleg, A. I. (2019). Ethics guidelines for trustworthy AI. B-1049 Brussels.

[2] Felzmann, H., Villaronga, E. F., Lutz, C., & Tamò-Larrieux, A. (2019). Transparency you can trust: Transparency requirements for artificial intelligence between legal norms and contextual concerns. *Big Data & Society*, 6(1), https://doi.org/10.1177/2053951719860542

[3] Schmidt, V. A. (2013). Democracy and legitimacy in the European Union revisited: Input, output and 'throughput'. *Political studies*, 61(1), 2-22. https://doi.org/10.1111/j.1467-9248.2012.00962.x

[4] Böschen, S., & Sigwart, H. J. (2020). Demokratisierung von Technikfolgenabschätzung? Zum Problem der Verhältnisbestimmung von Öffentlichkeit und wissenschaftlicher Expertise. *TATuP-Zeitschrift für Technikfolgenabschätzung in Theorie und Praxis/Journal for Technology Assessment in Theory and Practice*, *29*(3), 18-23. https://doi.org/10.14512/tatup.29.3.18

[5] Sørensen, E., & Torfing, J. (2007). Introduction Governance Network Research: Towards a Second Generation. *Theories of Democratic Network Governance*, 1-21. https://doi.org/10.1057/9780230625006_1

[6] Trein, P., Thomann, E., & Maggetti, M. (2019). Integration, functional differentiation and problem-solving in multilevel governance. *Public Administration*, 97(2), 339-354. https://doi.org/10.1111/padm.12595

[7] Kalyvas, A. (2005). Popular sovereignty, democracy, and the constituent power. Constellations, 12(2), 223-244. https://doi.org/10.1111/j.1351-0487.2005.00413.x

[8] Wenman, M. (2013). *Agonistic democracy: Constituent power in the era of globalisation*. Cambridge University Press. https://doi.org/10.1057/cpt.2014.2

[9] White, J., & Ypi, L. (2017). The politics of peoplehood. *Political Theory*, *45*(4), 439-465. https://doi.org/10.1177/0090591715608899

[10] Saward, M. (2006). The representative claim. *Contemporary Political Theory*, 5(3), 297-318. https://doi.org/10.1057/palgrave.cpt.9300234

[11] *The Constructivist Turn in Political Representation*. (2019). Edinburgh University Press.

[12] Gomart, E., & Hajer, M. (2003). Is That Politics? Social Studies of Science and Technology: Looking Back, Ahead, 33–61. https://doi.org/10.1007/978-94-010-0185-4_3

[13] Brown, M. B. (2009). *Science in democracy: Expertise, institutions, and representation*. MIT Press.

[14] Chilvers, J., & Kearnes, M. (2015). *Remaking participation: Science, environment and emergent publics*. Routledge.

[15] Hayat, S. (2019). *Representation as proposition: Democratic representation after the constructivist turn. The constructivist turn in political representation*, 121-40.

[16] Laurent, B. (2011). Technologies of democracy: Experiments and demonstrations. *Science and Engineering Ethics*, *17*, 649-666. https://doi.org/10.1007/s11948-011-9303-1

[17] Chilvers, J., Pallett, H., & Hargreaves, T. (2018). Ecologies of participation in socio-technical change: The case of energy system transitions. *Energy Research & Social Science*, 42, 199-210. https://doi.org/10.1016/j.erss.2018.03.020

[18] Lezaun, J., & Soneryd, L. (2007). Consulting citizens: Technologies of elicitation and the mobility of publics. *Public understanding of science*, 16(3), 279-297. https://doi.org/10.1177/0963662507079371

[19] Felt, U., & Fochler, M. (2010). Machineries for making publics: Inscribing and de-scribing publics in public engagement. *Minerva*, *48*, 219-238. https://doi.org/10.1007/s11024-010-9155-x

[20] Soneryd, L., & Amelung, N. (2016). Translating Participation: Scenario Workshops and Citizens' Juries across Situations and Contexts. In J.-P. Voß & R. Freeman (Eds.), *Knowing Governance. The Epistemic Construction of Political Order* (pp. 155-174). Basingstoke: Palgrave Macmillan. https://doi.org/10.1057/9781137514509_7

[21] Konopásek, Z., Soneryd, L., & Svačina, K. (2018). Lost in translation: Czech dialogues by Swedish design. *Science & Technology Studies*, 31(3), 5-23. https://doi.org/10.23987/sts.65543

[22] Voß, J. P., Schritt, J., & Sayman, V. (2022). Politics at a distance: Infrastructuring knowledge flows for democratic innovation. *Social Studies of Science*, 52(1), 106-126. https://doi.org/10.1177/03063127211033990

[23] Bond, A. H., & Gasser, L. G. (1988). Readings in distributed artificial intelligence, San Mateo, Calif. *M. Kaufmann*.

[24] Wooldridge, M. (2009). *An introduction to multiagent systems*. John wiley & sons.

[25] Dostatni, E., Diakun, J., Grajewski, D., Wichniarek, R., & Karwasz, A. (2016). Multi-agent system to support decision-making process in design for recycling. *Soft Computing*, *20*, 4347-4361. https://doi.org/10.1007/s00500-016-2302-z

[26] Xie, J., & Liu, C. C. (2017). Multi-agent systems and their applications. *Journal of International Council on Electrical Engineering*, 7(1), 188-197. https://doi.org/10.1080/22348972.2017.1348890

[27] Jong, S. D. E., Tuyls, K., & Verbeeck, K. (2008). Fairness in multi-agent systems. The Knowledge Engineering Review, 23(2), 153–180. https://doi.org/10.1017/s026988890800132x

[28] Chen, J. Y. C., Lakhmani, S. G., Stowers, K., Selkowitz, A. R., Wright, J. L., & Barnes, M. (2018). Situation awareness-based agent transparency and human-autonomy teaming effectiveness. Theoretical Issues in Ergonomics Science, 19(3), 259–282. https://doi.org/10.1080/1463922x.2017.1315750

[29] Rădulescu, R., Mannion, P., Roijers, D. M., & Nowé, A. (2019). Multi-objective multi-agent decision making: a utility-based analysis and survey. *Autonomous Agents and Multi-Agent Systems*, 34(1), 10. ArXiv:1909.02964[Cs]. https://doi.org/10.48550/arXiv.1909.02964

[30] Zhang, Y., Siwen, Y., & Xu, X. (2008, 12-14 Oct. 2008). *An Administrative Decision-Making Support System Based on Multi-Agent Technology.* Paper presented at the 2008 4th International Conference on Wireless Communications, Networking and Mobile Computing.

[31] Krupitzer, C., Müller, S., Lesch, V., Züfle, M., Edinger, J., Lemken, A., Schäfer, D., Kounev, S., & Becker, C. (2020). A Survey on Human Machine Interaction in Industry 4.0. ArXiv: 2002.01025 [Cs]. https://doi.org/10.48550/arXiv.2002.01025

[32] Kim, Y., & Matson, E. T. (2016). A realistic decision making for task allocation in heterogeneous multi-agent systems. *Procedia Computer Science*, *94*, 386-391.

[33] Nixdorf, S., Ansari, F., & Schlund, S. (2022). Reciprocal Learning in human-Machine Collaboration: A multi-agent system framework in Industry 5.0. In *Digitization of the work environment for sustainable production*, 207–225. https://doi.org/10.30844/wgab_2022_11

[34] Hanga, K. M., & Kovalchuk, Y. (2019). Machine learning and multi-agent systems in oil and gas industry applications: A survey. *Computer Science Review*, 34, 100191. https://doi.org/10.1016/j.cosrev.2019.08.002

[35] Kim, Y. G., Lee, S., Son, J., Bae, H., & Chung, B. D. (2020). Multi-agent system and reinforcement learning approach for distributed intelligence in a flexible smart manufacturing system. *Journal of Manufacturing Systems*, *57*, 440–450. https://doi.org/10.1016/j.jmsy.2020.11.004

[36] Zhang, P., Hao, J., Wang, W., Tang, H., Ma, Y., Duan, Y., & Zheng, Y. (2020). KoGuN: Accelerating Deep Reinforcement Learning via Integrating Human Suboptimal Knowledge. *CoRR*, ArXiv:2002.07418[Cs]. https://doi.org/10.48550/arXiv.2002.07418

[37] Kagermann, H., Wahlster, W., & Helbig, J. (2013). Securing the future of German manufacturing industry: Recommendations for implementing the strategic initiative INDUSTRIE 4.0, Final report of the Industrie 4.0 Working Group. In Forschungsunion & acatech (Issue April).

[38] Winkelhaus, S., Grosse, E. H., & Morana, S. (2021). Towards a conceptualisation of Order Picking 4.0. Computers & Industrial Engineering, 159, Article 107511. https://doi.org/10.1016/j.cie.2021.107511

[39] Breque, M., De Nul, L., & Petridis, A. (2021). Industry 5.0: Towards a sustainable, human-centric and resilient European industry. https://doi.org/10.2777/308407

[40] Dixson-Decl`eve, S., Balland, P.-A., Bria, F., Dunlop, K., Giovannini, E., Tataj, D., Hidalgo, C., Huang, A., Isaksson, D., Martins, F., Mir Roca, M., Morlet, A., Renda, A., & Schwaag Serger, S. (2022). Industry 5.0: A Transformative Vision for Europe – Governing systemic transformations towards a sustainable industry, Publications Office of the European Union. https://doi.org/https://data.europa.eu/doi/10.2777/17322

[41] Xie, J., Gao, L., Peng, K., Li, X., & Li, H. (2019). Review on flexible job shop scheduling. *IET Collaborative Intelligent Manufacturing*, 1(3), 67-77. https://doi.org/10.1049/iet-cim.2018.0009

[42] Kress, D., Müller, D. & Nossack, J. (2019). A worker constrained flexible job shop scheduling problem with sequence-dependent setup times. *OR Spectrum*, 41, 179–217. https://doi.org/10.1007/s00291-018-0537-z

[43] Zhang, J., Ding, G., Zou, Y., Qin, S., & Fu, J. (2019). Review of job shop scheduling research and its new perspectives under Industry 4.0. *Journal of Intelligent Manufacturing*, 30, 1809-1830. https://doi.org/10.1007/s10845-017-1350-2

[44] Liu, Y., Wang, L., Wang, X. V., Xu, X., & Zhang, L. (2019). Scheduling in cloud manufacturing: state-of-the-art and research challenges. *International Journal of Production Research*, 57(15-16), 4854-4879. https://doi.org/10.1080/00207543.2018.1449978

[45] Destouet, C., Tlahig, H., Bettayeb, B., & Mazari, B. (2023). Flexible job shop scheduling problem under Industry 5.0: A survey on human reintegration, environmental consideration and resilience improvement. *Journal of Manufacturing Systems*, 67, 155-173. https://doi.org/10.1016/j.jmsy.2023.01.004

[46] Zhang, S., Du, H., Borucki, S., Jin, S., Hou, T., & Li, Z. (2021). Dual Resource Constrained Flexible Job Shop Scheduling Based on Improved Quantum Genetic Algorithm. *Machines*, 9(6), 108. https://doi.org/10.3390/machines9060108

[47] Blandino, G. (2023). How to Measure Stress in Smart and Intelligent Manufacturing Systems: A Systematic Review. *Systems*, 11(4), 167. https://doi.org/10.3390/systems11040167

[48] Di Pasquale, V., De Simone, V., Radano, M., Miranda, S. (2022). Wearable devices for health and safety in production systems: a literature review, IFAC-PapersOnLine, Vol. 55, pp. 341-346, https://doi.org/10.1016/j.ifacol.2022.09.410

[49] Akbar, F., Mark, G., Pavlidis, I., & Gutierrez-Osuna, R. (2019). An empirical study comparing unobtrusive physiological sensors for stress detection in computer work. *Sensors* (Switzerland), 19(17), 3766. https://doi.org/10.3390/s19173766

[50] Heikkilä, P., Schmalfuß, F., Honka, A., Kaasinen, E., Mach, S., & Väänänen, K. (2018). Quantified Factory Worker—Expert Evaluation and Ethical Considerations of Wearable Self-tracking Devices. *ACM International Conference Proceeding Series*, 202–211. https://doi.org/10.1145/3275116.3275119

[51] Leone, A., Rescio, G., Siciliano, P., Papetti, A., Brunzini, A., & Germani, M. (2020). Multi sensors platform for stress monitoring of workers in smart manufacturing context. *2020 IEEE International Instrumentation and Measurement Technology Conference (I2MTC*), 3–7. https://doi.org/10.1109/i2mtc43012.2020.9129288

[52] Montesinos, V., Dell'Agnola, F., Arza, A., Aminifar, A., & Atienza, D. (2019). Multi-Modal Acute Stress Recognition Using Off-the-Shelf Wearable Devices. *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC).* https://doi.org/10.1109/embc.2019.8857130

[53] Sanchez, W., Martinez, A., Hernandez, Y., Estrada, H., & Gonzalez-Mendoza, M. (2018). A predictive model for stress recognition in desk jobs. *Journal of Ambient Intelligence and Humanized Computing,* 14, 17–29. https://doi.org/10.1007/s12652-018-1149-9

[54] Pfeiffer, J., Pfeiffer, T., Meißner, M., & Weiß, E. (2020). Eye-tracking-based classification of information search behavior using machine learning: evidence from experiments in physical shops and virtual reality shopping environments. Information Systems Research, 31(3), 675–691. https://doi.org/10.1287/isre.2019.0907

[55] Peißl, S., Wickens, C. D., & Baruah, R. (2018). Eye-Tracking Measures in Aviation: A Selective Literature Review. *The International Journal of Aerospace Psychology*, 28 (3–4), 98–112. https://doi.org/10.1080/24721840.2018.1514978

[56] Rosch, J. L., & Vogel-Walcutt, J. J. (2013). A review of eye-tracking applications as tools for training. *Cognition, Technology & Work*, 15(3), 313–327. https://doi.org/10.1007/s10111-012-0234-7

[57] Stork, S., StoBel, C., Muller, H. J., Wiesbeck, M., Zah, M. F., & Schubo, A. (2007). A Neuroergonomic Approach for the Investigation of Cognitive Processes in Interactive Assembly Environments. In *RO-MAN 2007 - The 16th IEEE International Symposium on Robot and Human Interactive Communication*, 750–755. https://doi.org/10.1109/ROMAN.2007.4415185

[58] Bovo, R., Binetti, N., Brumby, D. P., & Julier, S. (2020). Detecting Errors in Pick and Place Procedures Detecting. *Proceedings of the 25th International Conference on Intelligent User Interfaces*, 536–545. https://doi.org/10.1145/3377325.3377497

[59] Straeter, O. (2020). An Objective Measure for Detecting Workload and Errors in Monotone, Repetitive or Fatigue-Causing Environments Using Pupil Variation. *Communications in Computer and Information Science,* 1318, 3–12. Springer International Publishing. https://doi.org/10.1007/978-3-030-62302-9_1

[60] Paletta, L., Dini, A., Murko, C., Yahyanejad, S., & Augsdörfer, U. (2019). Estimation of situation awareness score and performance using eye and head gaze for human-robot collaboration. *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications*, 3322504(c), 1–3. https://doi.org/10.1145/3314111.3322504
[61] Haslgrübler, M., Gollan, B., Thomay, C., Ferscha, A., & Heftberger, J. (2019). Towards skill recognition using eye-hand coordination in industrial production. *Proceedings of the 12th ACM International Conference on PErvasive Technologies Related to Assistive Environments*, 11–20. https://doi.org/10.1145/3316782.3316784
[62] Tuncer, M. C., Ozkan, N. F., & Ulutas, B. H. (2020). A systematic approach to assess visual inspection tasks: An assembly line case. International Journal of Productivity and Quality Management, 29(3), 418. https://doi.org/10.1504/IJPQM.2020.106024
[63] Zheng, T., Glock, C.H., Grosse, E.H. (2022). Opportunities for using eye tracking technology in manufacturing and logistics: Systematic literature review and research agenda, Computers & Industrial Engineering, 171. https://doi.org/10.1016/j.cie.2022.108444
[64] Demerouti E., Bakker A.B., Nachreiner F., Schaufeli W.B. (2001). The job demands-resources model of burnout. Journal of Applied psychology 86(3), 499–512. https://doi.org/10.1037/0021-9010.86.3.499
[65] Carod-Artal F.J., Vázquez-Cabrera C. (2013). Burnout syndrome in an international setting. *Burnout for Experts*. Springer. 15–35. https://doi.org/10.1007/978-1-4614-4391-9_2
[66] Boksem, M. A. S., Meijman, T. F., & Lorist, M. M. (2006). Mental fatigue, motivation and action monitoring. *Biological Psychology*, 72(2), 123–132. https://doi.org/10.1016/j.biopsycho.2005.08.007
[67] Hockey, G. R. J., John Maule, A., Clough, P. J., & Bdzola, L. (2000). Effects of negative mood states on risk in everyday decision making. *Cognition and Emotion*, 14(6), 823–855. https://doi.org/10.1080/026999300050156654
[68] Linder, J. A., Doctor, J. N., Friedberg, M. W., Reyes Nieva, H., Birks, C., Meeker, D., & Fox, C. R. (2014). Time of Day and the Decision to Prescribe Antibiotics. *JAMA Internal Medicine*, 174(12), 2029. https://doi.org/10.1001/jamainternmed.2014.5225
[69] Porcelli, A. J., & Delgado, M. R. (2017). Stress and Decision Making: Effects on Valuation, Learning, and Risk-taking. Current opinion in behavioral sciences, 14, 33–39. https://doi.org/10.1016/j.cobeha.2016.11.015
[70] Klein, G. (2008). Naturalistic Decision Making. *Human Factors*, 50(3), 456–460. https://doi.org/10.1518/001872008X288385
[71] Galván, A., & Rahdar, A. (2013). The neurobiological effects of stress on adolescent decision making. *Neuroscience*, 249, 223–231. https://doi.org/10.1016/j.neuroscience.2012.09.074
[72] Driskell, J. E., Salas, E., & Johnston, J. (1999). Does stress lead to a loss of team perspective? Group Dynamics: Theory, *Research, and Practice*, 3(4), 291–302. https://doi.org/10.1037/1089-2699.3.4.291
[73] Miller, M. E., & Spatz, E. (2022). A unified view of a human digital twin. *Human-Intelligent Systems Integration*, 4(1–2), 23–33. https://doi.org/10.1007/s42454-022-00041-x
[74] Lutze, R. (2020). Digital Twin Based Software Design in eHealth - A New Development Approach for Health / Medical Software Products. *2020 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC)*. https://doi.org/10.1109/ice/itmc49519.2020.9198546
[75] Chakshu, N. K., Sazonov, I., & Nithiarasu, P. (2020). Towards enabling a cardiovascular digital twin for human systemic circulation using inverse analysis. *Biomechanics and Modeling in Mechanobiology*, 20(2), 449–465. https://doi.org/10.1007/s10237-020-01393-6
[76] Barricelli, B. R., Casiraghi, E., Gliozzo, J., Petrini, A., & Valtolina, S. (2020). *Human Digital Twin for Fitness Management. IEEE Access*, 8, 26637–26664. https://doi.org/10.1109/access.2020.2971576
[77] Caputo, F., Greco, A., Fera, M., & Macchiaroli, R. (2019). Digital twins to enhance the integration of ergonomics in the workplace design. *International Journal of Industrial Ergonomics*, 71, 20–31. https://doi.org/10.1016/j.ergon.2019.02.001
[78] Greco, A., Caterino, M., Fera, M., & Gerbino, S. (2020). Digital Twin for Monitoring Ergonomics during Manufacturing Production. *Applied Sciences*, 10(21), 7758. https://doi.org/10.3390/app10217758

[79] Sharotry, A., Jimenez, J. A., Wierschem, D., Mendez Mediavilla, F. A., Koldenhoven, R. M., Valles, D., Koutitas, G., & Aslan, S. (2020). A Digital Twin Framework for Real-Time Analysis and Feedback of Repetitive Work in the Manual Material Handling Industry. *2020 Winter Simulation Conference (WSC).* https://doi.org/10.1109/wsc48552.2020.9384043

[80] Onan Demirel, H., Irshad, L., Ahmed, S., & Tumer, I. Y. (2021). Digital Twin-Driven Human-Centered Design Frameworks for Meeting Sustainability Objectives. *Journal of Computing and Information Science in Engineering*, 21(3). https://doi.org/10.1115/1.4050684

[81] Boschert, S., Rosen, R. (2016). Digital Twin—The Simulation Aspect. In: Hehenberger, P., Bradley, D. (Eds.) *Mechatronic Futures*. Springer. https://doi.org/10.1007/978-3-319-32156-1_5

[82] Paletta, L. (2021). Attention Analysis for Assistance in Assembly Processes, In Ed., Ratchev, S., Smart Technologies for Precision Assembly (Eds.), *9th IFIP WG 5.5 International Precision Assembly Seminar, IPAS 2020, Virtual Event, December 14–15, 2020*, Revised Selected Papers Springer International Publishing, https://www.springerprofessional.de/attention-analysis-for-assistance-in-assembly-processes/19030464

[83] Paletta, L., Ganster, H., Schneeberger, M., Pszeida, M., Lodron, G., Pechstädt, K., Spitzer, M., & Reischl, C. (2021). Towards Large-Scale Evaluation of Mental Stress and Biomechanical Strain in Manufacturing Environments Using 3D-Referenced Gaze and Wearable-based Analytics. *Electronic Imaging*, 33(6), 310-1-310–317. https://doi.org/10.2352/issn.2470-1173.2021.6.iriacv-310

[84] Paletta, L., Pszeida, M., Ganster, H., Fuhrmann, F., Weiss, W., Ladstätter, S., Dini, A., Murg, S., Mayer, H., Brijacak, I., and Reiterer, B. (2019). Gaze based Human Factors Measurements for the Evaluation of Intuitive Human-Robot Collaboration in Real-time. *2019 24th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA).* https://doi.org/10.1109/etfa.2019.8869270

[85] Felsberger, A.; Oberegger, B.; Reiner, G. (2016): A Review of Decision Support Systems for Manufacturing Systems. SamI40 workshop at i-KNOW.

[86] Nemati, H. R., Steiger, D. M., Iyer, L. S., & Herschel, R. T. (2002). Knowledge warehouse: an architectural integration of knowledge management, decision support, artificial intelligence and data warehousing. *Decision Support Systems*, 33(2), 143–161. https://doi.org/10.1016/s0167-9236(01)00141-5

[87] Negnevitsky, M (2005). *Artificial intelligence: a guide to intelligent systems*. Pearson Education Limited

[88] Kapanoglu, M., & Miller, W. A. (2004). An evolutionary algorithm-based decision support system for managing flexible manufacturing. *Robotics and Computer-Integrated Manufacturing*, *20*(6), 529-539. https://doi.org/10.1016/j.rcim.2004.07.008

[89] Marchetta, M., & Forradellas, R. (2007). Artificial intelligence planning for generative computer aided process planning. *19th International Conference on Production Research.*

[90] Leo Kumar, S. P. (2017). State of the art-intense review on artificial intelligence systems application in process planning and manufacturing. *Engineering Applications of Artificial Intelligence*, 65, 294-329. https://doi.org/10.1016/j.engappai.2017.08.005

[91] Onwubolu, G. C., & Clerc, M. (2004). Optimal path for automated drilling operations by a new heuristic approach using particle swarm optimization. *International Journal of Production Research*, 42(3), 473–491. https://doi.org/10.1080/00207540310001614150

[92] Solimanpur, M., Vrat, P., & Shankar, R. (2004). Ant colony optimization algorithm to the inter-cell layout problem in cellular manufacturing. *European Journal of Operational Research*, 157(3), 592–606. https://doi.org/10.1016/s0377-2217(03)00248-0

[93] Kapoor, V., & Tak, S. S. (2005). Fuzzy Application to the Analytic Hierarchy Process for Robot Selection. *Fuzzy Optimization and Decision Making*, 4(3), 209–234. https://doi.org/10.1007/s10700-005-1890-3

[94] Son, C. (2016). Intelligent rule-based sequence planning algorithm with fuzzy optimization for robot manipulation tasks in partially dynamic environments. *Information Sciences*, 342, 209–221. https://doi.org/10.1016/j.ins.2015.08.020

[95] Zhao, W., Li, J., Zhang, C., Liang, Z., & Wang, B. (2006). Feature-based cost estimation for composite structures with Fuzzy Multi-Attribute Utility Theory. *International Journal of Industrial and Systems Engineering*, *1*(1/2), 284-300. https://doi.org/10.1504/IJISE.2006.009061

[96] Chu, T.-C., & Lin, Y.-C. (2003). A Fuzzy TOPSIS Method for Robot Selection. *The International Journal of Advanced Manufacturing Technology*, 21(4), 284–290. https://doi.org/10.1007/s001700300033

[97] Ghahramani, M., Qiao, Y., Zhou, M. C., O'Hagan, A., & Sweeney, J. (2020). AI-based modeling and data-driven evaluation for smart manufacturing processes. *IEEE/CAA Journal of Automatica Sinica*, 7(4), 1026–1037. https://doi.org/10.1109/jas.2020.1003114

[98] Guo, R., Mo, R., & Sun, H. B. (2012). Study on Supplier Selection and Evaluation in Serviced-Oriented Manufacturing Network Based on NN. *Advanced Materials Research*, 479–481, 644–647. https://doi.org/10.4028/www.scientific.net/amr.479-481.644

[99] Wu, D., Jennings, C., Terpenny, J., Gao, R. X., & Kumara, S. (2017). A Comparative Study on Machine Learning Algorithms for Smart Manufacturing: Tool Wear Prediction Using Random Forests. *Journal of Manufacturing Science and Engineering*, 139(7). https://doi.org/10.1115/1.4036350

[100] Waubert de Puiseau, C., Meyes, R., & Meisen, T. (2022). On reliability of reinforcement learning based production scheduling systems: a comparative survey. *Journal of Intelligent Manufacturing*, 33(4), 911–927. https://doi.org/10.1007/s10845-022-01915-2

[101] Samsonov, V., Ben Hicham, K., & Meisen, T. (2022). Reinforcement Learning in Manufacturing Control: Baselines, challenges and ways forward. *Engineering Applications of Artificial Intelligence*, 112, 104868. https://doi.org/10.1016/j.engappai.2022.104868

[102] Samsonov, V., Kemmerling, M., Paegert, M., Lütticke, D., Sauermann, F., Gützlaff, A., ... & Meisen, T. (2021, February). Manufacturing Control in Job Shop Environments with Reinforcement Learning. *Proceedings of the 13th International Conference on Agents and Artificial Intelligence.* https://doi.org/10.5220/0010202405890597

[103] Zhang, C., Song, W., Cao, Z., Zhang, J., Tan, P. S., & Chi, X. (2020). Learning to dispatch for job shop scheduling via deep reinforcement learning. *Advances in Neural Information Processing Systems*, 33, 1621-1632. ArXiv:2010.12367[Cs]. https://doi.org/10.48550/arXiv.2010.12367

[104] Tassel, P., Gebser, M., & Schekotihin, K. (2021). A reinforcement learning environment for job-shop scheduling. ArXiv:2104.03760[Cs]. https://doi.org/10.48550/arXiv.2104.03760

[105] Kar, Arpan Kumar (2015): A hybrid group decision support system for supplier selection using analytic hierarchy process, fuzzy set theory and neural network. *Journal of Computational Science,* 6, 23–33. https://doi.org/10.1016/j.jocs.2014.11.002

[106] Rabta, B., & Reiner, G. (2012). Batch sizes optimisation by means of queueing network decomposition and genetic algorithm. *International Journal of Production Research*, 50(10), 2720–2731. https://doi.org/10.1080/00207543.2011.588618

[107] Shamim, A., Hussain, H., & Shaikh, M. U. (2010). A framework for generation of rules from decision tree and decision table. *2010 International Conference on Information and Emerging Technologies*, 1–6. https://doi.org/10.1109/ICIET.2010.5625700

[108] Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes, 50(2*), 179–211. https://doi.org/10.1016/0749-5978(91)90020-T

[109] Ajzen, I. (2014). The theory of planned behaviour is alive and well, and not ready to retire [Peer commentary on the paper "Time to retire the theory of planned behaviour" by FF Sniehotta, J. Presseau, & V. Araújo-Soares*]. Health Psychology Review, 9(2*), 131–137. https://doi.org/10.1080/17437199.2014.883474

[110] Davis, F. D. (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology*. MIS Quarterly, 13(3)*, 319–340. JSTOR. https://doi.org/10.2307/249008

[111] Venkatesh, V., & Davis, F. D. (2000). A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies. *Management Science*, 46(2), 186–204. https://doi.org/10.1287/mnsc.46.2.186.11926

[112] Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User Acceptance of Information Technology: Toward a Unified View*. MIS Quarterly*, 425–478. https://ssrn.com/abstract=3375136

[113] Felzmann, H., Villaronga, E. F., Lutz, C., & Tamò-Larrieux, A. (2019). Transparency you can trust: Transparency requirements for artificial intelligence between legal norms and contextual concerns. *Big Data & Society, 6(1*), 1–14. https://doi.org/10.1177/2053951719860542

[114] Mayer, R. C., & Davis, J. H. (1995). An Integrative Model of Organizational Trust. *The Academy of Management Review,* 20(3), 709–734. https://www.jstor.org/stable/258792

[115] Söllner, M., Hoffmann, A., Hoffmann, H., Wacker, A., & Leimeister, J. M. (2012). Understanding the Formation of Trust in IT Artifacts. *Proceedings of the International Conference on Information Systems*, 1–18.

[116] Dietvorst, B. J., Simmons, J. P., & Massey, C. (2014). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology,* 144(1), 114–126. https://doi.org/10.1037/xge0000033

[117] Dietvorst, B. J., Simmons, J. P., & Massey, C. (2018). Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them. *Management Science,* 64(3), 1155–1170. https://doi.org/10.1287/mnsc.2016.2643

[118] Dzindolet, M. T., Peterson, S. A., Pomranky, R. A., Pierce, L. G., & Beck, H. P. (2003). The role of trust in automation reliance*. International Journal of Human-Computer Studies,* 58(6), 697–718. https://doi.org/10.1016/S1071-5819(03)00038-7

[119] Lee, P. (2016, March 25). Learning from Tay's introduction. The Official Microsoft Blog. https://blogs.microsoft.com/blog/2016/03/25/learning-tays-introduction/

[120] O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Penguin UK.

[121] Verma, P. (2022, July 19). *These robots were trained on AI. They became racist and sexist*. Washington Post. https://www.washingtonpost.com/technology/2022/07/16/racist-robots-ai/

[122] Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion, 58,* 82–115. https://doi.org/10.1016/j.inffus.2019.12.012

[123] Rai, A. (2020). Explainable AI: From black box to glass box. *Journal of the Academy of Marketing Science, 48(1)*, 137–141. https://doi.org/10.1007/s11747-019-00710-5

[124] Murdoch, W. J., Singh, C., Kumbier, K., Abbasi-Asl, R., & Yu, B. (2019). Definitions, methods, and applications in interpretable machine learning. *Proceedings of the National Academy of Sciences, 116(44),* 22071–22080. https://doi.org/10.1073/pnas.1900654116

[125] Miller, T. (2018). Explanation in Artificial Intelligence: Insights from the Social Sciences. ArXiv:1706.07269[Cs]. https://doi.org/10.48550/arXiv.1706.07269

[126] Páez, A. (2019). The Pragmatic Turn in Explainable Artificial Intelligence (XAI). *Minds and Machines,* 29(3), 441–459. https://doi.org/10.1007/s11023-019-09502-w

[127] Verordnung (EU) 2016/679 des Europäischen Parlaments und des Rates vom 27. April 2016 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten, zum freien Datenverkehr und zur Aufhebung der Richtlinie 95/46/EG (Datenschutz-Grundverordnung), Pub. L. No. 2016/679 (2016). https://eur-lex.europa.eu/legal-content/DE/TXT/HTML/?uri=CELEX:32016R0679

[128] *High-level expert group on artificial intelligence.* (2023, May 25). https://digital-strategy.ec.europa.eu/en/policies/expert-group-ai

[129] AI HLEG. (2019). *Ethics guidelines for trustworthy AI.* European Commission. https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai

[130] AI HLEG. (2019). *Policy and investment recommendations for trustworthy Artificial Intelligence.* https://digital-strategy.ec.europa.eu/en/library/policy-and-investment-recommendations-trustworthy-artificial-intelligence

[131] *The Digital Services Act package* (2023, May 23). https://digital-strategy.ec.europa.eu/en/policies/digital-services-act-package

[132] Zerilli, J., Bhatt, U., & Weller, A. (2022). How transparency modulates trust in artificial intelligence. *Patterns,* 3(4), 100455. https://doi.org/10.1016/j.patter.2022.100455

[133] Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1(9), 9. https://doi.org/10.1038/s42256-019-0088-2

[134] Hoff, K. A., & Bashir, M. (2015). Trust in Automation: Integrating Empirical Evidence on Factors That Influence Trust. *Human Factors, 57(3*), 407–434. https://doi.org/10.1177/0018720814547570

[135] Ross, J. M., Szalma, J. L., Hancock, P. A., Barnett, J. S., & Taylor, G. (2008). The Effect of Automation Reliability on User Automation Trust and Reliance in a Search-and-Rescue Scenario. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 52(19), 1340–1344. https://doi.org/10.1177/154193120805201908

[136] Glikson, E., & Woolley, A. W. (2020). Human Trust in Artificial Intelligence: Review of Empirical Research. *Academy of Management Annals, 14(2)*, 627–660. https://doi.org/10.5465/annals.2018.0057

[137] Miller, T. (2018). Explanation in Artificial Intelligence: Insights from the Social Sciences. *ArXiv:1706.07269 [Cs].* https://doi.org/10.48550/arXiv.1706.07269

[138] Venkatesh, V., Brown, S. A., & Sullivan, Y. W. (2016). Guidelines for conducting mixed-methods research: An extension and illustration. *Journal of the Association for Information Systems, 17(7)*, 435–494. https://doi.org/10.17705/1jais.00433

[139] Siau, K., & Wang, W. (2018). Building Trust in Artificial Intelligence, Machine Learning, and Robotics. *Cutter Business Technology Journal*, 31(2), 47-53. Cutter Consortium.

[140] Schmidt, P., Biessmann, F., & Teubner, T. (2020). Transparency and trust in artificial intelligence systems. *Journal of Decision Systems, 29(4)*, 260–278. https://doi.org/10.1080/12460125.2020.1819094

[141] Springer, A. (2019). Accurate, Fair, and Explainable: Building Human-Centered AI [UC Santa Cruz]. https://escholarship.org/uc/item/4d80t5j5

[142] Kizilcec, R. F. (2016). How much information?: Effects of transparency on trust in an algorithmic interface. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems,* 2390–2395. https://doi.org/10.1145/2858036.2858402

[143] Larsson, S., & Heintz, F. (2020). Transparency in artificial intelligence. *Internet Policy Review, 9(2).* https://policyreview.info/concepts/transparency-artificial-intelligence

[144] Felzmann, H., Villaronga, E. F., Lutz, C., & Tamò-Larrieux, A. (2019). Transparency you can trust: Transparency requirements for artificial intelligence between legal norms and contextual concerns. *Big Data & Society, 6(1),* 1–14. https://doi.org/10.1177/2053951719860542

[145] Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence, 1(9),* 9. https://doi.org/10.1038/s42256-019-0088-2

[146] Bahill, A. T., Adler, D., & Stark, L. (1975). Most naturally occurring human saccades have magnitudes of 15 degrees or less. *Investigative ophthalmology*, *14*(6), 468–469.

[147] Bulling, A., Ward, J. A., Gellersen, H., & Tröster, G. (2011). Eye Movement Analysis for Activity Recognition Using Electrooculography. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(4), 741–753. https://doi.org/10.1109/tpami.2010.86

[148] Saucedo-Martínez, J. A., Pérez-Lara, M., Marmolejo-Saucedo, J. A., Salais-Fierro, T. E., & Vasant, P. (2017). Industry 4.0 framework for management and operations: a review. *Journal of Ambient Intelligence and Humanized Computing*, 9(3), 789–801. https://doi.org/10.1007/s12652-017-0533-1

[149] Steinfeld, A., Fong, T., Kaber, D., Lewis, M., Scholtz, J., Schultz, A., & Goodrich, M. (2006). Common metrics for human-robot interaction. *Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction.* https://doi.org/10.1145/1121241.1121249

[150] Huang, C. M., & Mutlu, B. (2016). Anticipatory robot control for efficient human-robot collaboration. *2016 11th ACM/IEEE international conference on human-robot interaction (HRI)*, 83-90. https://doi.org/10.1109/hri.2016.7451737

[151] Paletta, L., Brijacak, I., Reiterer, B., Pszeida, M., Ganster, H., Fuhrmann, F., Weiss, W., Ladstatter, S., Dini, A., Murg, S., & Mayer, H. (2019). Gaze-based Human Factors Measurements for the Evaluation of Intuitive Human-Robot Collaboration in Real-time. *2019 24th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*. https://doi.org/10.1109/etfa.2019.8869270

[152] Manic, M., Wijayasekara, D., Amarasinghe, K., Rodriguez-Andina, J.J. (2016). Building energy management systems: The age of intelligent and adaptive buildings. *IEEE Industrial Electronics Magazine* 10(1), 25–39. https://doi.org/10.1109/mie.2015.2513749

[153] Tushar, W., Wijerathne, N., Li, W.T., Yuen, C., Poor, H.V., Saha, T.K., Wood, K.L. (2018). Internet of things for green building management: Disruptive innovations through low-cost sensor technology and artificial intelligence. *IEEE Signal Processing Magazine* 35(5), 100–110. https://doi.org/10.1109/msp.2018.2842096

[154] Jung, W., Jazizadeh, F. (2019). Human-in-the-loop hvac operations: A quantitative review on occupancy, comfort, and energy-efficiency dimensions. *Applied Energy* 239, 1471–1508. https://doi.org/10.1016/j.apenergy.2019.01.070

[155] Erickson, V.L., Carreira-Perpinan, M. A., Cerpa, A.E. (2011). Observe: Occupancy-based system for efficient reduction of hvac energy. *Proceedings of the 10thACM/IEEE international conference on information processing in sensor networks.* 258–269. IEEE.

[156] Zeiner H., Paletta L., Unterberger R. (2021). Cognitive Living Spaces by using IoT Devices and Ambient Biosensor Technologies . In: Ayaz H., Asgher, U., Paletta, L. (eds.) *Advances in Neuroergonomics and Cognitive Engineering – Proceedings of the AHFE 2021 Conferences on Neuroergonomics and Cognitive Engineering, Industrial Cognitive Ergonomics and Engineering Psychology, and Cognitive Computing and Internet of Things*. https://doi.org/10.1007/978-3-030-80285-1_47

[157] Wolkoff, P. (2018). Indoor air humidity, air quality, and health–an overview. *International journal of hygiene and environmental health* 221(3), 376–390. https://doi.org/10.1016/j.ijheh.2018.01.015

[158] Tanir, F., & Mete, B. (2022). Impacts of the Indoor Air Quality on the Health of the Employee and Protection against These Impacts. *IntechOpen*. https://doi.org/10.5772/intechopen.102708

[159] Jiang, C., Masood, M.K., Soh, Y.C., Li, H. (2016). Indoor occupancy estimation fromcarbon dioxide concentration. *Energy and Buildings* 131, 132–141. https://doi.org/10.1016/j.enbuild.2016.09.002

[160] Chen, Z., Jiang, C., Xie, L. (2018). Building occupancy estimation and detection: A re-view. *Energy and Buildings* 169, 260–270. https://doi.org/10.1016/j.enbuild.2018.03.084

[161] Diamond A. (2013). Executive functions. *Annual review of psychology*, *64*, pp. 135–168. https://doi.org/10.1146/annurev-psych-113011-143750

[162] O'Donovan, J., & Smyth, B. (2005). Trust in Recommender Systems. *Proceedings of the 10th International Conference on Intelligent User Interfaces*, 167–174. https://doi.org/10.1145/1040830.1040870

[163] Cheng, M., Yin, C., Zhang, J., Nazarian, S., Deshmukh, J., & Bogdan, P. (2021). A General Trust Framework for Multi-Agent Systems. *Proceedings of the 20th International Conference on Autonomous Agents and MultiAgent Systems*, 332–340.

[164] Ramchurn, S. D., Jennings, N. R., Sierra, C., & Godo, L. (2004). Devising a trust model for Multi-Agent Interactions Using Confidence and Reputation. *Applied Artificial Intelligence*, 18(9–10), 833–852. https://doi.org/10.1080/0883951049050904509045

[165] Sabater, J., & Sierra, C. (2001). REGRET: Reputation in Gregarious Societies. *Proceedings of the 5th International Conference on Autonomous Agents*, 194–195. https://doi.org/10.1145/375735.376110

[166] Yu, B., & Singh, M. P. (2003). Detecting Deception in Reputation Management. *Proceedings of the Second International Joint Conference on Autonomous Agents and Multiagent Systems*, 73–80. https://doi.org/10.1145/860575.860588

[167] Muller, G., & Vercouter, L. (2005). Decentralized Monitoring of Agent Communications with a Reputation Model. In: R. Falcone, S. Barber, J. Sabater-Mir, & M. P. Singh (Eds.), *Trusting Agents for Trusting Electronic Societies*, 144–161.

[168] Sierra, C., & Debenham, J. (2005). An Information-Based Model for Trust. *Proceedings of the Fourth International Joint Conference on Autonomous Agents and Multiagent Systems*, 497–504. https://doi.org/10.1145/1082473.1082549

[169] Sabater-Mir, J., Paolucci, M., & Conte, R. (2006). Repage: REPutation and ImAGE Among Limited Autonomous Partners. *Journal of Artificial Societies and Social Simulation*, 9(2), 1–3. https://ideas.repec.org/a/jas/jasssj/2005-17-3.html

[170] Pereira, A., Rodrigues, N., Barbosa, J., & Leitão, P. (2013). Trust and risk management towards resilient large-scale Cyber-Physical Systems. *2013 IEEE International Symposium on Industrial Electronics*, 1–6.

https://doi.org/10.1109/isie.2013.6563837

[171] Psarommatis, F., & Kiritsis, D. (2022). A hybrid Decision Support System for automating decision making in the event of defects in the era of Zero Defect Manufacturing. *Journal of Industrial Information Integration*, 26, 100263. https://doi.org/10.1016/j.jii.2021.100263

[172] Mattioli, J., Pedroza, G., Khalfaoui, S., & Leroy, B. (2022). Combining Data-Driven and Knowledge-Based AI Paradigms for Engineering AI-Based Safety-Critical Systems. In Workshop on Artificial Intelligence Safety (SafeAI).

[173] Heinrich, L. J., Heinzl, A., & Riedl, R. (2011). *Wirtschaftsinformatik: Einführung und Grundlegung*. Springer. https://doi.org/10.1007/978-3-642-15426-3

[174] Wilde, T., & Hess, T. (2006). *Methodenspektrum der Wirtschaftsinformatik: Überblick und Portfoliobildung* (Arbeitsbericht No. 2).

[175] Karagiannis, D., Buchmann, R., Burzynski, P., Reimer, U., & Walch, M. (2016). Fundamental Conceptual Modeling Languages in OMiLAB. *Domain-Specific Conceptual Modeling: Concepts, Methods and Tools*, 3–30. https://doi.org/10.1007/978-3-319-39417-6_1

[176] Heinrich, L. J., Heinzl, A., & Riedl, R. (2011). *Wirtschaftsinformatik*. Springer-Lehrbuch. https://doi.org/10.1007/978-3-642-15426-3

[177] OMILAB Team. (2020). A Digital Innovation Environment powered by Open Models Laboratory (1.0). Zenodo. https://doi.org/10.5281/zenodo.3899990

[178] Karagiannis, D., Buchmann, R. A., & Utz, W. (2022). The OMiLAB Digital Innovation environment: Agile conceptual models to bridge business value with Digital and Physical Twins for Product-Service Systems development. *Computers in Industry*, 138, 103631. https://doi.org/10.1016/j.compind.2022.103631

[179] Muck, C., & Palkovits-Rauter, S. (2021). Conceptualizing design thinking artefacts: the Scene2Model storyboard approach. *Domain-Specific Conceptual Modeling: Concepts, Methods and ADOxx Tools*, 567-587. https://doi.org/10.1007/978-3-030-93547-4_25

[180] Grieves, M. (2022). Intelligent digital twins and the development and management of complex systems. *Digital Twin*, 2, 8. https://doi.org/10.12688/digitaltwin.17574.1

[181] https://www.boc-group.com/en/adonis/

[182] Bourdieu, P. (1982). Der Sozialraum und seine Transformationen. *Die feinen Unterschiede – Kritik der gesellschaftlichen Urteilskraft,* 171–210.

[183] Giddens, A. (1984). *Die Konstitution der Gesellschaft. Grundzüge einer Theorie der Strukturierung*

[184] Olbrych, S., Kemmerling, M., Zhou, H.A., Lütticke, D., Schmitt, R.H. (2023). A Conceptual Framework for Production Process Parameter Optimization with Modular Hybrid Simulations. In Masci, P., Bernardeschi, C., Graziani, P., Koddenbrock, M., Palmieri, M. (Eds.), *Software Engineering and Formal Methods. SEFM 2022 Collocated Workshops. SEFM 2022. Lecture Notes in Computer Science*, 13765. Springer. https://doi.org/10.1007/978-3-031-26236-4_2