# Highlights

## Identifying Key AI Challenges in Make-To-Order Manufacturing Organisations: A Multiple Case Study

Jonatan Flyckt, Tony Gorschek, Daniel Mendez, Niklas Lavesson

- AI challenges for make-to-order manufacturers are identified via multiple case study.
- AI tasks for operational effectiveness and efficiency are identified and prioritised.
- Process control and customer behaviour emerge as key AI areas.
- Explainable AI is highlighted as important to ensure user trust in AI systems.
- Most AI tasks feasible with current data and technologies, but open challenges exist.

Identifying Key AI Challenges in Make-To-Order Manufacturing Organisations: A Multiple Case Study

Jonatan Flyckt[a,b,*], Tony Gorschek[b,c], Daniel Mendez[b,c], Niklas Lavesson[b]

[a]*Herenco AB, Skolgatan 24, Jönköping, 553 16, Sweden*
[b]*Blekinge Institute of Technology, Valhallavägen 10, Karlskrona, 371 79, Sweden*
[c]*fortiss GmbH, Guerickestraße 25, Munich, 808 05, Germany*

---

**Abstract**

Artificial Intelligence can make manufacturing organisations more effective and efficient, but it is not clear which AI tasks hold the greatest potential. Make-to-order manufacturers must constantly adapt to customers' unique and rapidly changing needs, and therefore have different challenges than make-to-stock manufacturers. Our ambition is to develop an AI-enabled software system to support manufacturing organisations in improving their processes. To this end, we first seek to understand the data and technology requirements for key AI-enabled tasks in a make-to-order setting and determine the level of performance and explainability needed to address them. We perform a multiple case study of five make-to-order packaging manufacturers, interviewing personnel from sales, production, and supply chain to identify and prioritise operational challenges suitable for AI approaches. Demand forecasting emerges as the most important task, followed by predictive maintenance, quality inspection, complex decision risk estimation, and production planning. Participants emphasise the importance of explainable techniques to ensure trust in the systems. The results highlight a need for a greater control of the production process and a better understanding of customer needs. Although most of the tasks could be solved with current techniques, some, such as intermittent demand forecasting and complex decision risk estima-

[⋆]This research was supported by the KKS foundation through the S.E.R.T. Research Profile project www.rethought.se at Blekinge Institute of Technology).

[*]Corresponding author

*Email addresses:* `jonatan.flyckt@bth.se` (Jonatan Flyckt), `tony.gorschek@bth.se` (Tony Gorschek), `daniel.mendez@bth.se` (Daniel Mendez), `niklas.lavesson@bth.se` (Niklas Lavesson)

tion, would require further development. The study clarifies the potential of AI-enabled systems in make-to-order manufacturing and outlines the steps required to realise it.

*Keywords:* multiple case study, artificial intelligence, manufacturing, make-to-order, data requirements

## 1. Introduction

Artificial Intelligence (AI) holds great potential to increase the efficiency and effectiveness for manufacturing organisations by automating routine tasks and improving decision-making. However, little is yet known about the extent to which manufacturing companies make use of AI techniques, and what the capabilities and requirements are for AI-enabled software systems in the domain. Manufacturing companies are increasingly relying on software, and AI-enabled software creates unique challenges when it comes to, for example, required data quality (and quantity) as well as the integration of AI components into existing software systems due to their non-deterministic nature. Organisations face different challenges in adopting data-driven practices depending on their manufacturing strategy: Make-to-order (MTO) manufacturing organisations produce orders when customers place orders, whereas make-to-stock (MTS) organisations produce to stock in anticipation of demands (Akinc and Meredith, 2015). Many MTO companies employ mass customisation, which means they allow customers to customise their products with e.g. specific branding or marketing campaigns in a rapid manner (Pine II et al., 1993). This raises the specific question of how to use AI-enabled systems, for example to optimise the manufacturing processes, but it also creates distinct operational challenges.

To address these challenges, we address the implementation of AI-enabled systems that shall optimise MTO manufacturing systems in our organisational environment. As a first step, we examine how AI can support MTO manufacturing companies through a multiple case study conducted in five packaging manufacturing companies. Our aim is to understand the potential and challenges for the implementation of AI-enabled systems in our partner companies. Previous qualitative studies have studied AI implementation in the manufacturing domain from a business perspective (Dubey et al., 2020; Peretz-Andersson et al., 2024), or seeking to understand the reasons behind the implementation (Hadid et al., 2024; Nabati et al., 2022; Helo and Hao, 2022). Although some studies have focused on mapping the current use of AI technologies (Hadid et al., 2024; Dubey et al., 2020), there are few studies that focus on which key AI tasks manufacturers would implement if they had the ability to do so.

We conduct a series of semi-structured group interviews in five companies, each with participants from production, supply chain, and sales. All five companies employ mass customisation and an MTO manufacturing strategy, facing similar challenges despite differences in customer contracts, manufacturing processes, and physical storage capabilities. The goal of the group interviews is to identify which AI tasks holds the greatest potential to enhance operational effectiveness and efficiency.

Having identified key AI tasks, we then examine the system requirements and challenges associated with implementing these AI solutions: What performance is required from the AI-generated insights to to address challenges and improve current practices? Do AI systems need to provide explanations for their insights to be actionable? What level of integration with other systems is needed?

Because AI systems are heavily dependent on good data, we also explore the characteristics of the available data: How should the data be processed to accurately represent reality for AI algorithms? Are the available data sufficient to address the tasks, or do additional internal or external data need to be collected to solve the tasks?

The following section provides a background of the use of AI in manufacturing, defines relevant terminology, and reviews existing case studies on AI practices in the manufacturing domain. The Research Methodology section details the research questions (RQs), study design, sample selection, data collection, and method of analysis. We then present the results from the group interviews and analyse them from the perspective of each identified AI task. We also assess whether current AI technologies are mature enough to address the goals of the companies. Finally, we summarise the findings and conclude with suggestions for future research.

## 2. Background

AI techniques have been used in the manufacturing domain for decades (Zeba et al., 2021). The main usage has been in tasks such as predictive maintenance, fault detection in production lines, and scheduling and production planning (Chmielarz, 2019; Zeba et al., 2021). Zeba et al. (2021) identifies two distinct time-periods for AI in manufacturing: 1979-2010, which focused mostly on flexible manufacturing systems, data mining, control systems, and decision support, and 2011 and onwards, which expanded to also include areas such as machine learning, genetic algorithms, deep learning, and Internet of Things (IoT).

### *2.1. Terminology*

**AI** has many different definitions. Russell and Norvig (2020) define AI as the study of systems that perceive their environment and perform actions. The definition can be further categorised into systems which act or think either rationally or human-like. In this paper we follow the definition of AI as systems that act rationally to achieve the best outcome, or best expected outcome under uncertainty (Russell and Norvig, 2020). This definition encompasses reactive agents, real-time planners, decision-theoretic systems, and machine learning models. Such systems may involve prediction, combinatorial optimisation, decision making based on expected utility, or descriptive techniques like clustering. We use the term **AI task** to describe a challenge or problem which can be addressed with AI techniques.

Some AI tasks involve **predictions**, where machine learning models are trained on historical data to make estimations for new, unseen inputs (Shmueli, 2010). **Descriptive** machine learning models instead summarise data in various ways (Shmueli, 2010). **Combinatorial optimisation** problems are commonly addressed by using heuristics approximations and constraints together with e.g., genetic algorithms, simulated annealing, or other heuristic optimisation algorithms (Kunapareddy and Allaka, 2020).

AI success is measured using task-specific metrics. Common metrics for machine learning systems are accuracy, recall, and precision for classification tasks, and mean absolute error and mean squared error for regression tasks (Flach, 2012). For tasks that rely on heuristic approximation, success is often assessed by the heuristics themselves, complemented by qualitative analysis where the user's subjective judgement: *"I know it when I see it"* determines whether the results are satisfactory (Luke, 2013). In this study we use the term **performance** as an umbrella term for all these metrics. **Explainable AI** (XAI) can also be used to assess the correctness of machine learning models. With XAI, the model's internal reasoning can be explained either on a per-sample basis (local explanations) or on a per-model basis (global explanations) (Ribeiro et al., 2016). XAI can help users understand why models produce their output, increasing trust in the model (Ribeiro et al., 2016).

AI can be used to support **data-driven decision making**, i.e., basing decisions on the analysis of data instead of relying solely on intuition or experience (Provost and Fawcett, 2013). Decisions that are repeated frequently and at scale are especially interesting for our study as they can improve decision making a lot even from small improvements, but the term data-driven decision making can refer to any decision for which discoveries need to be made within data (Provost and Fawcett, 2013). We focus on AI tasks that improve **efficiency** and **effectiveness**. Effectiveness refers to achieving the desired result, e.g., preventing faulty products from reaching customers. Efficiency involves achieving goals while minimising the spending of resources such as time, money, or energy. Efficiency in manufacturing can be measured for example with **setup times**, i.e., the time it takes to prepare a machine for production (Allahverdi et al., 1999). Overall Equipment Effectiveness (**OEE**) is a common manufacturing metric which measures both efficiency and effectiveness, and includes factors such as machine availability, machine performance, and quality of output (Nakajima, 1988).

Manufacturers can either adopt a **make-to-stock** (MTS) manufacturing strategy, producing material in anticipation of customer demand (Akinc and Meredith, 2015), or a **make-to-order** (MTO) strategy, producing only after the customer places orders (Akinc and Meredith, 2015; Pine II et al., 1993). Other manufacturing strategies exist, but we will focus on challenges related to MTO manufacturing. Many MTO companies employ **mass customisation**, allowing customers to tailor the function or cosmetic appearance of their products (Pine II et al., 1993). This makes stock production challenging, as the produced goods may become obsolete if the customers wish to change their product design. MTO manufacturers typically offer **base products** that customers can customise in various ways. We will use the term **article** to refer to a unique product which is mass customised, either being derived from a base product or being a unique design. Although some of the companies in this case study sometimes take risks to manufacture goods in anticipation of orders or have **safety stock** agreements set up with customers where they keep small stocks of materials, their main strategy is MTO.

### *2.2. Related work*

Previous research on AI in the manufacturing domain, including both case studies (Peretz-Andersson et al., 2024; Jiang et al., 2024; Helo and Hao, 2022; Nabati et al., 2022) and surveys (Dubey et al., 2020; Hadid et al., 2024) have taken both an organisational and technical focus, typically with an MTS perspective. Peretz-Andersson et al. (2024) focused on the resource orchestration perspective in their multiple case study examining Swedish manufacturing SMEs, and Dubey et al. (2020) studied the role of entrepreneurial orientation to provide insights into why companies adopt AI technologies. We aim to complement this organisational focus with a technical focus which addresses both current practices and the goals of implementing AI techniques specifically in MTO organisations.

Several studies have focused on technical aspects of AI in MTS manufacturing organisations: Nabati et al. (2022) sought to understand how manufacturing companies could optimise their energy usage with machine learning technologies. Helo and Hao (2022) investigated the role of AI in supply chain management through an exploratory multiple case study. Hadid et al. (2024) studied the implementation of AI tasks in Japanese manufacturing companies to better understand how AI has been implemented to drive product innovation, and the reasons why some manufacturing companies choose to not invest in AI technologies.

Mass customisation and large variabilities in customer behaviour mean that MTO manufacturing has its own set of unique challenges and requires more adaptive and reactive capabilities. Because previous studies (Hadid et al., 2024; Dubey et al., 2020) have primarily adopted an MTS perspective, and given the unique challenges in MTO manufacturing, we now investigate AI challenges from an MTO perspective. Our goal is to understand which AI tasks would have the biggest effect on operational efficiency and effectiveness, i.e., which AI tasks would MTO manufacturers like to implement if they were able to. Additionally, we investigate the feasibility of the identified tasks from both data quality and state-of-the-art technology perspectives, providing new insights into what challenges would need to be addressed for MTO organisations to achieve their AI goals. To the best of our knowledge, no case studies have focused on AI task identification and requirements analysis for MTO manufacturers specifically.

## 3. Research methodology

We used a qualitative approach with semi-structured group interviews to address the objectives of this multiple case study. Figure 1 shows the workflow of the entire study.

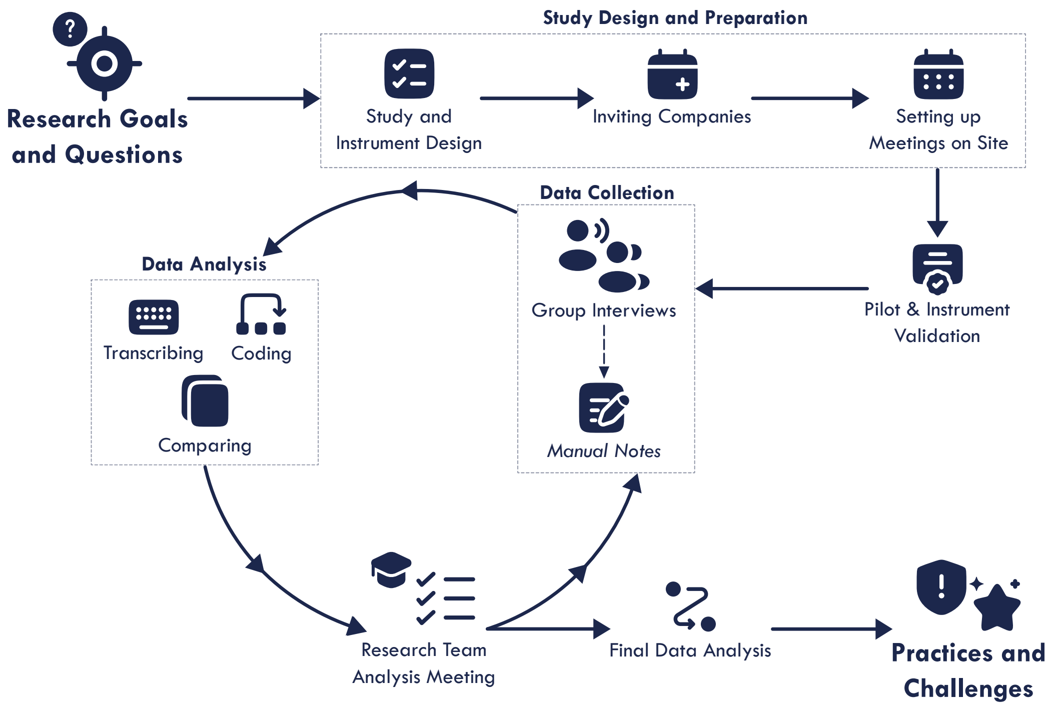


Figure 1: Overview of the study workflow.

### *3.1. Research questions*

The study was guided by the following research questions:

- **RQ1:** Which key AI tasks would MTO manufacturing companies wish to implement to enhance their operational effectiveness and efficiency?
  - **RQ1.1:** How are the tasks currently addressed, and what limitations exist in the present methods and processes?
  - **RQ1.2:** How would the current workflows be transformed with the AI-generated information?
- **RQ2:** What are the expectations and requirements of AI-based systems in terms of performance, explainability, and integration with existing systems?
- **RQ3:** How feasible is the implementation of the identified AI tasks, considering the characteristics of available data and maturity of state-of-the-art AI technologies?

- **RQ3.1:** What data are available currently, and how should they be processed or transformed to accurately represent reality?
- **RQ3.2:** What additional internal or external data need to be collected to support the AI tasks?
- **RQ3.3:** Are current AI technologies mature enough to address the identified tasks, and what challenges need to be addressed to ensure successful implementation?

The goal was to identify how MTO manufacturers can use AI to become more effective and efficient and to determine whether companies and current technologies are mature enough to address the challenges. The first step to achieve this was to identify the areas of operation which could be improved the most with AI, and in what state the company would be after implementing the AI techniques (RQ1). We also sought to understand what the companies expected from the AI systems, and what level of performance was needed to solve the challenges (RQ2), as well as the capabilities of current state-of-the-art AI technologies to address the challenges (RQ3). Because data availability and quality are often crucial to solving AI tasks, we also wanted to understand which data the companies were recording, and what new data collection procedures they would need to introduce (RQ3).

### *3.2. Participating companies and sample selection*

Five companies participated in the study (Table 1). All companies belong to the same corporate group, where each company operates largely independently and with its own management, production sites, customer base, and operational practices. The first author is employed at the parent company, which led to the identification of the initial need for the study: Stakeholders across the subsidiary companies had expressed that it is difficult to realise the potential of AI and to understand which tasks are feasible. All the companies that accepted participation in the study were included.

The companies employ mass customisation to varying degrees, often by purchasing customised semi-finished goods from external suppliers. The mass customisation creates a relatively large product variety as each customisation creates a unique product. The customised products are either produced from customer-specific tools or derived from a range of base products. The companies are all from the private sector, and mainly sell to other businesses, who use the packaging solutions to accommodate their own products. Although we did not assess AI maturity using a specific framework, it can be generally described as follows: The companies are interested in and curious about AI, however, they are still in the early stages of exploration. The use of AI has mainly been through pilot projects in isolated areas, and both technical capabilities and infrastructure are still under development.

| **Company** | **Manufacturing Process** | **Mass Customisation** | **# Employees at company** | **# Unique Articles** | **# Base Products** |
|---|---|---|---|---|---|
| A | Plastic injection moulding | Almost exclusively | ~150 | ~3000 | ~50 |
| B | Plastic injection moulding | On selected products | ~60 | ~400 | ~90 |
| C | Plastic blow moulding | On selected products | ~85 | ~1200 | ~130 |
| D | Plastic & Metal extrusion | Almost exclusively | ~160 | ~1700 | ~140 |
| E | Plastic injection moulding & Metal pressing/printing | Almost exclusively | ~150 | ~2000 | ~90 |

Table 1: Overview of the companies in the study. All companies are packaging manufacturers operating in the private sector.

Employees from the sales, production, and supply chain departments participated in the study. These roles were selected to cover as large a part of the operations as possible while maintaining a relatively small group to allow for natural discussion in the group interviews. The CEO of each company was the first point of contact. In some cases, they recommended suitable participants based on the study description, and in other cases the participants were contacted based on their title/role in the company. The participants were contacted by email, in which the study objective and approach was explained, and two-hour group interviews were conducted. A total of 14 participants partook in the study (Table 2). Because some of the roles were vacant, the most similar roles were selected to participate. For Company A, the Sales and Marketing Manager filled the supply chain role in addition to the sales role, having previously worked in the purchasing department at the company.

Table 2: Group interview participants.
[a]Previously in the purchasing department at the company.

| **Company** | **Role** | **Years at company** |
|---|---|---|
| A | Sales and Marketing Manager[a] | 15 |
| A | Factory Manager | 11 |
| B | Sales and Marketing Manager | 2 |
| B | Production Manager | 3 |
| B | Supply Chain Manager | 28 |
| C | Sales | 25 |
| C | Factory Manager | 5 |
| C | Warehouse Manager | 37 |
| D | Sales and Marketing Manager | 4 |
| D | Factory Manager | 3 |
| D | Strategic Purchaser | 13 |
| E | Key Account Manager | 9 |
| E | Production Controller | 24 |
| E | Supply Chain Manager | 2 |

### *3.3. Study and instrument design*

A multiple case study approach was used to allow for cross-case analysis and to ensure robust data to answer the research questions (Eisenhardt, 1989). We chose to perform semi-structured group interviews using a facilitator from the research team for two main reasons: First, the companies had a varying degree of experience with AI, and the semi-structured approach allowed us to adjust the questions slightly to discover relevant challenges, whereas a more rigid procedure could cause participants to interpret the questions differently. Second: This gave us freedom to ask additional questions discovered during previous interviews or as the discussion progressed during the interview, allowing a better grounding of the theory and providing additional insights (Eisenhardt, 1989). A group interview setting allows participants to both query each other and explain themselves, which offers insight into how much the participants agree, and if they do not, allows for deeper investigation of why (Morgan, 1996). The group interviews were loosely guided by the questions in Table 3, intended to provoke discussion and reflection. The role of the facilitator in the group interviews was to steer the discussion in the direction of AI and data-driven decision-making and to ensure that any questions or misunderstandings could be addressed directly.

Table 3: The questions that guided the group interviews.

| Addresses | Question |
|---|---|
| Context | What is your role and responsibility at the company? |
| Context | How long have you worked for the company? |
| Context | Can you explain briefly how the company operates? |
| RQ1 | What are the main challenges you face which keep you from being effective or efficient? |
| RQ1 | How are the challenges addressed today? |
| RQ1 | What tools (e.g., Excel, ERP system, etc.) do you use to address the challenges in a data-driven manner? |
| RQ1 | What type of manual predictions or estimations do you perform? *By manual, we mean employees using domain knowledge and know-how instead of automated and data-driven approaches.* |
| RQ1 | How accurate are these approaches, and what are the shortcomings? |
| RQ1 | What priority ranking do the identified AI tasks have? |
| RQ2 | How accurate would an AI-based system which addresses the task need to be? |
| RQ2 | Would an AI-based decision support system need to explain its reasoning to be useful? |
| RQ2 | Would the AI system need to integrate with existing systems? |
| RQ2 | What type of output would you expect from an AI system which addresses the task? |
| RQ2 | How would you use the output from an AI system which addresses the task if it were available? |
| RQ3 | What roadblocks exist which could be limiting factors in addressing the challenges with a data-driven approach? |
| RQ3 | What data do you store today relating to the challenges? |
| RQ3 | Where do the data come from and how are they collected? |
| RQ3 | Do the data adequately represent reality? |
| RQ3 | What data do you think are lacking to solve the challenges in a data-driven manner? |
| RQ3 | Could you collect the required data in the organisation today, or would external data sources be needed? |

### *3.4. Data collection*

The data collection comprised a series of two-hour group interviews which were held on-site in four cases, and via remote video meeting in one case. We performed a pilot group interview (Case E) in which the instrument and study designs were fine-tuned for the subsequent group interviews. During this pilot interview we noted that the discussion diverged too much towards non-AI automation techniques or pure system support. We therefore started the subsequent group interviews by showing a simple AI example from the retail industry (Figure 2), and refined the interview questions to more clearly focus on data-driven approaches. The first author was the interview facilitator for all group interviews. All group interviews were conducted in the native tongue of the participants, which was Swedish for four cases and English for one case. Information was captured through note-taking rather than audio recordings, as audio recordings could create an artificial situation that limited participants' expression.

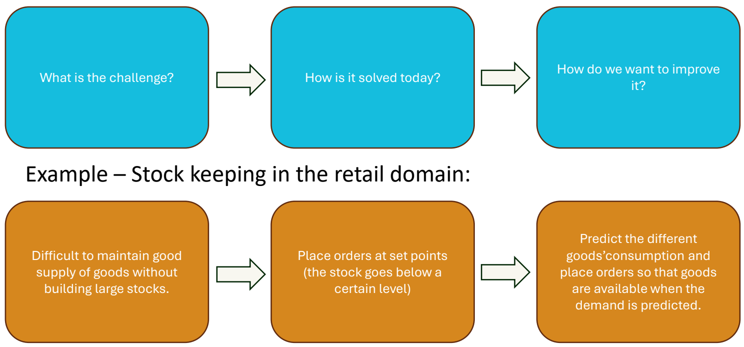


Figure 2: The example AI use case from the retail domain shown to group interview participants before the interviews. The example has been translated from Swedish to English.

Each group interview started with an explanation of the goals of the study and the purpose of the group interview. All participants were informed that all data would be anonymised, both in regard to the company and participant names. The participants were shown the example AI case from the retail industry (Figure 2) to illustrate what information we wanted to extract during the interview: *What is the challenge? - How is it solved today? - How do we want to improve it?*. We also communicated the interview goal of identifying 3-4 AI tasks that could enhance operational efficiency and effectiveness. At the end of the group interviews, the participants were asked to rank the identified tasks from most to least important for their organisation, which allowed for a better understanding of the overall prioritisation of tasks.

After the group interviews, we identified follow-up questions from information that was discovered in the interviews at the other companies. This ensured comparable information between the companies. Based on the questions, we conducted follow-up interviews through remote video chats. These interviews were preceded by emails detailing the gaps of the initial data collection to allow the interviewees to prepare and look up current practices in the organisation if needed. The participants were also asked to send supplementary artefacts (e.g., Excel files with manual demand forecasting) where it was needed for the research team to gain a better understanding of the current way of working. For the cases where the participants were unable to fully answer questions related to data composition or availability, we examined the relevant data sources (e.g., ERP systems) directly. We could then use triangulation (Archibald, 2016) to combine the participant responses with our own analysis and domain expertise to fill in gaps or verify claims, ensuring the answers were grounded not only in participants' perceptions but also in actual data. The participants were given the chance to study the information and conclusions pertaining to their company before the study was finalised.

In addition to the interview data and other artifacts, we conducted a targeted literature search to address RQ3.3, which assesses the maturity of current AI technologies to solve the identified tasks. This search aimed to identify relevant state-of-the-art approaches, review papers, and implementation studies for each AI task. Although this was not a systematic literature search, references were selected based on their relevance as judged by the research team, as well as based on recency and influence within the academic community.

### *3.5. Analysis procedure*

We performed a thematic analysis (Braun and Clarke, 2006) on the notes from the group interviews in which we identified themes that were similar across the interviews. The interview notes were transcribed, and colour-coded per AI task and visualised in a 2-dimensional matrix with the research questions and identified tasks on either axis. This helped us identify which research questions had been answered to satisfaction for each case. The comparative case analysis (Eisenhardt, 1989) also helped us identify information that was present in some of the interviews but absent in others, for example: One company noted that they wanted explanations of how to adjust production parameters from a sensor quality inspection system, whereas the other companies simply noted that they wanted to receive a warning.

The research team met and identified the additional data needed to adequately answer the research questions based on the analysis. From this followed remote video interviews where the participants were asked to fill in the gaps. The analysis procedure was performed iteratively (Figure 1) until the research team was satisfied that the research questions had been fully answered for each company and AI task.

### *3.6. Validity procedure*

We identified several potential threats to the validity of the results, primarily concerning **construct validity**, and implemented measures to mitigate them. One such example is *groupthink*, which can cause the group to gravitate toward a dominant voice, causing potentially relevant information to be missed or omitted (Janis, 1972). Groupthink was mitigated by the interview facilitator actively inviting contrasting opinions and encouraging all participants to critically think about the topics. By showing the participants an AI example at the start of the interview, we risked conceptually *priming* them towards identifying similar tasks (Vaidya et al., 1999). However, the pilot study indicated that this was needed. We attempted to mitigate the effects by picking an example from a different domain (retail). *Evaluation apprehension* involves not disclosing accurate information due to fear of being judged (Zhou et al., 2019). We tried to limit the effects of evaluation apprehension by taking notes instead of recording, as well as informing all participants that the results would be anonymised in the study. Note-taking introduced another construct validity threat by potentially capturing information inaccurately, but we judged it to be more important to mitigate the effects of evaluation apprehension.

An **internal validity** threat to the study is that the research team's conclusions could be biased towards certain findings, causing the interview facilitator to subconsciously lead the discussion towards certain topics. The research team could also introduce a bias towards a specific type of AI task by choosing participants from certain roles while excluding others. We tried to mitigate this by selecting as broad a range of roles involved in daily operations as possible.

An **external validity** limitation is that because all companies in this study have fewer than 200 employees, the findings may not be generalisable beyond that. Larger companies may have a more mature digital infrastructure and different kinds of data and system integration challenges.

## 4. Results and analysis

This section will discuss and analyse the results of the case study from the perspective of the five AI tasks identified during the group interviews: Demand Forecasting, Predictive Maintenance, Quality Inspection, Decision Risk Estimation, and Production Planning. These tasks were not predetermined but emerged inductively during the group interviews (described in the Data collection section). From the answers to the questions in Table 3, specifically those addressing RQ1, the interview facilitator helped steer the discussion from the identified challenges to AI solutions that would address them. After the data collection, the research team categorised similar challenges across the companies into the five AI tasks described here.

We answer all research questions for each AI task subsequently. We cover which AI tasks were identified as most important by the companies, how they solve the task currently, and how the way of working would change if the AI tasks were implemented (RQ1). We also explore the level of performance needed for an implemented task to provide value (RQ2). For RQ3, we discuss the AI tasks from a perspective of data availability and quality, and whether current AI technologies could adequately address the companies' wishes. The data availability and requirements are informed both by the participants' responses during the group interviews and by the research team's direct examination of the data sources, as described in the Data collection section. This approach allowed us to verify claims and fill in gaps through triangulation. Current capabilities of AI technologies are addressed through a literature search by the research team, as detailed in the Data collection section.

### *4.1. Demand forecasting*

#### *4.1.1. Most important AI tasks (RQ1)*

One of the most important areas for all companies in the study is demand forecasting, i.e., predicting when the customers will place orders for finished goods. Companies A, B, and C are forced to swap tools often due to the lack of anticipation of demands for materials using the tools. This reduces machine uptime, as machines remain idle during tool changes. In contrast, the main issue for companies D and E is that they are forced to react in sub-optimal ways because they are unable to anticipate demand patterns, which leads to an unstable production pace.

Low-volume orders affect production efficiency the most because the unique tools require long setup times in relation to the size of the production batches. The companies are forced to staff extra shifts to meet the demands effectively. The Sales and Marketing Manager at Company A expanded: *Most of the articles we sell have a low volume, and these articles cause a lot of headache in the planning. They still need to be produced in a timely manner, so forecasting these low-volume articles would most improve our production efficiency. The high-volume orders have a relatively predictable pattern that is easier for* *us to adapt to today, so these are less important for us to forecast.* Here, demand forecasts could help the companies become more proactive in the production planning.

The companies must sometimes take risks and produce goods before orders are placed by the customers to meet delivery times, but they lack the required decision support from demand forecasts to take the risks effectively and understand the effects on the contribution margin. Demand forecasting has been used in make-to-order manufacturing to inform an understanding of the level of risk associated with keeping items in stock for too long or running out of stock for critical material sub-components (Radke et al., 2013).

Sharing forecasts with customers could create a win-win situation: The customers would understand more about their own purchasing patterns and get better service and deliveries, and the companies would use the demand forecasts to make data-driven decisions to better meet delivery deadlines with fewer resources spent. Sahin and Robinson (2005) found that such practices within MTO companies could lower costs substantially, mainly because the variability and unpredictability in customer demands were lowered when coordinating the customers' ordering with the production processes.

*How are the tasks solved today (RQ1.1).* Larger customers often provide forecasts that are fairly accurate at the yearly level and useful for budgeting: total quantities for all articles aggregated together are 95% correct on a monthly basis, and 99 % correct on a yearly basis. However, these forecasts are usually not accurate enough to support production planning or purchasing decisions: On a monthly basis, the article-level forecasting error is usually between 50 and 60 %. Company B mentioned that although 45 % of their revenue is covered by these forecasts, only 5 % of the total number of articles that they produce are covered.

Sales staff commonly estimate future demand by hand by observing previous year's demand patterns and leveraging knowledge about the customer. For one of the larger customers to Company C, the manual forecasting error was between 15 and 46 % on a monthly basis, which the Factory Manager noted is accurate enough to support some production decisions: *The tricky thing is to run the production as effectively as possible without having many large machine setup processes. To accomplish this, we need to estimate a bit when the customers will place orders. It is not possible to only use the sales orders as a basis for production, as we would fall too much behind the schedule.* Company A reported that their manual forecasts are about 90%

accurate for large-volume articles, and 80% for low-volume articles—allowing for some leniency in the timing. This is enough to take calculated risks in production and purchasing, but the time constraints prohibit them from doing this on a large scale. Here, AI-based demand forecasts could scale in ways not feasible for human employees.

*How would the way of working change (RQ1.2).* Companies A, B, and C want to use demand forecasts to anticipate orders and inform better decisions in purchasing and production. This could reduce raw material costs and increase machine uptime, as similar goods could be produced at the same time without having to swap the tools in the machines as often.

Despite their make-to-order strategy, the companies are sometimes forced to produce towards stock outside of customer agreements. Because producing to stock is a risk, the companies would like to use demand forecasts as a decision support to determine when to do so, as motivated by the Factory Manager at Company A: *If we had forecasts with strong enough performance, we could keep stocks down a lot in addition to keeping a higher machine uptime. Stocks are often produced to the safety stock level, but these materials lie in stock for upwards of a year sometimes. These materials could have been produced much later, but right now production orders are created as soon as we drop below the safety stock levels.*

Companies D and E would use demand forecasts to make the production more stable and predictable, as noted by the Factory Manager at Company D: *Demand forecasting affects basically everything. We can break it down to materials, sub-materials, suppliers, and customers. We would like to know in advance whether the order intake will increase or decrease in general. Then we could set the machine pace based on anticipated customer behaviour instead of the historical machine pace.* Demand forecasting has been found to make production more stable and predictable when including customers in the planning (Sahin and Robinson, 2005), as well as in general by allowing make-to-order manufacturers to anticipate fluctuations and plan ahead to spend resources more optimally (Gansterer, 2015).

#### *4.1.2. Expectation and requirements of AI-based system (RQ2)*

The required forecasting performance varies depending on the task: For decision-making and risk-taking in over- or under-producing towards safety stock, Company A estimated that roughly 90 % correct forecasts would be required, whereas a 50 % correct forecast would be sufficient for high-volume materials, for which orders tend to arrive in a steady flow. Company C estimated that roughly 80 % forecast correctness on a monthly basis is needed to use for production planning, whereas Company B argued that forecasts will be useful as long as they beat the customer-supplied forecasts they use to support production planning today (50-75 % correct). This variation likely stems from the differences in risk that the forecasts would support: For more risky tasks such as safety stock decisions, the companies want to play it safe unless the forecasts are very good. The M4 forecasting competition saw Symmetric Mean Absolute Percentage Error (sMAPE) errors as low as 11-12 % for the best algorithms, which indicates that the companies' wish for 80 - 90 % correctness might be possible (Makridakis et al., 2020). However, the individual characteristics of each time-series greatly affects which performance is possible and required, and because performance expectations were expressed informally during the interviews, the companies may allow for different degrees of relaxation on timing requirements.

Most companies wished to focus on monthly forecasts; weekly forecasts would be too granular. Roughly 85 % of the articles at the companies have intermittent demand patterns on a monthly basis, meaning that the demand occurs sporadically and most of the months have no demand (Syntetos et al., 2005). Predicting the demand quantity would be the most important for most low-volume articles, but demand timing is not unimportant. However, for articles with few other similar articles, the demand timing becomes more important to help reduce setup times by producing all the rare articles in sequence. The Factory Manager at Company D reasoned that *the timing would probably be more important than the actual volume to make the production more efficient. If we know the demand intervals, we can have a predictable production pattern which we can also extend to storage, transport trucks, planning, purchasing, etc.* The Sales Representative at Company C explained that the intermittent patterns mostly exist in low-volume articles: *The larger orders are more top-of-mind, whereas the smaller orders are more difficult to anticipate. If we could get some support telling us when small orders are incoming that would be very helpful. Both the timing and size of demands are equally important.*

A forecasting system should output the predicted customer orders for the following weeks or months. The companies agreed that each article should be forecasted separately, but that it should be possible to aggregate the forecasts to manufacturing tools or product groups for decision support. To fully solve the demand forecasting challenges, there would need to be integration towards the ERP system as well as towards decision support systems that use the forecasts to solve other tasks more efficiently. For example, the forecasts could be integrated into production planning systems to suggest efficient production sequences, or into smart purchasing systems to lower material costs.

Several of the companies argued that it would be nice to have explainable predictions, as this would help ensure user trust in the model and improve decision making. One interviewee noted that this is almost a must-have for people to trust it. The wish for explainable AI is a common trend in organisations adopting AI and is seen as a crucial area of research (Weitz et al., 2023).

#### *4.1.3. Feasibility of identified tasks (RQ3)*

*Data requirements and availability (RQ3.1 and 3.2).* Basic demand forecasting uses historical sales orders to predict future demand. The sales order data are typically aggregated on the desired level of detail (daily, weekly, monthly, etc.), and structured as time-series for each unique article and customer. The companies' ERP systems contain sales order data from several decades back, which need to be filtered to ensure uniform quantities and accurate demand dates.

Customers frequently change their designs, leading to new material numbers in the ERP systems without links to earlier versions. As a result, some of the companies would have to rely on organisational knowledge to reconstruct product histories and build longer time-series for model training. Without consistent version control, models may miss recurring patterns, which could reduce forecast accuracy. The companies also expressed a desire to aggregate forecasts by tools, customers, or material sub-components, but not all relevant data, e.g., product segments, are consistently recorded. Such data gaps are common for organisations starting with AI because it is difficult for non-technical staff to anticipate and store data that may be valuable for AI or other data-driven purposes later (Weber et al., 2023).

Because many of the companies' sales are seasonal and dependent on external factors such as holidays or weather conditions, exogenous variables which express these factors could improve forecasting performance. Customer-provided forecasts could also be incorporated to improve model performance. Although such external factors could improve performance, they would also add a layer of complexity to the overall forecasting architecture, potentially making it more difficult to model interactions with exogenous variables across many different time-series.

*Maturity of AI technologies (RQ3.3).* Demand forecasting can be accomplished with statistical machine learning models such as ARIMA (Box and Jenkins, 1970), or with deep learning models such as LSTMs (Hochreiter and Schmidhuber, 1997; Salinas et al., 2020), CNNs (Chen et al., 2020), and Transformers (Vaswani et al., 2017; Lim et al., 2021). Roughly 10 to 15 % of the companies' demand time-series are smooth, and more than 85 % are intermittent on a monthly basis based on the categorisation criteria defined by Syntetos et al. (2005). Deep learning models can work well for smooth time-series, as well as for intermittent time-series when predicting average demand rates (Salinas et al., 2020; Kourentzes, 2013). Statistical models can also predict average demand rates for intermittent demand (Croston, 1972; Teunter et al., 2011; Shale et al., 2006). However, the companies noted that both the timing and size of demands are important to support decisions. Because current models only support average demand rates, there is a gap between the expectation of the companies and current state-of-the-art.

Time-series models can be both local (trained on and for one single time-series) and global (trained on multiple time-series for use on all of them as single time-series). Local models have traditionally been used in forecasting (Salinas et al., 2020), but recent work has shown that global models can be both more complex and better at generalisation than local models (Montero-Manso and Hyndman, 2021). Because of the several thousand time-series at each company, it may make more sense to construct global or semi-global models which train on all or a subset of all time-series. Global models have also been shown to be able to transfer the knowledge well to single time-series models (Ye and Dai, 2021). Semi-global models have been shown to have superior performance to both local and fully global models in some cases (Bandara et al., 2020; Oriona et al., 2023). However, how to best assign time-series to training clusters is an open challenge which should be addressed in future work.

Both statistical and deep learning approaches can leverage exogenous variables to improve forecasting performance (Lim et al., 2021; Vagropoulos et al., 2016), which several of the companies noted as potentially important. The companies also expressed a desire for models to explain their reasoning to help users trust the predictions. Transformers have been used in time-series forecasting (Lim et al., 2021) and have inherent explainability due to the attention mechanism (Vaswani et al., 2017). Model-agnostic methods like SHAP (Lundberg and Lee, 2017) and LIME (Ribeiro et al., 2016) can also explain predictions both on a model scale and for individual predictions. SHAP has been used to explain individual time-series predictions with good results both for univariate (Zhang et al., 2024) and multivariate (Ozyegen et al., 2022) time-series.

Beyond algorithms, the maturity of organisational practices also influences forecasting performance: Kalchschmidt (2012) identifies three theoretical perspectives: universalistic, contingency, and configurational, based on a study of more than 500 global manufacturing companies. The study found that although some best-practice aspects improved outcome regardless of the specific context (universalistic view), better performance was achieved when forecasting practices were tailored to specific organisational conditions, such as which production systems were used or the product complexity and flora (contingency view), and when the practices were implemented coherently as part of a structured, end-to-end forecasting process (configurational view). This suggests that AI-based forecasting systems should be embedded into the organisational workflows and not treated as stand-alone tools.

### *4.2. Predictive maintenance*

#### *4.2.1. Most important AI tasks (RQ1)*

Four out of the five companies have issues with machines breaking down unexpectedly. They want to detect faults proactively and prioritise maintenance using data-driven methods. By predicting machines' maintenance needs based on sensors or throughput rates, the number of machine failures could be reduced, allowing for a more reliable production and higher throughput for the factories. For Company B, roughly 10 % of the machines are unavailable at any given time due to unexpected failures. Predictive maintenance has been shown to increase OEE levels by 10 - 30 % in previous work in e.g., tube filling and foundry moulding (Natanael and Sutanto, 2022; Mohan et al., 2023). However, predictive maintenance is not a silver bullet, as other factors like quality and machine performance also affect OEE (Obiuto et al., 2024).

It is difficult to keep control of everything with the current approaches, as explained by the Production Controller at Company E: *Machine failures can happen due to inexperience or lack of staff, but also because there is a poor system support for maintenance. A machine failure directly affects the subsequent machines in a production line, and eventually it affects delivery and* *customer satisfaction.* A predictive maintenance system could help maintenance crews prioritise which machines should be given the most attention, lowering the dependency on experienced personnel. Because machine failures often occur due to machine parts inadvertently being disturbed during the machine setting process, visual explanations of the predictions could help inexperienced workers understand how failures arise, and how to prevent them.

*How are the tasks solved today (RQ1.1).* Preventive maintenance schedules dictate which machines need service at which time. Company B performs a daily maintenance where machines are cleaned and greased, during which roughly 80 % of faults are found and remedied. Detecting the remaining faults requires continuous, automated monitoring. A few machines are equipped with IoT sensors for manual inspection and threshold-driven anomaly detection, but they require intimate worker knowledge of the machines. It is difficult for inexperienced employees to set correct threshold values for each machine, whereas an anomaly detection predictive maintenance approach could detect faults without prior knowledge about the machine (Givnan et al., 2022). Faults are sometimes found when machine operators notice weird sounds, shaking, or faulty outputs, but the companies found it difficult to estimate the success level of these approaches. These methods are common but suboptimal, and sound alone may be insufficient in noisy, complex environments (Tsuji et al., 2021).

Faulty machines disturb the production plans, and forces planners to react and change the plans, which can lead to sub-optimal decisions. A reactive maintenance can take anything from five minutes to several weeks if new machine parts are needed. This causes uncertainty for the planners who are forced to react to constantly changing conditions, as explained by the Supply Chain Manager at Company B: *If a machine breaks down, we have to spend a long time to figure out what materials to run instead. Many different things need to fit together in the production. The way we react to this is often to bring in staff on weekends to meet delivery deadlines, and we have to order expensive components with express delivery.* Predictive maintenance in the form of e.g., Remaining Useful Life (RUL) has been used to improve production planning before: Maierhofer et al. (2024) and Bencheikh et al. (2022) argue that maintenance predictions should be integrated into the production planning process, and Bencheikh et al. (2022) found that doing so improved scheduling efficiency, reduced unplanned downtime, and optimised resource utilisation.

*How would the way of working change (RQ1.2).* Predictive maintenance would shift companies from reactive to proactive maintenance. When anomalies occur, maintenance teams would be alerted early and could be dispatched to hopefully prevent a total machine failure. RUL estimations would focus on a longer time-horizon, helping production planners take machine health factors into account when scheduling production orders, and enabling maintenance teams to plan their work more efficiently. An increased trust in the machine availability could also lower setup times, as planners would not need to account for potential future machine faults.

Predictive maintenance systems reduce the need for manual diagnosis and enable new forms of coordination, which would alter the role of maintenance teams (Von Enzberg et al., 2020). However, their implementation poses organisational challenges, including the need for clear communication, data-sharing protocols, and worker retraining to build trust in the systems (Ciocoiu et al., 2017). Many SMEs hesitate to adopt predictive maintenance due to limited awareness of its potential or uncertainty about how to transform and store relevant data (Khan et al., 2022). Technical barriers such as outdated machinery, missing sensors, or difficulties in extracting sensor data also play a role.

#### *4.2.2. Expectation and requirements of AI-based system (RQ2)*

A predictive maintenance system would need to send warnings to the maintenance crew in some way, ideally integrated into the current maintenance systems. The requirements vary significantly depending on the purpose: For production planning purposes a 6-week heads-up would be required according to the Supply Chain Manager at Company B, whereas even a 5-minute heads-up on machine failures could be useful according to the Production Manager at the same company. State-of-the-art RUL estimation performance is roughly 15 to 25 % relative to the mean lifecycle length on baseline datasets (Serradilla et al., 2022). However, the performance depends heavily on the type of machine, operational conditions, sensor data quality, and how the predictive task is framed. Any improvement over current practices would be beneficial to the companies.

For the short-term warnings, a precision of roughly 75 to 80 % would be required. The interviewees noted that some false positives would be acceptable, but that too many would undermine trust in the system. State-of-the-art systems achieve precision values ranging from 12 to 99 % depending on how the time-leniency for false and true positives are defined (Carrasco et al., 2021). AUC scores between 0.5 and 0.85 have been attained in previous work (Carrasco et al., 2021), indicating that it should be possible to balance successful fault detection while keeping the number of false positives low.

Explainability is essential: the system should indicate which parameters triggered the alert and how they deviate from normal. Without such explanations the anomaly detection may not be actionable (Pashami et al., 2023). However, explanations are often too focused on what caused the trigger, rather than providing actionable decision support for maintenance workers (Pashami et al., 2023), which motivates further study in the topic. For the longer-range RUL predictions, explanations would mainly serve to increase user trust in the system, as their purpose is to aid in scheduling regular maintenance rather than detecting faults.

*4.2.3. Feasibility of identified tasks (RQ3)*

*Data requirements and availability (RQ3.1 and 3.2).* Although normal wear and tear can be detected by experienced humans when machines make unusual sounds or vibrations, other relevant machine health factors such as electricity spikes or variations in heat, humidity, and pressure are more difficult to detect. IoT sensors could detect such factors, and some such sensors are equipped today, either placed on the machines or internally as a part of the machine infrastructure. However, it is not always clear how to access these internal data, especially for older machines. Newer machines often have APIs that support continuous data extraction.

Capturing machine health factors requires installing sensors, but the companies currently lack systems for storing or streaming sensor data for predictive use. Additionally, some AI approaches require labelled failure events to train supervised models. As a result, there would be a lead time before it becomes possible to train predictive maintenance models on the data. This creates a tension between collecting failure data and avoiding failures. One mitigation strategy is to use near-failures as proxies. Alternatively, many predictive maintenance approaches can use simulation data or unsupervised methods such as autoencoders as a replacement for run-to-failure data (Kim et al., 2021).

Data collection and preparation are often the biggest challenges in predictive maintenance implementation, requiring extensive work to make machine sensor data available, and to clean, merge, and engineer features. According to Achouch et al. (2022), 70–90 % of total project time is spent on these tasks. They argue that a projected return on investment (ROI) should be carefully evaluated against the anticipated data collection and integration costs, which can vary based on sensor strategy, organisational knowledge, and external consulting needs.

*Maturity of AI technologies (RQ3.3).* Predictive maintenance usually relies on sensors which measure important aspects of machine health for each specific machine, where sensor data are generally fused together for increased model performance (Gawde et al., 2024). Predictive maintenance can be approached in several ways:

- Anomaly detection: Predict future sensor values based on historical values, using supervised learning and models such as RNNs and CNNs. If the predicted values fall outside of a predetermined threshold, an alert is triggered (Namuduri et al., 2020). Unsupervised learning with autoencoders can also be used for anomaly detection, in which the autoencoder learns to reconstruct healthy samples during training (Tian et al., 2022), and large reconstruction error during inference indicates anomalies.

- Remaining Useful Life (RUL): Predict the time left until the machine fails as a regression task by connecting machine sensor data to the time left until a machine failure (Babu et al., 2016; Namuduri et al., 2020).

- Classification based on historical failures: Use supervised learning to classify a sensor sample. The goal variable could be e.g., *will a failure occur within a week?* (Namuduri et al., 2020)

The companies want early warnings to prioritise maintenance outside of scheduled intervals. RUL prediction could support this by estimating when failures are likely, whereas unsupervised anomaly detection could provide real-time alerts when machines behave out of the ordinary. Anomaly detection requires less historical data, whereas accurate RUL prediction depends on labelled failure events. Another challenge is machine heterogeneity, as models trained on one machine may not generalise to others. Transfer learning may help by adapting models to new data distributions (Azari et al., 2023).

Common XAI techniques such as SHAP (Lundberg and Lee, 2017) and LIME (Ribeiro et al., 2016) can be used for predictive maintenance to explain predictions on a per-sample basis (Cummins et al., 2024). To help provide actionable decision support on how to adjust machines when something is wrong, counterfactuals could explain which specific factors would have to be changed to make the predictions non-anomalous (Cummins et al., 2024; Gawde et al., 2024). These model-agnostic explanations can be applied post hoc.

### *4.3. Quality inspection*

#### *4.3.1. Most important AI tasks (RQ1)*

Three of the five companies want to improve quality inspection to avoid scrapping batches or shipping faulty goods. Faulty goods shipped to customers incur significantly higher costs than faults detected at the factory (Wang et al., 2024). The Sales and Marketing Manager at Company B wanted quality assessment to be both objective, because consistency across employees is difficult, and automatic, because manual inspections are time-consuming.

Companies C and D noted that product quality varies due to difficulty in tracking process parameters between batches. It can also be an issue of key-person dependency, as noted by the Sales Representative from Company C: *Someone who has been working with the machine for 40 years has an intuition for what it sounds and looks like when it is running right, but what if this person quits or is not available? We have a key-person dependency and would like for it to be more data-driven.* Process parameter tracking across batches has been shown to produce more even results in other manufacturing domains (Chen and Zhao, 2016) and could likely be employed for packaging manufacturing as well.

The number of quality faults and the tolerance for them depends on the type of product and the stability of the production process, and this differs between the companies. Quality inspection was a top priority for both Company C and D (Table 4). Company C noted that 7.5 % of batches have some kind of quality issue. Company D reported 2–6% quality issues depending on the product type; for Company B, the rate was 1

*How are the tasks solved today (RQ1.1).* Simple visual inspection systems are employed at the companies today, but they detect only certain faults and often have low accuracy. Most of the quality inspection is performed manually by humans, e.g., by comparing a finished product to a master sample. However, because these inspections are infrequent, many faulty goods are often produced before the faults are detected. Human assessments are inconsistent both between inspectors and even across repeated evaluations by the same person (Öberg and Åstrand, 2013). AI-based inspection systems in domains like casting have increased detection rates from around 80% with humans to over 99% with automatic inspection (Sundaram and Zeid, 2023).

The Factory Manager at Company C highlighted the need for real-time inspection to minimise time and material waste: *We perform manual controls, as well as some automated controls such as weighing and leak testing the products, but most of these tests are done in two-hour intervals, so if a fault occurs right after an inspection we will waste two hours of production. We would like constant monitoring with both production parameter values from sensors as well as visual quality inspection with cameras. If these could determine when something is out of the ordinary, we could warn the operators and stop the faulty process early. Because our process is quite unstable, the quality can vary quite a bit throughout a production batch, and a lot of these issues are difficult to catch.* Although some machines are equipped with IoT sensors today, neither Company C nor D have automatic tracking of process parameters across batches and rely solely on operator expertise.

*How would the way of working change (RQ1.2).* The Production Manager at Company B noted that an implementation of an accurate visual quality inspection would allow them to provide a more consistent quality to customers and avoid subjective assessments. Several full-time roles could also be freed up to focus on other tasks instead. A solid visual AI-based quality inspection would also scale quality control without needing to hire more staff or compromising on inspection frequency. Implementation of automatic fault detection systems has been shown to reduce reliance on manual inspections, help to onboard new operators more quickly, and enable preventive instead of reactive quality control processes (Tortorella et al., 2023).

With a process parameter anomaly detection system in place, any deviations could be swiftly resolved to avoid faulty production batches, as the Factory Manager from Company D noted: *What would be great is if we could use sensors to find the correct process parameters that ensure a consistent production quality. Today the control is done outside of the production line. Quality inspection is placed at the end of the production line, which means we may produce several hundred products which need to be discarded. If we* *can trace successful production batches to which process parameters were used at the time, we can use that information to help guide us when we manufacture the same goods again.* The company would be able to move from fault detection during or after the manufacturing process to a preventive system which warns before the quality issues happen.

#### *4.3.2. Expectation and requirements of AI-based system (RQ2)*

Two types of quality inspection were discussed during the interviews: vision-based and process parameter-based. Both types would need to provide warnings to operators directly at the machine, as well as integrate towards production follow-up systems to synchronise with production order data. A process parameter-based quality inspection would have similar requirements as a predictive maintenance system: There should be a tolerance range learned or provided for each produced article and a warning when values drift outside the range. The Factory Manager at Company D emphasised that the system would need to provide operators with suggestions on how to adjust parameters to improve the product quality, and that simple warnings would not be enough to make the predictions actionable.

A visual quality inspection system would have to match or exceed the performance of current manual inspections to be trusted. Ideally, it would outperform humans to help prevent faulty goods from reaching customers. Many modern systems already achieve equal or better accuracy than humans, with better efficiency and consistency (Reyna et al., 2022). The companies also want more frequent, ideally continuous, inspections to detect faults earlier and reduce waste and costs. Company B estimated that false positives must remain below 50%, meaning that at least half of the alerts should reflect actual faults. Although task complexity affects the performance possibilities, some computer vision systems have demonstrated near-perfect precision (Singh and Desai, 2023).

#### *4.3.3. Feasibility of identified tasks (RQ3)*

*Data requirements and availability (RQ3.1 and 3.2).* In general, AI-based quality inspection would require a significant investment into sensors and data infrastructure for the companies. Visual quality inspection would primarily rely on camera sensors that capture images of finished products on the production line. Depending on the task, these may include standard colour cameras, thermal/infrared for heat detection, stereo or LIDAR for depth, and ultraviolet or x-ray for identifying defects outside the visual spectrum. Line-scan cameras are typically used for moving conveyor lines. Such installations introduce logistical challenges, including lighting conditions, occlusion, and sensor degradation over time. Selecting the appropriate sensor type requires domain expertise to ensure that fault types can be accurately detected. A significant hurdle is the need for labelled data (e.g., bounding boxes or segmentation masks), which are time-consuming to produce. Because the companies have large product ranges and frequent design changes, labelled data would be even more challenging. As a result, unsupervised learning approaches may be more practical.

Process parameter monitoring would rely on machine settings data collected over time and linked to both the specific article being produced and a corresponding quality metric. This metric could come from human visual inspection or customer returns. Although product returns are tracked today, the companies reported difficulty linking them to specific production batches. Establishing such links would be essential, both to understand more about what causes quality deviations and for data labelling purposes. In addition to process parameters, external sensor readings could be incorporated, with similar data requirements to those described for predictive maintenance in section 4.2.3.

*Maturity of AI technologies (RQ3.3).* Visual quality inspection can be approached through a supervised (Chen et al., 2018), self-supervised (Xu et al., 2022) or unsupervised (Mei et al., 2018) learning fashion. Supervised learning, which typically requires labelled datasets such as segmentation masks or bounding boxes, may be impractical for the companies due to the time and effort required to label faults across many different products. Although consensus labelling can improve quality by requiring multiple experts to agree on a label, it remains costly and labour-intensive (Sagodi et al., 2022).

Unsupervised methods such as autoencoders offer a more scalable alternative by learning to reconstruct non-defective samples and triggering warnings when the reconstruction error is high (Chen et al., 2018). This approach avoids the need for fault annotations and may be particularly suitable when comparing finished goods to master samples or known defect-free images.

Regardless of the algorithm, deploying vision-based inspection in manufacturing environments involves several practical challenges. Sagodi et al. (2022) identify key best practices:

- **Managing expectations:** Overestimating AI capabilities, such as assuming it can detect faults beyond expert ability, or underestimating the need for representative training data.

- **Roles needed:** Cross-functional teams consisting of domain experts, AI and software engineers, hardware specialists, and production IT.
- **Data collection:** The data need to be collected in conditions representative of actual production. This is time-consuming, especially as faults may occur rarely.
- **Data labelling:** When required, fault labelling demands domain expertise and significant effort.

Process parameter quality inspection would have similar implementations and challenges to predictive maintenance, but instead of differentiating between healthy and unhealthy machines, the models would learn to differentiate between good and bad process parameters. This could be done unsupervised with autoencoders or supervised as a regression task by predicting the future values of sensors, triggering warnings if the predictions fall outside a pre-defined range. Explainable predictions would be critical to make the predictions actionable for operators, and this would also require educating operators in how to interpret such outputs. Counterfactual explanations can be used to achieve these goals (Cummins et al., 2024; Gawde et al., 2024), as described in section 4.2.3.

### *4.4. Decision risk estimation*

#### *4.4.1. Most important AI tasks (RQ1)*

Several of the companies need to take risks in production to meet delivery times and other customer needs, but over-producing goods risks the goods becoming obsolete due to customers changing designs, and the companies lack the capabilities to assess the associated risks on a large scale. Some customers have safety stock agreements, but it is a delicate balance to understand when to fill these safety stocks. Instead, they wish to sometimes take calculated risks and produce more material than required towards stock.

Purchasing material sub-components involves many parameters, where complex decision risk estimation could help, as the Sales and Marketing Manager from Company A reflected on: *Everything has become very complex as we have grown as a company, and we lack both internal knowledge and system support to deal with it. We need to purchase several thousand unique* *material sub-components, and the only decision support the purchaser has is the orders placed by the customer. If they are to make a guess about future demand, they must take that risk themselves, which they may be wary to do.* Organisational complexity has increased in recent years, and AI has become increasingly more relevant to address these complexities (Saba et al., 2021).

*How are the tasks solved today (RQ1.1).* Risks are currently taken sparingly or not at all, explained by the Strategic Purchaser from Company D: *We have taken risks previously with purchasing material sub-components at a lower bulk price from complex price lists. Taking these risks has ended up with poor results, so we avoid it now. If we had some more information about what the customer will do, that would be great.*

Company A uses ad-hoc human judgement to determine when purchasing more in bulk would be beneficial, based on observations of historical patterns and product design swaps. The success rate for this is high: roughly 96 % for purchasing and 99 % for production. However, many opportunities are missed. The company makes extra purchases for roughly 4 % of purchases but estimate that they should do so for up to 30 %. Additional opportunities also exist in production, where they estimate that they could run extra material for as many as 50 % of high-volume articles but only do so for roughly 15 % currently. Human-only decision making tends to lead to missed opportunities and over-conservative approaches, as humans have cognitive biases which prevent accurate risk assessments (Carter et al., 2007). An AI-based system could overview all purchases and make data-driven decisions, potentially failing more often, but overall lowering costs due to finding the opportunities better.

*How would the way of working change (RQ1.2).* With accurate decision risk estimations on a large scale, the companies would be able to take risks in more cases and therefore improve efficiency. Human employees would shift towards an oversight role focusing on verifying the output of the AI system (Braun et al., 2024), and time could be freed up for working on other tasks. The systems would assess the risk of obsolescence from decisions to increase purchasing or production quantities, and would systematically evaluate factors such as customer consistency, design swap frequencies, and historical and forecasted demand patterns.

*4.4.2. Expectation and requirements of AI-based system (RQ2)*

A decision risk estimation system should assess the risk of obsolescence from over-producing or over-purchasing and model the human intuition for such decisions in algorithms. Similar AI systems have been successfully applied in other domains, such as credit risk assessment (Qadi et al., 2021). The system must perform at least as well as humans, but with the added benefit of evaluating hundreds or thousands of cases without bias or cognitive fatigue. This scalability is a key advantage and a recurring theme in this study, as most identified AI tasks are feasible for humans but too time-consuming to perform at scale. AI also offers an opportunity to reduce human bias in risk decisions. The necessary data are available in the ERP system, and aside from data access, no additional system integration is required for these tasks.

Model explainability was considered desirable: one interviewee noted that it is probably enough that explainability exists on the model level to help users understand the logic behind the reasoning to build trust in the system. However, the Sales and Marketing Manager noted that per-sample explanations would allow the user to make more informed decisions, as this would help users identify when the model has made an error due to a lack of context knowledge. Explainable predictions can shine a light on models learning weird or incorrect patterns which become obvious when a human sees the explanations but are difficult to detect solely from the models' predictions (Ribeiro et al., 2016).

*4.4.3. Feasibility of identified tasks (RQ3)*

*Data requirements and availability (RQ3.1 and 3.2).* It is difficult to define any general data requirements for decision risk estimation as a whole, as the tasks identified in the group interviews are quite case-specific. The companies wished to:

1. Estimate the expected monetary outcome of making purchases outside of the current known demand.
2. Estimate the expected monetary outcome of producing materials to stock outside of the current known demand.

The risk of purchasing customer-specific sub-components can be modelled using historical data to estimate the likelihood of obsolescence from product design changes. With no labelled ground truth, the target variable must be inferred from historical design swaps and related features. Model input features describing customer behaviour, such as design change frequency and consistency, can be developed from ERP data. External sources like market trends may also be helpful. Supplier price lists can then be combined with obsolescence probabilities to support purchasing decisions.

Assessing the risk of producing beyond known demand would use similar data but would require estimating the potential reward, such as the value of improved machine uptime. This requires accurate material base data and a correctly modelled setup process. The cost of the surplus finished goods (the risk) can then be weighed against the expected reward and obsolescence probability to guide decisions.

*Maturity of AI technologies (RQ3.3).* Expected Monetary Value (EMV) is commonly used to calculate risks associated with decisions. EMV is calculated by first estimating the probability of success ($P_S$) and the probability of failure ($P_F$), where $(P_S) + (P_F) = 1$. The monetary value of success ($V_S$) and the monetary cost of failure ($V_F$) are also calculated. The EMV can then be calculated with $EMV = P_S \cdot V_S - P_F \cdot V_F$. (Wood and Khosravanian, 2015)

The decision risk estimation tasks are case-specific and rely heavily on feature engineering for both dependent and independent variables. The key challenge is modelling the dependent variable ($P_S$), which can then be predicted using machine learning (Flach, 2012). Machine learning is well suited for this task, as it can evaluate many cases and consider more complex factors than humans. Independent variables such as customer predictability, time since last product change, and average product volume, should be developed in close collaboration with domain experts at the companies.

For both purchasing and production planning, the cost of failure is the value of goods rendered obsolete. For production planning, the model must estimate the monetary value resulting from increased machine uptime. If estimated accurately, the model could help optimise decisions and generate monetary value over time. Interviewees also emphasised the importance of explainable predictions. These would help users detect over-reliance on individual features and reduce trust issues often associated with black-box models. Explainability methods like SHAP (Lundberg and Lee, 2017) and LIME (Ribeiro et al., 2016) can be applied to any trained model on a per-sample basis.

### *4.5. Production planning*

#### *4.5.1. Most important AI tasks (RQ1)*

Two of the companies found it difficult to plan the production and meet delivery times while taking into account factors such as machine availability, order priorities, and inventory levels. Frequent disruptions like machine breakdowns or missing materials force planners to reactively adjust schedules. The Supply Chain Manager at Company B expressed the need for more flexible production planning: *I am interested in improving our flexibility in the planning. A lot of it is done today in the head of the planner, which is not sustainable if we hire new planners who don't have the knowledge. If we could have a system which could react when things happen and re-plan based on the new information, that would be of great help. It is too complex to do optimally manually, even with a long experience planning at the company.*

*How are the tasks solved today (RQ1.1).* Today, planners rely on their experience and basic digital tools to make manual scheduling decisions. This creates a person-dependency on a few experienced employees, which poses a risk for any organisation (Khlaponin et al., 2021). Although interviewees found it difficult to quantify the potential gains from better planning tools, one estimated that better production plans could increase machine uptime by 1–5%. This aligns with findings from related studies (Kozinski et al., 2023), suggesting that the estimate is reasonable. AI-based planning systems can quickly compute near-optimal plans and adapt to changing conditions (Wittrock, 1988; Kunapareddy and Allaka, 2020).

The Key Account Manager at Company E noted that production planning was too focused on solving local issues, instead of looking at the bigger picture: *The production is normally planned to the efficiency of each machine. We would love to be able to take a more holistic view where we avoid the warehouse being filled up inefficiently, potentially waiting for other orders to be completed before finished goods can be shipped.*

*How would the way of working change (RQ1.2).* The planner wants the ability to run the planning tool whenever unforeseen disruptions occur. Sang Chan Park et al. (1997) found that dynamic, flexible planning systems outperform static ones, particularly when disruptions such as unexpected priority orders or machine breakdowns are frequent. The tool's output could serve as a decision support, helping the planners combine system suggestions with their operational knowledge. This would shift their role from reactive to proactive and free up time for other tasks. As a result, delivery accuracy and machine uptime would improve, and dependence on individual experts would decrease. More consistent and reliable planning would also benefit downstream functions like logistics and customer service, enabling more accurate delivery estimates and increasing the company's competitive edge.

### *4.5.2. Expectation and requirements of AI-based system (RQ2)*

A production planning system must perform at least as well as a human planner but at a larger scale. The companies primarily want to automate the production planning to gain a more holistic view and to better handle frequent disruptions. Although off-the-shelf tools and improved data can support some of this, company-specific challenges often go beyond their capabilities. One such example is incorporating demand forecasts or predictive maintenance outputs tailored to the factory's constraints. The system should understand delivery deadlines, machine capabilities, and setup time factors to maximise machine uptime while meeting delivery schedules.

The system must also integrate with the ERP system and potentially connect to production follow-up, maintenance, and forecasting systems. The Supply Chain Manager at Company B noted the need for transparency: if the system delays one order to fulfil three others, the planner must understand the reasoning to assess whether the recommendation should be followed or not. Recent methods show that explainability can be embedded in planning tools, e.g., by using decision trees to interpret genetic algorithm outputs (Wang and Chen, 2024), or structured search trees in reinforcement learning applications (Weichert et al., 2023).

### *4.5.3. Feasibility of identified tasks (RQ3)*

*Data requirements and availability (RQ3.1 and 3.2).* A production planning system that supports frequent re-planning and handles disruptions would need to model the setup process, production chain, and delivery requirements. This requires structured production order data, material master data, and machine data from the ERP system, including which tools fit which machines, and which machines can produce which goods. Additionally, delivery constraints must be clearly defined. Currently, much of this information resides with key individuals rather than in structured systems. Capturing and modelling these data would reduce reliance on individual expertise and enable better production planning.

*Maturity of AI technologies (RQ3.3).* AI-based production planning is a mature field, with decades of research and real-world implementations (Wittrock, 1988). Techniques such as Simulated Annealing, Ant Colony Optimisation, and Genetic Algorithms can solve planning problems using heuristic objective functions (Kunapareddy and Allaka, 2020). The companies want reactive planning that can be re-run as conditions change, which is something that current AI technologies can support. The main challenge lies in accurately modelling the objective function and constraints such as machine capabilities, resource availability, and delivery deadlines. For large, dynamic systems where classical methods may be slow or prone to local optima, reinforcement learning offers an alternative for generating high-quality approximate solutions (Chen et al., 2024).

Objective functions should be developed iteratively with stakeholders to ensure that all constraints are captured, and outcomes align with business needs. In practice, it is often not possible to optimise for only a single objective. For example, planners may need to choose between delivering an order one day late or incurring high setup costs, which requires a trade-off between the two goals. To handle such cases, multi-objective optimisation methods are often necessary. Pareto optimisation can generate a set of solutions without one objective taking over the solution space completely, e.g., as fitness functions in genetic algorithms. This lets stakeholders assess trade-offs case-by-case without needing to predefine weights (Ojstersek et al., 2020). Starting simple and refining the objective function over time encourages planner ownership and feedback, increasing the chance of successful adoption. This approach is supported by Oluyisola et al. (2022), who propose a structured methodology for smart production planning in make-to-order settings, emphasising iterative refinement, system integration, and architectural stability.

### *4.6. Summary of results and implementation considerations*

Table 4 shows the differences and similarities in the priorities for the companies in the study. The companies were asked to rank the importance of each identified AI task in descending order where the number 1 denotes the most important task. Demand forecasting and predictive maintenance were top priorities for most companies. This underscores the companies' desire to predict future patterns to stabilise their processes, improve reliability, and handle unpredictability. An overall theme in the identified tasks is a frustration with not being able to trust plans fully and being surprised by e.g., a sudden large intake of orders or several machines breaking down and halting production. The companies wish to get more control of their process to fulfil promises to the customers more effectively.

Table 4: Identified AI tasks and the prioritisation of their importance across the companies. The tasks are ranked in descending order of importance, where 1 denotes the highest priority.

| **Company** | **Demand forecasting** | **Predictive maintenance** | **Quality inspection** | **Decision risk estimation** | **Production planning** |
|---|---|---|---|---|---|
| A | 2 | 3 | – | 1 | – |
| B | 1 | 3 | 4 | – | 2 |
| C | 2 | 3 | 1 | – | – |
| D | 1 | – | 2 | 3 | – |
| E | 3 | 1 | – | – | 2 |

The lower-priority tasks are the ones that are more specific to the companies' way of working. For example: Quality inspection was prioritised much higher for the companies with a more unstable production process (Companies C and D), whereas decision risk estimations were prioritised highly by the companies with a high degree of mass customisation and a large flora of material sub-components (companies A and D).

According to Hadid et al. (2024), the main motivations for companies adopting AI technologies are to improve operational efficiency, detect defects, and automate processes. Kovič et al. (2024) found that the most adopted AI technologies were related to process management, quality control, and maintenance. These findings largely align with what the companies in this study selected as the most important AI tasks. However, the strong focus on demand forecasting for our companies stands out when compared to the related literature. This can be explained by the make-to-order and mass customisation strategy adopted by the companies, and the many difficulties that arise from supplying several thousand unique articles.

In addition to the technical feasibility of the identified AI tasks, the study revealed several organisational challenges related to data and knowledge that affect implementation. One such example is the lack of versioning data for product series in the ERP systems, which limits the length and continuity of time-series data for demand forecasting. Other examples include missing links between customer returns and specific production batches for quality assurance purposes, as well as incomplete machine or tool meta data for production planning tasks. Addressing such challenges may require companies to adopt master data management (MDM) strategies, improving data governance, and formalising data collection processes. MDM would help create unified master data structures for customers, products, and suppliers, and to consolidate information from different subsystems (Loshin, 2010; Zhao et al., 2020). These issues are not AI-specific, but they are critical prerequisites for successful AI adoption.

We summarise our findings related to the Research questions in Table 5 (RQ1), Table 6 (RQ2), and Table 7 (RQ3).

Table 5: Current and expected transformed approaches for each AI task (RQ1).

| **AI task** | **Current practice (RQ1.1)** | **Transformed practice with AI (RQ1.2)** |
|---|---|---|
| Demand forecasting | Manual forecasting and rough customer forecasts, only accurate in aggregate for budgeting. | Use forecasts to inform production and purchasing, reducing setup times and costs. |
| Predictive maintenance | Preventive scheduled maintenance and ad-hoc manual detection. Limited IoT use. | Use RUL and anomaly detection to plan maintenance and reduce downtime. |
| Quality inspection | Infrequent and sometimes subjective manual inspections. | Automated and continuous inspections using computer vision and process data to reduce errors. |
| Decision risk estimation | Based on human judgement, often avoided. | Model-based risk evaluation at scale to support decisions in purchasing and production. |
| Production planning | Manual planning reliant on human expertise. | Algorithmic scheduling, robust to disruptions and constraints. |

Table 6: Performance, explainability, and integration expectations for each AI task (RQ2).

| **AI task** | **Performance requirements** | **Explainability requirements** | **Integration needs** |
|---|---|---|---|
| Demand forecasting | 50 - 90% correctness depending on use case. Timing & quantity both important. | Desirable for user trust and validation of results. | ERP systems and decision support tools for production and purchasing. |
| Predictive maintenance | $\geq$ 75% precision on warnings while maintaining low false positives. Any gain in RUL estimation useful. | Critical for alerts to be actionable. | Maintenance systems and alert dashboards. |
| Quality inspection | At least as good as manual inspection while false positives $\leq$ 50%. | Very relevant for alerts to be actionable. | Production monitoring systems. |
| Decision risk estimation | At least as good as human judgement but on a larger scale. | Desirable for user trust and validation of results. | ERP systems. |
| Production planning | At least as good as human planners but with more re-planning flexibility. | Desirable for user trust and validation of results. | ERP and planning systems. |

Table 7: Feasibility of implementation for each AI task (RQ3).

| AI task | Available data (RQ3.1) | Additional required data (RQ3.2) | Technology maturity (RQ3.3) |
| --- | --- | --- | --- |
| Demand forecasting | Historical ERP sales data. | Product version history, potentially exogenous parameters, e.g., weather events. | Mature for smooth demand. Lacking for intermittent demand. |
| Predictive maintenance | Maintenance logs, sporadic sensor data. | Additional sensors and machine failure logs. | Mature for both RUL/anomaly detection. Explainability under-explored. |
| Quality inspection | Sporadic process parameter sensor data. | Continuous sensor/camera data, links between batches and faults. | Computer Vision mature. Process parameter maturity similar to predictive maintenance. |
| Decision risk estimation | ERP data on materials, customers, and production. | Behavioural features on customers and defined target features. | Methods mature, but high customisation and feature engineering needed. |
| Production planning | ERP data on production orders and materials. | Modelled constraints and setup times for use in cost functions. | Technically mature. |

## 5. Conclusion and key takeaways

This study has identified key AI challenges for the MTO manufacturing domain. Through a multiple case study, we explored which AI tasks hold the largest potential for increasing effectiveness and efficiency, and what the data and technology requirements are for AI-enabled software systems that address the tasks. A series of group interviews were conducted at five packaging manufacturing companies to identify operational areas where practices could be improved with data-driven approaches.

Among the important AI tasks identified (demand forecasting, predictive maintenance, quality inspection, decision risk estimation, and production planning), demand forecasting emerged as the most important task overall, but the reasons for wanting demand forecasts differed between the companies. This shows that demand forecasting has diverse strategic use-cases for MTO companies, which must constantly adapt to changing customer needs. Overall, the results showed a desire for the companies to gain control over their manufacturing process to serve their customers better and with fewer resources spent. Predictive maintenance and quality inspection being mentioned by several of the companies highlights that the companies want to trust their processes more than they currently do. Addressing these tasks often requires improved data quality or additional data sources, such as sensor data or annotated images, but many tasks can be supported by already existing sales or production data.

A theme throughout all identified tasks was ensuring user trust in AI systems: The companies wished for AI systems capable of explaining their reasoning for them to understand when the system is making informed decisions and when they are lacking sufficient context. The companies further noted that explainable predictions were a necessity to make full use of the decision support in some cases, such as recommending process parameter adjustments for quality inspection. Explainable AI outputs have been shown to benefit from transforming model explanations into more understandable clear-text explanations using Large Language Models (LLMs) (Mavrepis et al., 2024). Future studies could explore how such explanations could be integrated into decision support systems to improve explainability for non-technical users within MTO manufacturing organisations. For many of the identified AI tasks in this study, a human-level performance would be sufficient if the systems performed time-consuming tasks consistently and effectively. However, the companies noted that much better performance would be required in some cases (e.g., demand forecasting and predictive maintenance) to fully solve their challenges.

We further acknowledge that in our study, we focused on small- and medium-sized companies, all operating in the manufacturing sector. In that sense, it is difficult to draw generally applicable conclusions and recommendations that might apply in other context configurations (e.g. companies in different sectors and of different sizes). Future work should therefore investigate to which extent the identified tasks and their prioritisation, as well as data management and system integration capabilities, may or may not differ in different contexts. Moreover, because there is limited research on the implementation of AI tasks specifically in MTO settings, future work should include pilot projects or controlled experiments to quantitatively measure the impact of the proposed AI tasks in operational environments. Additionally, as organisational AI maturity increases, future work could examine aspects of systems engineering such as requirements engineering, test case generation, or product design support for AI systems in an MTO context.

AI techniques are generally mature enough to address the identified challenges. However, some AI tasks would need to be studied further, e.g., demand forecasting models for intermittent time-series which can predict both the timing and size of demands, custom decision risk estimations for complex purchasing tasks, or explainable predictions capable of both detecting issues and recommending actions to solve the issues. Future work should focus on these open issues, and to study how these tasks can be implemented in an MTO context. Addressing the challenges holds the potential to impact decision making, conserve resources, and enable better responsiveness to customer needs in MTO settings. We hope that this study contributes to a better understanding of the potential of AI for MTO manufacturing organisations, and that, as a result, more companies will invest in and build AI competence within their organisations to realise this potential.

*Key takeaways*

- **Crucial to understand customers' behaviours** - Resources are spent unnecessarily to meet customer demands. The companies wish to use AI to predict certain behaviours to avoid being caught off-guard by unexpected orders or changes to the customers' products.

- **Better control over manufacturing processes needed** - Machines break down unexpectedly, and quality control sometimes overlooks faults that AI could potentially detect.

- **Human-level performance required but surpassing it would be desirable** - In most cases, performance on par with humans is sufficient to enhance current processes and free up workers' time. However, the companies also believe that AI could exceed human capabilities and enable better decision making in tasks too complex for humans to handle adequately.

- **Explainable AI should not be an afterthought** - User trust in the AI systems emerged as a recurring theme. Without transparent models and consistent, explainable predictions, users may ultimately avoid using the AI systems.

- **Data quality improvements are needed** - Although the tasks could be partially addressed with existing data, additional data need to be collected and stored, and data quality needs to be improved.

- **Demand forecasting relevant for all companies, but for different reasons** - All companies are interested in demand forecasting, but for different purposes. Some wish to smooth production across all goods for a more predictable workflow, whereas others want to group similar goods proactively to increase machine up-times. The required forecasting performance ranged widely between 50 % and 90 % correct predictions.
- **Current state-of-the-art can solve most identified tasks** - Most of the identified AI tasks can be tackled with existing technologies, but some require further development. Open problems (particularly in time-series analysis and explainable predictions) must be addressed to resolve challenges around demand forecasting and quality inspection, and certain tasks may require custom AI solutions.

### CRediT authorship contribution statement

**Jonatan Flyckt:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Visualization, Writing – Original Draft. **Tony Gorschek:** Conceptualization, Methodology, Writing – Review & Editing. **Daniel Mendez:** Conceptualization, Methodology, Writing – Review & Editing. **Niklas Lavesson:** Conceptualization, Methodology, Writing – Review & Editing.

### Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used ChatGPT to improve the readability and language of the manuscript. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

### Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Jonatan Flyckt is currently employed by the parent company of the companies participating in the study.

## Acknowledgements

This work was supported by the KKS foundation through the S.E.R.T. Research Profile project at Blekinge Institute of Technology. We additionally thank all participating companies and all group interview participants for contributing to the study as well as the reviewers for their valuable feedback that helped strengthen the manuscript.

## Data availability

Due to the competitive nature of the data, the study participants were assured that raw data would remain confidential and would not be shared.

## References

Achouch, M., Dimitrova, M., Ziane, K., Sattarpanah Karganroudi, S., Dhouib, R., Ibrahim, H., Adda, M., 2022. On Predictive Maintenance in Industry 4.0: Overview, Models, and Challenges. Applied Sciences 12, 8081. doi:10.3390/app12168081.

Akinc, U., Meredith, J.R., 2015. Make-to-forecast: customization with fast delivery. International Journal of Operations & Production Management 35, 728–750. doi:10.1108/IJOPM-12-2012-0567.

Allahverdi, A., Gupta, J.N., Aldowaisan, T., 1999. A review of scheduling research involving setup considerations. Omega 27, 219–239. doi:10.1016/S0305-0483(98)00042-5.

Archibald, M.M., 2016. Investigator triangulation: A collaborative strategy with potential for mixed methods research. Journal of Mixed Methods Research 10, 228 – 250. doi:10.1177/1558689815570092.

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

Babu, G.S., Zhao, P., Li, X.L., 2016. Deep convolutional neural network based regression approach for estimation of remaining useful life. Lecture Notes in Computer Science 9642, 214 – 228. doi:10.1007/978-3-319-32025-0_14.

Bandara, K., Bergmeir, C., Smyl, S., 2020. Forecasting across time series databases using recurrent neural networks on groups of similar series: A clustering approach. Expert Systems with Applications 140, 112896. doi:10.1016/j.eswa.2019.112896.

Bencheikh, G., Letouzey, A., Desforges, X., 2022. An approach for joint scheduling of production and predictive maintenance activities. Journal of Manufacturing Systems 64, 546–560. doi:10.1016/j.jmsy.2022.08.005.

Box, G.E.P., Jenkins, G.M., 1970. Time Series Analysis: Forecasting and Control. Holden-Day, San Francisco.

Braun, M., Greve, M., Brendel, A.B., Kolbe, L.M., 2024. Humans Supervising Artificial Intelligence – Investigation of Designs to Optimize Error Detection. Journal of Decision Systems 33, 674–699. doi:10.1080/12460125.2023.2260518.

Braun, V., Clarke, V., 2006. Using thematic analysis in psychology. Qualitative Research in Psychology 3, 77–101. doi:10.1191/1478088706qp063oa.

Carrasco, J., López, D., Aguilera-Martos, I., García-Gil, D., Markova, I., García-Barzana, M., Arias-Rodil, M., Luengo, J., Herrera, F., 2021. Anomaly detection in predictive maintenance: A new evaluation framework for temporal unsupervised anomaly detection algorithms. Neurocomputing 462, 440–452. doi:10.1016/j.neucom.2021.07.095.

Carter, C.R., Kaufmann, L., Michel, A., 2007. Behavioral supply management: a taxonomy of judgment and decision-making biases. International Journal of Physical Distribution & Logistics Management 37, 631–669. doi:10.1108/09600030710825694.

Chen, H., Zhao, Y.F., 2016. Process parameters optimization for improving surface quality and manufacturing accuracy of binder jetting additive manufacturing process. Rapid Prototyping Journal 22, 527–538. doi:10.1108/RPJ-11-2014-0149.

Chen, J., Liu, Z., Wang, H., Nunez, A., Han, Z., 2018. Automatic Defect Detection of Fasteners on the Catenary Support Device Using Deep Convolutional Neural Network. IEEE Transactions on Instrumentation and Measurement 67, 257–269. doi:10.1109/TIM.2017.2775345.

Chen, Y., Ding, J., Chen, Q., 2024. A Reinforcement Learning Based Large-Scale Refinery Production Scheduling Algorithm. IEEE Transactions on Automation Science and Engineering 21, 6041–6055. doi:10.1109/TASE.2023.3321612.

Chen, Y., Kang, Y., Chen, Y., Wang, Z., 2020. Probabilistic forecasting with temporal convolutional neural network. Neurocomputing 399, 491–501. doi:10.1016/j.neucom.2020.03.011.

Chmielarz, G., 2019. Present state and future application of smart technologies in manufacturing processes. Production Engineering Archives 24, 14–19. doi:10.30657/pea.2019.24.04.

Ciocoiu, L., Siemieniuch, C.E., Hubbard, E.M., 2017. From preventative to predictive maintenance: The organisational challenge. Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit 231, 1174–1185. doi:10.1177/0954409717701785.

Croston, J., 1972. Forecasting and stock control for intermittent demands. Journal of the Operational Research Society 23, 289 – 303. doi:10.1057/jors.1972.50.

Cummins, L., Sommers, A., Ramezani, S.B., Mittal, S., Jabour, J., Seale, M., Rahimi, S., 2024. Explainable Predictive Maintenance: A Survey of Current Methods, Challenges and Opportunities. IEEE Access 12, 57574–57602. doi:10.1109/ACCESS.2024.3391130.

Dubey, R., Gunasekaran, A., Childe, S.J., Bryde, D.J., Giannakis, M., Foropon, C., Roubaud, D., Hazen, B.T., 2020. Big data analytics and artificial intelligence pathway to operational performance under the effects of entrepreneurial orientation and environmental dynamism: A study of manufacturing organisations. International Journal of Production Economics 226, 107599. doi:10.1016/j.ijpe.2019.107599.

Eisenhardt, K.M., 1989. Building Theories from Case Study Research. The Academy of Management Review 14, 532. doi:10.2307/258557.

Flach, P., 2012. Machine Learning: The Art and Science of Algorithms that Make Sense of Data. Cambridge University Press.

Gansterer, M., 2015. Aggregate planning and forecasting in make-to-order production systems. International Journal of Production Economics 170, 521–528. doi:10.1016/j.ijpe.2015.06.001.

Gawde, S., Patil, S., Kumar, S., Kamat, P., Kotecha, K., 2024. An explainable predictive maintenance strategy for multi-fault diagnosis of rotating machines using multi-sensor data fusion. Decision Analytics Journal 10, 100425. doi:10.1016/j.dajour.2024.100425.

Givnan, S., Chalmers, C., Fergus, P., Ortega-Martorell, S., Whalley, T., 2022. Anomaly Detection Using Autoencoder Reconstruction upon Industrial Motors. Sensors 22, 3166. doi:10.3390/s22093166.

Hadid, W., Horii, S., Yokota, A., 2024. Artificial intelligent technologies in Japanese manufacturing firms: an empirical survey study. International Journal of Production Research , 1–27doi:10.1080/00207543.2024.2358409.

Helo, P., Hao, Y., 2022. Artificial intelligence in operations management and supply chain management: an exploratory case study. Production Planning & Control 33, 1573–1590. doi:10.1080/09537287.2021.1882690.

Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Computation 9, 1735–1780. doi:10.1162/neco.1997.9.8.1735.

Janis, I., 1972. Victims of groupthink. Houghton, Mifflin, Boston [u.a.].

Jiang, X., Jiang, X., Sun, W., Fan, W., 2024. How Do Manufacturing Firms Manage Artificial Intelligence to Drive Iterative Product Innovation? IEEE Transactions on Engineering Management 71, 6090–6102. doi:10.1109/TEM.2023.3259396.

Kalchschmidt, M., 2012. Best practices in demand forecasting: Tests of universalistic, contingency and configurational theories. International Journal of Production Economics 140, 782–793. doi:10.1016/j.ijpe.2012.02.022.

Khan, M., Ahmad, A., Sobieczky, F., Pichler, M., Moser, B.A., Bukovsky, I., 2022. A Systematic Mapping Study of Predictive Maintenance in SMEs. IEEE Access 10, 88738–88749. doi:10.1109/ACCESS.2022.3200694.

Khlaponin, Y., Izmailova, O., Qasim, N., Krasovska, H., Krasovska, K., 2021. Management risks of dependence on key employees: Identification of personnel, in: CEUR Workshop Proceedings, pp. 295–308.

Kim, D., Lee, S., Kim, D., 2021. An Applicable Predictive Maintenance Framework for the Absence of Run-to-Failure Data. Applied Sciences 11, 5180. doi:10.3390/app11115180.

Kourentzes, N., 2013. Intermittent demand forecasts with neural networks. International Journal of Production Economics 143, 198–206. doi:10.1016/j.ijpe.2013.01.009.

Kovič, K., Tominc, P., Prester, J., Palčič, I., 2024. Artificial Intelligence Software Adoption in Manufacturing Companies. Applied Sciences 14, 6959. doi:10.3390/app14166959.

Kozinski, O., Kotyrba, M., Volna, E., 2023. Improving the Production Efficiency Based on Algorithmization of the Planning Process. Applied System Innovation 6, 77. doi:10.3390/asi6050077.

Kunapareddy, A., Allaka, G., 2020. An improved genetic algorithm for production planning and scheduling optimization problem. Intelligent Manufacturing and Energy Sustainability , 157 – 171doi:10.1007/978-981-15-1616-0_15.

Lim, B., Arık, S.d., Loeff, N., Pfister, T., 2021. Temporal fusion transformers for interpretable multi-horizon time series forecasting. International Journal of Forecasting 37, 1748–1764. doi:10.1016/j.ijforecast.2021.03.012.

Loshin, D., 2010. Master Data Management. Morgan Kaufmann, San Mateo.

Luke, S., 2013. Essentials of Metaheuristics. second ed., Lulu.

Lundberg, S.M., Lee, S.I., 2017. A unified approach to interpreting model predictions, in: Advances in Neural Information Processing Systems, p. 4766 – 4775. doi:10.48550/arXiv.1705.07874.

Maierhofer, A., Trojahn, S., Ryll, F., 2024. Concept for a Robust and Reliable Manufacturing and Logistics System that Combines Production Planning and Control with Predictive Maintenance. Procedia Computer Science 232, 3054–3062. doi:10.1016/j.procs.2024.02.121.

Makridakis, S., Spiliotis, E., Assimakopoulos, V., 2020. The M4 Competition: 100,000 time series and 61 forecasting methods. International Journal of Forecasting 36, 54–74. doi:10.1016/j.ijforecast.2019.04.014.

Mavrepis, P., Makridis, G., Fatouros, G., Koukos, V., Separdani, M.M., Kyriazis, D., 2024. XAI for All: Can Large Language Models Simplify Explainable AI? doi:10.48550/arXiv.2401.13110. arXiv:2401.13110 [cs].

Mei, S., Wang, Y., Wen, G., 2018. Automatic Fabric Defect Detection with a Multi-Scale Convolutional Denoising Autoencoder Network Model. Sensors 18, 1064. doi:10.3390/s18041064.

Mohan, R., Roselyn, J.P., Uthra, R.A., 2023. LSTM based artificial intelligence predictive maintenance technique for availability rate and OEE improvement in a TPM implementing plant through Industry 4.0 transformation. Journal of Quality in Maintenance Engineering 29, 763–798. doi:10.1108/JQME-07-2022-0041.

Montero-Manso, P., Hyndman, R.J., 2021. Principles and algorithms for forecasting groups of time series: Locality and globality. International Journal of Forecasting 37, 1632–1653. doi:10.1016/j.ijforecast.2021.03.004.

Morgan, D.L., 1996. Focus groups. Annual Review of Sociology 22, 129–152. doi:https://doi.org/10.1146/annurev.soc.22.1.129.

Nabati, E.G., Alvela Nieto, M.T., Bode, D., Schindler, T.F., Decker, A., Thoben, K.D., 2022. Challenges of manufacturing for energy efficiency: towards a systematic approach through applications of machine learning. Production 32. doi:10.1590/0103-6513.20210147.

Nakajima, S., 1988. Introduction to TPM: Total Productive Maintenance. Preventative Maintenance Series, Productivity Press.

Namuduri, S., Narayanan, B.N., Davuluru, V.S.P., Burton, L., Bhansali, S., 2020. Review—Deep Learning Methods for Sensor Based Predictive Maintenance and Future Perspectives for Electrochemical Sensors. Journal of The Electrochemical Society 167, 037552. doi:10.1149/1945-7111/ab67a8.

Natanael, D., Sutanto, H., 2022. Machine Learning Application Using Cost-Effective Components for Predictive Maintenance in Industry: A Tube Filling Machine Case Study. Journal of Manufacturing and Materials Processing 6, 108. doi:10.3390/jmmp6050108.

Öberg, A.E., Åstrand, E., 2013. The subjective judgement of weld quality and its effect on production cost, in: Jármai, K., Farkas, J. (Eds.), Design, Fabrication and Economy of Metal Structures, Springer Berlin Heidelberg, Berlin, Heidelberg. pp. 621–626.

Obiuto, N.C., Chiedu, E.O., Chikwendu, O.C., Benjamin, I., 2024. Enhancing Injection Moulding Productivity through Overall Equipment Effectiveness and Total Preventive Maintenance Approach. International Journal of Advances in Engineering and Management 6, 739–747.

Ojstersek, R., Brezocnik, M., Buchmeister, B., 2020. Multi-objective optimization of production scheduling with evolutionary computation: A review. International Journal of Industrial Engineering Computations , 359–376doi:10.5267/j.ijiec.2020.1.003.

Oluyisola, O.E., Bhalla, S., Sgarbossa, F., Strandhagen, J.O., 2022. Designing and developing smart production planning and control systems in the industry 4.0 era: a methodology and case study. Journal of Intelligent Manufacturing 33, 311–332. doi:10.1007/s10845-021-01808-w.

Oriona, Á.L., Manso, P.M., Fernández, J.A.V., 2023. Time series clustering based on prediction accuracy of global forecasting models. doi:10.48550/arXiv.2305.00473.

Ozyegen, O., Ilic, I., Cevik, M., 2022. Evaluation of interpretability methods for multivariate time series forecasting. Applied Intelligence 52, 4727 – 4743. doi:10.1007/s10489-021-02662-2.

Pashami, S., Nowaczyk, S., Fan, Y., Jakubowski, J., Paiva, N., Davari, N., Bobek, S., Jamshidi, S., Sarmadi, H., Alabdallah, A., Ribeiro, R.P., Veloso, B., Sayed-Mouchaweh, M., Rajaoarisoa, L., Nalepa, G.J., Gama, J., 2023. Explainable Predictive Maintenance. doi:10.48550/arXiv.2306.05120.

Peretz-Andersson, E., Tabares, S., Mikalef, P., Parida, V., 2024. Artificial intelligence implementation in manufacturing SMEs: A resource orchestration approach. International Journal of Information Management 77, 102781. doi:10.1016/j.ijinfomgt.2024.102781.

Pine II, B., Victor, B., Boynton, A., 1993. Making mass customization work. Harvard Business Review 71.

Provost, F., Fawcett, T., 2013. Data Science and its Relationship to Big Data and Data-Driven Decision Making. Big Data 1, 51–59. doi:10.1089/big.2013.1508.

Qadi, A.E., Diaz-Rodriguez, N., Trocan, M., Frossard, T., 2021. Explaining Credit Risk Scoring through Feature Contribution Alignment with Expert Risk Analysts. doi:10.48550/arXiv.2103.08359.

Radke, A.M., Tolio, T., Tseng, M.M., Urgo, M., 2013. A risk management-based evaluation of inventory allocations for make-to-order production. CIRP Annals 62, 459–462. doi:10.1016/j.cirp.2013.03.132.

Reyna, M., Delgado, G., Akundi, A., Luna, S., Chumacero, E., 2022. Product Digital Quality Inspection using Machine Vision Systems – A Categorical Review, in: 2022 17th Annual System of Systems Engineering Conference (SOSE), IEEE, Rochester, NY, USA. pp. 37–42. doi:10.1109/SOSE55472.2022.9812687.

Ribeiro, M.T., Singh, S., Guestrin, C., 2016. "why should i trust you?" explaining the predictions of any classifier, in: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, p. 1135 – 1144. doi:10.1145/2939672.2939778.

Russell, S., Norvig, P., 2020. Artificial Intelligence: A Modern Approach (4th Edition). Pearson.

Saba, D., Sahli, Y., Hadidi, A., 2021. The Role of Artificial Intelligence in Company's Decision Making. Springer International Publishing. pp. 287–314. doi:10.1007/978-3-030-52067-0_13.

Sagodi, A., Schniertshauer, J., Van Giffen, B., 2022. Engineering AI-Enabled Computer Vision Systems: Lessons From Manufacturing. IEEE Software 39, 51–57. doi:10.1109/MS.2022.3189904.

Sahin, F., Robinson, E.P., 2005. Information sharing and coordination in make-to-order supply chains. Journal of Operations Management 23, 579–598. doi:10.1016/j.jom.2004.08.007.

Salinas, D., Flunkert, V., Gasthaus, J., Januschowski, T., 2020. Deepar: Probabilistic forecasting with autoregressive recurrent networks. International Journal of Forecasting 36, 1181–1191. doi:10.1016/j.ijforecast.2019.07.001.

Sang Chan Park, Raman, N., Shaw, M., 1997. Adaptive scheduling in dynamic flexible manufacturing systems: a dynamic rule selection approach. IEEE Transactions on Robotics and Automation 13, 486–502. doi:10.1109/70.611301.

Serradilla, O., Zugasti, E., Rodriguez, J., Zurutuza, U., 2022. Deep learning models for predictive maintenance: a survey, comparison, challenges and prospects. Applied Intelligence 52, 10934–10964. doi:10.1007/s10489-021-03004-y.

Shale, E., Boylan, J., Johnston, F., 2006. Forecasting for intermittent demand: The estimation of an unbiased average. Journal of the Operational Research Society 57, 588 – 592. doi:10.1057/palgrave.jors.2602031.

Shmueli, G., 2010. To Explain or to Predict? Statistical Science 25. doi:10.1214/10-STS330.

Singh, S.A., Desai, K.A., 2023. Automated surface defect detection framework using machine vision and convolutional neural networks. Journal of Intelligent Manufacturing 34, 1995–2011. doi:10.1007/s10845-021-01878-w.

Sundaram, S., Zeid, A., 2023. Artificial Intelligence-Based Smart Quality Inspection for Manufacturing. Micromachines 14, 570. doi:10.3390/mi14030570.

Syntetos, A.A., Boylan, J.E., Croston, J.D., 2005. On the categorization of demand patterns. Journal of the Operational Research Society 56, 495–503. doi:10.1057/palgrave.jors.2601841.

Teunter, R.H., Syntetos, A.A., Babai, M.Z., 2011. Intermittent demand: Linking forecasting to inventory obsolescence. European Journal of Operational Research 214, 606–615. doi:https://doi.org/10.1016/j.ejor.2011.05.018.

Tian, R., Liboni, L., Capretz, M., 2022. Anomaly Detection with Convolutional Autoencoder for Predictive Maintenance, in: 2022 9th International Conference on Soft Computing & Machine Intelligence (ISCMI), IEEE, Toronto, ON, Canada. pp. 241–245. doi:10.1109/ISCMI56532.2022.10068441.

Tortorella, G.L., Anzanello, M.J., Fogliatto, F.S., Antony, J., Nascimento, D., 2023. Effect of Industry 4.0 technologies adoption on the learning process of workers in a quality inspection operation. International Journal of Production Research 61, 7592–7607. doi:10.1080/00207543.2022.2153943.

Tsuji, K., Imai, S., Takao, R., Kimura, T., Kondo, H., Kamiya, Y., 2021. A machine sound monitoring for predictive maintenance focusing on very low frequency band. SICE Journal of Control, Measurement, and System Integration 14, 27–38. doi:10.1080/18824889.2020.1863611.

Vagropoulos, S.I., Chouliaras, G., Kardakos, E., Simoglou, C., Bakirtzis, A., 2016. Comparison of sarimax, sarima, modified sarima and ann-based models for short-term pv generation forecasting, in: 2016 IEEE International Energy Conference, ENERGYCON 2016, pp. 1–6. doi:10.1109/ENERGYCON.2016.7514029.

Vaidya, C.J., Monti, L.A., Gabrieli, J.D.E., Tinklenberg, J.R., Yesavage, J.A., 1999. Dissociation between two forms of conceptual priming in alzheimer's disease. Neuropsychology 13, 516 – 524. doi:10.1037/0894-4105.13.4.516.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I., 2017. Attention is all you need, in: Proceedings of the 31st International Conference on Neural Information Processing Systems, p. 6000–6010.

Von Enzberg, S., Naskos, A., Metaxa, I., Köchling, D., Kühn, A., 2020. Implementation and Transfer of Predictive Analytics for Smart Maintenance: A Case Study. Frontiers in Computer Science 2, 578469. doi:10.3389/fcomp.2020.578469.

Wang, H., Wang, D., Fang, Z., Zeng, X.J., 2024. Impacts of inspection errors and warranty costs on a complex product supply chain under different development modes. International Journal of Production Economics 272, 109229. doi:10.1016/j.ijpe.2024.109229.

Wang, Y.C., Chen, T., 2024. Adapted techniques of explainable artificial intelligence for explaining genetic algorithms on the example of job scheduling. Expert Systems with Applications 237, 121369. doi:10.1016/j.eswa.2023.121369.

Weber, M., Engert, M., Schaffer, N., Weking, J., Krcmar, H., 2023. Organizational Capabilities for AI Implementation—Coping with Inscrutability and Data Dependency in AI. Information Systems Frontiers 25, 1549–1569. doi:10.1007/s10796-022-10297-y.

Weichert, D., Kister, A., Volbach, P., Houben, S., Trost, M., Wrobel, S., 2023. Explainable production planning under partial observability in high-precision manufacturing. Journal of Manufacturing Systems 70, 514–524. doi:10.1016/j.jmsy.2023.08.009.

Weitz, K., Dang, C.T., André, E., 2023. Do We Need Explainable AI in Companies? Investigation of Challenges, Expectations, and Chances from Employees' Perspective. doi:10.48550/arXiv.2210.03527.

Wittrock, R.J., 1988. An adaptable scheduling algorithm for flexible flow lines. Operations Research 36, 445 – 453. doi:10.1287/opre.36.3.445.

Wood, D.A., Khosravanian, R., 2015. Exponential utility functions aid upstream decision making. Journal of Natural Gas Science and Engineering 27, 1482–1494. doi:10.1016/j.jngse.2015.10.012.

Xu, R., Hao, R., Huang, B., 2022. Efficient surface defect detection using self-supervised learning strategy and segmentation network. Advanced Engineering Informatics 52, 101566. doi:10.1016/j.aei.2022.101566.

Ye, R., Dai, Q., 2021. Implementing transfer learning across different datasets for time series forecasting. Pattern Recognition 109, 107617. doi:10.1016/j.patcog.2020.107617.

Zeba, G., Dabić, M., Čičak, M., Daim, T., Yalcin, H., 2021. Technology mining: Artificial intelligence in manufacturing. Technological Forecasting and Social Change 171, 120971. doi:10.1016/j.techfore.2021.120971.

Zhang, Y., Sun, Q., Qi, D., Liu, J., Ma, R., Petrosian, O., 2024. Shaptime: A general xai approach for explainable time series forecasting, in: Arai, K. (Ed.), Intelligent Systems and Applications, Cham. pp. 659–673.

Zhao, C., Ren, L., Zhang, Z., Meng, Z., 2020. Master data management for manufacturing big data: a method of evaluation for data network. World Wide Web 23, 1407–1421. doi:10.1007/s11280-019-00707-8.

Zhou, X., Zhai, H.K., Delidabieke, B., Zeng, H., Cui, Y.X., Cao, X., 2019. Exposure to ideas, evaluation apprehension, and incubation intervals in collaborative idea generation. Frontiers in Psychology 10. doi:10.3389/fpsyg.2019.01459.