5-11-2023

# ARTIFICIAL INTELLIGENCE-AUGMENTED DECISION MAKING IN SUPPLY CHAIN MONITORING: AN ACTION DESIGN RESEARCH STUDY

Savindu Herath Pathirannehelage
*ETH Zurich*, sherath@ethz.ch

Jóhann G. Jóhannsson
*ETH Zurich*, jjohannsson@student.ethz.ch

Yash Raj Shrestha
*ETH Zurich*, yshrestha@ethz.ch

Georg F. von Krogh
*ETH Zurich*, gvkrogh@ethz.ch

# ARTIFICIAL INTELLIGENCE-AUGMENTED DECISION MAKING IN SUPPLY CHAIN MONITORING: AN ACTION DESIGN RESEARCH STUDY

*Research Paper*

Herath Pathirannehelage, S.*, ETH Zurich, Switzerland, sherath@ethz.ch

Johannsson, J. G.*, ETH Zurich, Switzerland, jjohannsson@ethz.ch

Shrestha, Y. R., ETH Zurich, Switzerland, yshrestha@ethz.ch

von Krogh, G. F., ETH Zurich, Switzerland, gvkrogh@ethz.ch

## Abstract

*Organizations are progressively adopting hybrid human-artificial intelligence (AI) systems in decision-making processes, with human decisions being augmented by AI insights. Among the promising AI applications in supply chain monitoring (SCMo) are predictive maintenance systems that predict potential device failures and augment maintenance decisions, allowing for timely and efficient interventions. Despite the growing proliferation of such systems, prescriptive knowledge encompassing technical, business, and organizational aspects on how to design, develop, and deploy them in actual operational environments remains limited. To address this shortcoming of evidence-based design principles in practice, our action design research developed a predictive maintenance system that predicts SCMo device failures and augments the maintenance decisions about those devices. By doing so, we outline generalizable design principles to guide prospective predictive maintenance systems in SCMo.*

*Keywords: AI-augmented decision making, Supply chain monitoring, Design principles, Action design research.*

## 1 Introduction

AI-augmented decision making (AIADM) is an umbrella term that encompasses diverse contexts where AI and humans are technically and behaviorally integrated to extend, enhance, and complement each other in decision-making processes (Keding and Meissner, 2021; Murray et al., 2020; Raisch and Krakowski, 2021). Hybrid human-AI systems have garnered significant interest in the information systems (IS) community, as well as in other disciplines and practices, over the last decade. A new stream of literature has emerged conceptualizing new forms of organizing human-AI systems from different vantage points. For example, Murray et al. (2020) emphasized the conjoined agency of humans and technologies. Lyytinen et al. (2020) proposed the concept of "metahuman systems" as a hybrid of humans and machines learning jointly while mutually reinforcing each other. They argued that such hybrid systems are better at learning than humans and machines individually. Recently, Baird and Maruping (2021) advanced this line of inquiry by discussing the paradigm shift of IS artifacts from passive tools to agentic artifacts that can assume responsibility for ambiguous tasks and produce preferred outcomes under uncertainty. These scholars emphasized the inadequacy of recognizing the primacy of human agency to study these new forms of relationships between humans and *agentic IS*

[*] These authors contributed equally to this work and are listed alphabetically.

*artifacts*, as the agentic primacy is ambiguous or fluid. Against this background, human-AI collaborative decision making poses many interesting puzzles for scholars in diverse disciplines (Bailey et al., 2022; Fuegener et al., 2022; Puranam, 2021; Von Krogh, 2018).

Although there have been various subfields of AI, machine learning (ML) systems have gained much prominence as AI's most ubiquitous branch in recent years (Shrestha et al., 2021). Organizations are increasingly adopting data-driven approaches informed by AI or ML algorithms to automate and/or augment their operational and strategic decisions (Shrestha et al., 2019; Stobierski, 2021). The McKinsey Global Survey on AI attests to the rapid adoption of AI by industries across different business functions and its significant and improving impact on the bottom line (Chui et al., 2021). Among these business functions, AI holds great potential for supply chain management (SCMa), due to its ability to process big data sets, detect complex relationships, and provide visualizations (Alicke et al., 2021). Supply chain monitoring (SCMo) refers to the process of continuously surveilling goods in transit and related activities to detect disruptions in the regulated control environments under which the goods are transported. This is crucial in cold chain management (CCM), as a change in the regulated environment can result in inferior quality, damage, or obsolescence. Such surveillance is carried out in real time via data streams from Internet of Things (IoT)-enabled devices attached to transport containers (Fink, 2020; Tsang et al., 2018). Data gathered through these devices coupled with data analytics has successfully mitigated supply chain disruptions and yielded significant quality improvements (Koot et al., 2021). However, failure of these condition-monitoring devices has severe financial and non-financial consequences for logistics companies and SCMo service providers by disrupting key services. Due to this mission-critical nature, the high cost of failures, and increased reliance on sensing devices, it is crucial to identify solutions for predicting the failure of the devices to take maintenance actions. AI offers enormous potential in building *predictive maintenance systems* to predict device failures and augment corrective/maintenance decisions (Fink, 2020). Accurate predictions and insights rendered through AI systems facilitate apt decisions and actions (Koot et al., 2021).

Despite the growing need for prescriptive design knowledge on developing maintenance systems that leverage AI to predict device failure, the research in this area remains either conceptual (Fink, 2020) or highly technical (Sun et al., 2019; Teo et al., 2019) and fails to recognize the business and organizational aspects within the design, development, and deployment phases. This limitation is of concern to IS scholars as well as practitioners. On the one hand, IS scholars need to understand the interactions between technology (predictive maintenance systems in our case) and its organizational contexts, since IS is a techno-organizational discipline (Padmanabhan et al., 2022). Failing to capture important organizational factors in the design and deployment of predictive maintenance systems leaves a lacuna in the IS field. Moreover, how to design, develop, and deploy predictive maintenance systems in the context of SCMo remains poorly understood. On the other hand, the same lack of knowledge prevents organizations from designing and deploying AI-powered predictive maintenance systems, limiting the potential benefits that AI can bring to their operations. Therefore, our research answers the following research question: *What are the design principles for an AI-based predictive maintenance system that predicts device failures and augments maintenance decisions in SCMo?*

In the following, we present our findings from an action design research (ADR) study conducted with a company that offers real-time SCMo solutions as a service. We design and deploy an AI system for the company that predicts the failure of devices that supervise supply chain networks and augments the maintenance decisions of those devices. We draw on a rich set of primary data for the design and development of the IS artifact, including interviews with the firm's executive members, analysis of their archival data, website, intranet, and field notes from weekly meetings. Additionally, we comprehensively evaluate our artifact through the corresponding proof of concept (POC), proof of value (POV), and proof of use (POU) by conducting field experiments and interviews with users and executives (Nunamaker et al., 2015; Venable et al., 2012).

We describe the artifact, its expected utility, and its design process within the organizational context. Our main contribution is the synthesis of a set of six design principles. These design principles encompass technical and organizational and business aspects relevant for designing and deploying AI-based predictive maintenance systems in SCMo. These principles contribute to IS scholarship studying the design and the intended and unintended organizational consequences of adopting such systems. The principles may also help guide practitioners in developing customized AI-based predictive maintenance systems for their SCMo contexts taking our case as an example. Overall, our ADR study not only fosters richer insights into the interactions of technology and organization (Altendeitering and Guggenberger, 2021), but also performs a dual mission of contributing to theory and providing practical insights.

The remainder of the paper is structured as follows. In Section 2, we review the literature on SCMo and predictive maintenance, AIADM, and artifactual contributions in IS. In Section 3, we describe our ADR method. In Section 4, we elaborate on our problem formulation. In Section 5, we detail the five building, intervention, and evaluation (BIE) cycles we conduct to design, develop, and evaluate the artifact. In Section 6, we formalize our design principles as our contribution to the literature. In Section 7, we conclude by detailing our contributions to theory and practice and avenues for future work.

# 2 Theoretical Background

## 2.1 Supply chain monitoring and predictive maintenance

SCMa builds on the foundational framework of logistics management with the key objective of planning and coordinating the flow of material and information in an integrated and effective manner from the producer to the end user (Seuring and Müller, 2008). Managing supply chains is becoming substantially more challenging due to interlinked physical flows resulting from complex product portfolios, market volatility, and rising sustainability concerns about supply chains that demand optimization and localization (Alicke et al., 2021). CCM in particular demands special attention to goods in transit that are perishable, sensitive, and of high-quality standard, such as pharmaceuticals and food products (Bishara, 2006; Han et al., 2021; Shih and Wang, 2016). Product quality, goodwill, safety, and brand image are key concerns for managers in these industries, and monitoring these important and sensitive products is essential (Tsang et al., 2018). SCMo is central to CCM to sustain the required quality standards by ensuring that goods are consistently maintained in prescribed optimal conditions using refrigeration and dehumidification systems (Bishara, 2006; Tsang et al., 2018). Monitoring helps logistics companies to evaluate supplier risks, increase product visibility and traceability with the related parties, detect disruptions promptly, and act swiftly to avoid subsequent damages (Koot et al., 2021).

Emerging technologies such as AI, IoT, cloud technologies, and (big) data analytics are increasingly adopted across industries (Bailey et al., 2022), including SCMo (Swanson et al., 2018). SCMo employs a suite of technologies, including radio frequency identification (RFID) and IoT to continuously track parameters such as location, temperature, lighting intensity, and humidity in real time (Abad et al., 2009). IoT refers to a structured network of smart, everyday objects equipped with intelligence, identification, and sensing technologies that can exchange data (Yang, 2014). Tsang et al. (2018) proposed an IoT-based risk monitoring system that applies IoT and AI techniques to prevent product loss and industrial accidents. Safety improvements and quality control have also been achieved through SCMo using both RFID and IoT technologies (Abad et al., 2009; Jia et al., 2012). The availability of vast amounts of data collected through IoT/RFID devices has coalesced with AI's ability to distill insights and predictions from these data to foster predictive maintenance and fault prediction systems (Fink, 2020; Kanawaday and Sane, 2018; Sogi and Mittal, 2020). Predictive maintenance helps determine the condition of a device to estimate when maintenance should be performed or when the device should be replaced (Fink, 2020; Kanawaday and Sane, 2018). Fault prediction is a closely related topic that focuses on when a device will fail (Catal and Diri, 2009). Leukel et al. (2021) reviewed the adoption of ML for failure prediction and found that a broad range of systems and domains have successfully adopted ML-based failure prediction solutions; nonetheless, this was without any specific relation to SCMa/SCMo.

## 2.2 AI-augmented decision making

AI is being deployed across a variety of organizational contexts at a remarkable rate to improve decision-making processes (Balakrishnan et al., 2020; Berente et al., 2021; Shrestha et al., 2019). The McKinsey

Global Survey on AI 2021 revealed that more than 57% of companies had adopted AI technologies in at least one business function to address complex decision-making problems (Berente et al., 2021; Chui et al., 2021). As organizations digitalize their operations, they invest in AI with the intention of improving decision making by leveraging AI-driven insights (Balakrishnan et al., 2020).

Recently, significant attention has been paid to structuring decision making in organizations while recognizing the agency of AI artifacts (i.e., when, where, and how to make and integrate decisions involving humans and AI, as well as who should do so) (Baird and Maruping, 2021; Murray et al., 2020; Shrestha et al., 2019). Scholarly attempts to address this issue can be broadly categorized into two streams, decision automation and augmentation (Raisch and Krakowski, 2021). Raisch and Krakowski (2021) defined automation as machines taking over human tasks completely, whereas augmentation refers to humans collaborating with machines (or AI) to perform tasks. The tension between these two concepts is termed the "automation-augmentation paradox." The literature concurring with the decision augmentation thesis is rife. Grønsund and Aanestad (2020) studied the use of AI to either automate or augment human work and concluded that augmentation (human-in-the-loop) could emerge as a strategic capability of firms. Shrestha et al. (2019) proposed a framework outlining how human and AI decision making may be combined to improve organizational decision making based on five contingency factors. Moreover, human-AI symbiosis was profoundly acknowledged by Murray et al.'s (2020) work on the conjoined agency of humans and technologies, Lyytinen et al.'s (2020) "metahuman systems" concept, and Baird and Maruping's (2021) new perspective recognizing the agentic nature of IS artifacts.

One approach to augment decision making using AI in condition monitoring systems is through predictive maintenance and failure prediction (Fink, 2020). Accurate predictions and insights, such as pattern recognition, classification, approximation, optimization, clustering, and process control (Toorajipour et al., 2021), allow for timely maintenance decisions and interventions (Koot et al., 2021). A couple of anecdotes illustrate such use. Amazon deploys fleets of AI-enabled warehouse robots to improve productivity and minimize errors. Goodyear uses AI sensors with IoT in its smart tires to monitor and control tire changes (Fosso Wamba et al., 2021). In the presence of such distributed device networks, leveraging condition monitoring systems may offer substantial benefits by improving uptime and performance, reducing maintenance costs, and signaling the optimal point and type of maintenance intervention to the decision maker (Fink, 2020).

## 2.3 Artifactual contributions in IS research

Ågerfalk and Karlsson (2020) viewed a research contribution as something that advances the understanding of the studied practice by virtue of its novelty and originality. They outlined three types of research contributions: theoretical, empirical, and artifactual. In our literature review, we focus on artifactual contributions of intervention-oriented IS research, such as design science research (DSR) (Hevner et al., 2004), action research (AR) (Baskerville and Wood-Harper, 1998), their combination – ADR (Sein et al., 2011), and collaborative practice research (Mathiassen, 2002). In ADR in particular, the artifacts are not merely IT systems but ensembles (including constructs, design principles, software tools and their features, process models, frameworks, and IS instantiations) shaped by the organizational context throughout their development and use (Sein et al., 2011). Hence, artifactual contributions serve the dual mission of IS research (i.e., advancing theory while responding to practitioner problems) well (Ågerfalk and Karlsson, 2020; Sein et al., 2011). Therefore, artifactual contributions are valued strongly by methodology- and design-oriented IS researchers and often feature in European IS research (Altendeitering and Guggenberger, 2021; Chen et al., 2022; Wache and Dinter, 2021).

Recently, editorials from premier IS journals have shown significant interest in AI-based artifacts (Abbasi et al., 2016; Ågerfalk et al., 2021; Padmanabhan et al., 2022; Rai, 2017). AI has caused a paradigm shift in IS research (Abbasi et al., 2016; Baird and Maruping, 2021), leading to the fundamental question of whether the theories and knowledge created about non-AI artifacts hold for AI-based artifacts (Herath Pathirannehelage et al., 2022). This has opened up research avenues for design science scholars to design and develop novel AI artifacts for prediction or description, decision support, and business process improvements and automation (Abbasi et al., 2016), leading to new constructs, design principles, process models, methods, and IS instantiations. For example, the cross-industry standard process for data mining (CRISP-DM) (Chapman et al., 2000) emerged as the most widespread process model for data analytics and is regarded as the system development lifecycle (SDLC) equivalent to developing traditional information systems (Abbasi et al., 2016). Abbasi et al. (2010) developed a predictive IT artifact following a design science approach. Herath Pathirannehelage et al. (2022) adopted an ADR paradigm to develop a techno-organizational process model for AI-based IS artifact design and deployment while identifying the challenges therein and proposing design principles to overcome them.

Nevertheless, scholarly attention to AI artifacts has predominantly remained conceptual (Fink, 2020) and technical (Bhavana et al., 2019; Sun et al., 2019; Teo et al., 2019) in predictive maintenance literature. Furthermore, AI artifact design and development for predictive maintenance in SCMa/SCMo has so far received scant attention (Mullarkey and Hevner, 2019). This research gap motivates our study to examine the interplay between AI-based predictive maintenance technology and its business and organizational use. We outline design principles that espouse crucial business and organizational aspects such as reconfigurability (Miah et al., 2019), domain knowledge integration, accountability, explainability, algorithm appreciation (Feuerriegel et al., 2022), cost efficiency, and system evaluation (Venable et al., 2012) when operationalizing AI-based predictive maintenance.

# 3 Action Design Research

The purpose of this paper is to develop design principles for an AI-based predictive maintenance system that predicts device failures and augments corrective/maintenance decisions in SCMo. To do so, we carefully followed the stages and principles of ADR conducted at a company offering SCMo as a service (Sein et al., 2011). ADR is a research method for generating prescriptive knowledge by building and evaluating IT artifacts in an organizational setting (Sein et al., 2011). It is a combination of DSR and AR; however, its striking advantage is that the artifact is not solely based on technical design but is also shaped by the organizational context throughout its design, development, and use (Orlikowski and Iacono, 2001). ADR's use in the IS community has increased in recent years (Chen et al., 2022; Golovianko et al., 2021; Mandviwalla and Flanagan, 2021), leading to a growing number of application settings in which it has been applied effectively and consistently (Mullarkey and Hevner, 2019).

ADR speaks to the core purpose of our research, namely developing an IS artifact (an AI-based predictive maintenance system in our case) while allowing for its emergence in the organizational context and evaluating its usefulness. Our ensemble artifact embodies design principles for predictive maintenance systems that can be "fundamental propositions that aid designers in achieving a successful transfer of requirements to design" (Möller et al., 2020, p. 3). We rigorously evaluated our ensemble artifact following the DSR evaluation approach suggested by Venable et al. (2012) and Nunamaker et al. (2015) by assessing its POC, POV, and POU. Therefore, our design principles, the process, and the instantiation should be sufficiently transparent for other SCMo firms to draw on this template to design and develop predictive maintenance systems to suit their contexts.

Following Sein et al. (2011), we proceeded along four stages: (1) problem formulation; (2) BIE; (3) reflection and learning, and (4) formalization of learning. **Problem formulation** identifies and conceptualizes a research opportunity based on existing theories and technologies. It is often triggered by a problem perceived in practice, but it is important to abstract and understand the practitioner problem as an instance of a class of problems to generalize the findings later (see ADR stage 4). The **BIE** phase is carried out as an iterative process in a target environment, and it interweaves the building of the artifact, intervention in the organization, and evaluation. The outcome of this phase is an IT artifact. The problem and the artifact are continually evaluated, and the design principles are articulated for the chosen class of systems. The **reflection and learning** stage emphasizes moving conceptually from building a solution instance to applying the learnings from the building phase to a broader class of problems. **Formalization of learning** means that the learnings from the specific situation of the ADR project should be generalized into general solution concepts. Figure 1 illustrates the research design while the ADR stages, ADR principles and justifications, and data collection in each stage are provided in Table 1.

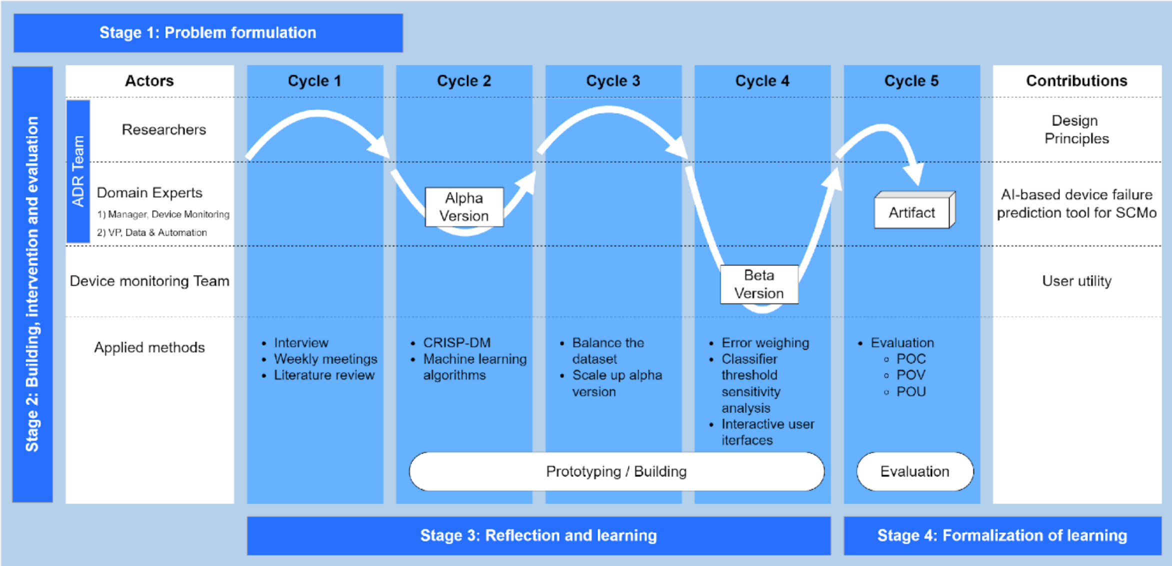


*Figure 1: Research design.*

| Stage | Principle | Justification | Data collection | Artifact |
|---|---|---|---|---|
| **Problem Formulation** | P1: Practice-inspired research | The study was motivated by inadequate knowledge on deploying AIADM in SCMo. | 1) Interview: vice-president (VP) of data and automation<br>2) Weekly meetings: head of device monitoring<br>3) Literature review, see Section 2 | Recognition: Shortcomings of the current SCMo process and lack of prescriptive knowledge to develop and deploy AIADM in SCMo |
| | P2: Theory ingrained artifact | The solution was based on CRISP-DM (Chapman et al., 2000) and pattern detection (Choudhury et al., 2021). | | |
| **BIE** | P3: Reciprocal shaping | Weekly exchanges between researchers and practitioners (ADR team) and artifact evaluations in the organization ensured that the artifact was influenced by both the technological and organizational contexts. | 1) Archival company data<br>2) Weekly meetings: head of device monitoring<br>3) Company website<br>4) Company intranet<br>5) Artifact evaluation: field experiments and interviews | Alpha version: ML models achieving superior predictive performance (AUC scores > 0.85). See Table 2.<br>Beta version: Neural network model with interactive user interfaces (UIs) with error weighting and classifier threshold sensitivity analysis, See Figure 4. |
| | P4: Mutually influential roles | The ADR team is comprised of researchers and practitioners to combine theoretical, technical, and practical perspectives. | | |
| | P5: Authentic and concurrent evaluation | The artifact underwent a threefold evaluation approach: POC, POV, and POU (Nunamaker et al., 2015). | | |
| **Reflection and learning** | P6: Guided emergence | The artifact was continuously reviewed and improved over five BIE cycles (see Figure 1). | 1) Interview: VP of data and automation<br>2) Weekly meetings: head of device monitoring<br>3) Literature review | Emerging version and realization: Beta version included the requirements that emerged in the BIE cycles. See Figure 1. |
| **Formalization of learnings** | P7: Generalized outcomes | A set of generalized design principles was articulated, positioning our case study as an instance. | 1) Interviews: VP of data and automation<br>2) Literature review | Ensemble version: An ensemble embodying design principles for predictive maintenance in SCMo. |

*Table 1. ADR stages, principles, justifications, and data collection in each stage.*

# 4 Problem Formulation

We utilized a unique opportunity to conduct our research in cooperation with a company offering real-time supply chain visibility solutions as a service. This company is fast growing and employs over 300 people. Its offering is the uninterrupted monitoring of goods in transit using IoT devices and providing clients with mission-critical insights, analytics, and operational service. The service seeks to facilitate customers with a better overview of their supply chain and in turn improve quality, reduce waste, and increase efficiency across it. Our case selection was motivated by three key factors. First, in terms of the technical foundation and data availability, the company has a robust IT infrastructure with the potential of collecting and storing massive amounts of data and producing analytics. Second, until the inception of our ADR study, the company had never developed an AI solution internally, making this project an inaugural one. Third, the company was willing to collaborate closely with us and shared our commitment, which is crucial for the success of an ADR project. Thus, the project scope from problem identification to deployment was cooperatively created from a management and operational perspective.

The practical problem that motivated our research was that the monitoring devices could stop collecting data unexpectedly. This might significantly decrease the value of the goods in transit and bring detrimental consequences to the client and the company. The project aimed to significantly reduce the number of devices that malfunctioned in operation by predicting device failure before it went into operation, better understanding the reason for the failure, and augmenting corrective/maintenance decisions at crucial points (predictive maintenance). Maintenance actions can significantly reduce device malfunctioning, saving the company and its clients significant financial and non-financial costs. Since effective decision augmentation relies on the interplay between humans, algorithms, and organizational protocols (Murray et al., 2020), the artifact design should embody acceptance, reconfigurability, accuracy, and interpretability of the algorithm output. In this context, we conceptualize our problem solution class as “AI-based predictive maintenance systems that predict device failures in SCMo.”

# 5 Building, Intervention, and Evaluation (BIE)

We applied the most prevalent process for guiding analytics projects, the CRISP-DM framework (Abbasi et al., 2016; Chapman et al., 2000), in the BIE phase of the ADR process to develop a modeling pipeline in which the models were evaluated and refined to create superior performance. CRISP-DM provided a solid framework for planning and executing our AI-based artifact development. It involves several iterative phases: (1) business understanding, (2) data understanding, (3) data preparation, (4) modeling, (5) evaluation, and (6) deployment (Chapman et al., 2000). We followed these iterative phases over five BIE cycles, as detailed in the following subsections.

## 5.1 BIE cycle 1: Business understanding to data preparation

The business objective of mitigating the adverse consequences of monitoring failures is outlined in Section 4. The technical goal was to predict the failure of monitoring devices using the proprietary datasets of the company using AI/ML algorithms. We defined the problem as a supervised ML binary classification problem (failing or functioning within a given time). We used the company’s relational database, which is a collection of tables with a wide range of variables, ranging from device events and status checks to client information, shipment information, and device measurements. The tables comprise 1.8 million to 33 billion data points per table. We developed an understanding of the variables present in these datasets, and then we used frequency plots for categorical variables and boxplots for continuous variables to explore the distributions of these features. Subsequently, we used the data quality verification metrics proposed by Chen et al. (2014), namely completeness, accuracy, timeliness, consistency, objectivity, security, relevance, clarity, format, convenience, and cost effectiveness, to ensure the quality of the data.

We created a timeline for data selection based on the fundamental idea of predicting future outcomes using past behavior while assuming the constancy of the data-generating process. We defined two periods, with an example of these illustrated in Figure 2. Period 1, the feature window, was the period used to create features that would then be used to predict future outcomes (i.e., failing or not). Period 2, the target window or ground truth, was the period we predicted. The periods spanned 20 days (Period 1) and 10 days (Period 2) to ensure that Period 1 would capture the recent data needed to make a comprehensive prediction and that Period 2 would then include sufficient devices (especially failed devices, as that was a minority class) to build a model. We created a master data table, which included features from Period 1 (feature engineering is described below) and the target variable from Period 2. Each row in the master table represents a device and includes an array of features that could predict the target variable. The master table was then split into a 70% training set and a 30% test set to train the models and test them with unseen data (Choudhury et al., 2021). The data selection described above was refined iteratively through evaluations with the domain experts.

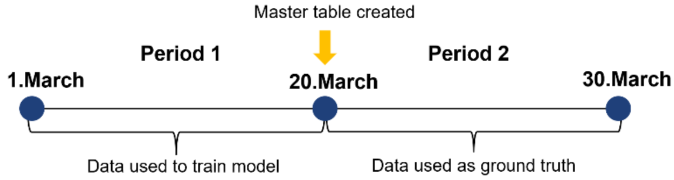


*Figure 2. An example timeline for data selection.*

Feature engineering refers to constructing, deriving, and transforming data into variables that can be used as input features in the model (Chapman et al., 2000). The domain experts and the researchers began by scoping the potential predictors and how they could affect the target variable. Following these discussions, the research team engineered multiple features and used them in modeling (see Section 5.2). We iteratively evaluated the predictive performance of the models based on the engineered features while removing redundant and highly correlated variables (Choudhury et al., 2021).

## 5.2 BIE cycle 2: Modeling the alpha version

The sole objective of the alpha version was to conduct model selection; that is, to develop multiple models using general-purpose ML algorithms, compare their predictive performance, and pick the best ones. We explored general-purpose ML algorithms (Choudhury et al., 2021), logistics regression (Berkson, 1944), decision trees (Quinlan, 1986), random forest (Breiman, 2001), gradient boosting (Friedman, 2001), and neural networks (Rosenblatt, 1958).

Tuning models involves finding the optimal hyperparameters for each model for a given problem and dataset by trying different hyperparameter value combinations and choosing the one that performs best on validation data. We chose the hyperparameters that would maximize *the area under the curve (AUC) of the receiver operating characteristic (ROC)*—our performance metric to evaluate different models—as the optimal hyperparameters with 10-fold cross-validation.

When comparing models, it is crucial to choose a model performance metric that is appropriate for the problem and objective. Different metrics can work better in different settings (Choudhury et al., 2021). Popular metrics for classification problems include log-loss score, accuracy (percentage of correct predictions for the test data), precision (proportion of positive examples that are truly positive), recall/sensitivity (number of true positives over the total number of positives), F1 score (combining precision and recall using the harmonic mean), and AUC (Choudhury et al., 2021; Lantz, 2013). Even though accuracy is effective for many classification problems (Lantz, 2013), it was futile in our setting due to the class imbalance (failing devices were a minority class) and false negatives (FNs; i.e., not predicting a device that would fail) being more costly. This cost could also vary depending on the type of goods in transit and the customer. Therefore, we adopted the AUC score, since the ROC curve plots the true positive rate (TPR) against the false positive rate (FPR) at different classification thresholds, providing an aggregate measure of performance across those thresholds (Bradley, 1997).

Finally, we compared the models among the different ML algorithms we implemented. The AUC scores of these models are tabulated in Table 2. An AUC score of 1 symbolizes a perfect model, while one of

0.5 represents a random guess. The reported AUC scores illustrate the superiority of the predictive performance of our models against a random guess.

| Model | AUC score | Model | AUC score | Model | AUC score |
|---|---|---|---|---|---|
| Logistic regression | 0.809 | Gradient boosting | 0.865 | Neural network | 0.850 |
| Decision tree | 0.615 | Random forest | 0.897 | | |

*Table 2. AUC scores of the models.*

The random forest model provided the best AUC score, with 0.897, while the neural network model also performed relatively well, with an AUC score of 0.850. However, neural networks have the capacity to achieve high predictive performance when they are trained with big data (Choudhury et al., 2021). Therefore, we picked the random forest and neural network model to continue into the beta version.

## 5.3 BIE cycle 3: Improvements to the alpha version

In consultation with domain experts, we identified two preliminary improvement actions to adapt our models better to the real organizational context: (1) addressing the class imbalance and (2) scaling up the alpha version. Class imbalance happens in a dataset when a target class has a very small number of instances relative to other classes. In the presence of this problem, a trivial classifier might fail to detect a minority class due to its low incidence rate in the data (Chawla et al., 2002). However, the minority class was the class that was important in our problem. We tried three methods to balance the datasets, namely oversampling, under-sampling, and the synthetic minority over-sampling technique (SMOTE) (Chawla et al., 2002). However, neither model improved by following class balancing measures when compared with the AUC scores of the alpha version. Class balancing measures can improve performance, but they are not guaranteed to do so (Chawla et al., 2002). We therefore abandoned the class balancing, but we scaled up the chosen models (neural networks and random forest) by training them with five times more data points than in BIE cycle 2 to better reflect the actual operational scenario. This led to the development of the beta version.

## 5.4 BIE cycle 4: Development of the beta version

To develop the beta version, the alpha version was refined with the intention of extending the best-performing neural network model to better meet the existing functional scenarios (see Section 5.3) and the needs of the users of the system. The domain experts and the device monitoring team (users) suggested integrating the costs associated with the prediction errors (cost analysis) and customer criticality (classifier threshold sensitivity). Both FNs and false positives (FPs) carry costs, but the cost of an FN could range from \$10K to \$3.5M in the case of a discarded shipment (but not always), whereas an FP would involve only the labor cost to quality check (~\$50). Table 3 shows the assumed error costs for the beta version.

| Error | Cost driver | Cost (\$) |
|---|---|---|
| FN | Average costs incurred from a failed device | 1000 |
| FP | Labor cost to quality check | 50 |

*Table 3. Baseline cost assumptions for errors (FNs and FPs).*

The number of FNs and FPs was based on the output of the classification model. The output of a classification problem is the predicted probability values between 0 and 1. This value is compared with the threshold value to map the predictions to one class or the other depending on whether the prediction is higher or lower than the threshold (Lantz, 2013). Therefore, this threshold value is instrumental in deciding the number of FNs and FPs. However, decision makers can alter the threshold value depending on the desired outcome for the problem at hand. For example, for a highly important customer with critical materials in the logistics chain, it would be sensible to have a lower threshold value to reduce the FNs, as a failed monitoring device would have severe consequences.

To accommodate the dynamism in this decision-making environment, we set up an interactive UI with a confusion matrix and classifier threshold sensitivity analysis, in which the decision maker could use their cost analysis as input and alter the threshold according to the customer, type of good, and situation to see what effect the decision had on the output (see Figure 3). The decision maker could alter the threshold bar to observe how it would affect the prediction, cost, and the different metrics. The output was a confusion matrix with the number of observations in each category, the total cost and associated cost for each category, along with an arsenal of model performance metrics.

The beta version of the artifact comprised the neural network classifier model as the back-end predictor, while the interactive front end provided an adjustable cost analysis for prediction errors and the classifier threshold sensitivity analysis to augment the corrective/maintenance decisions in SCMo.

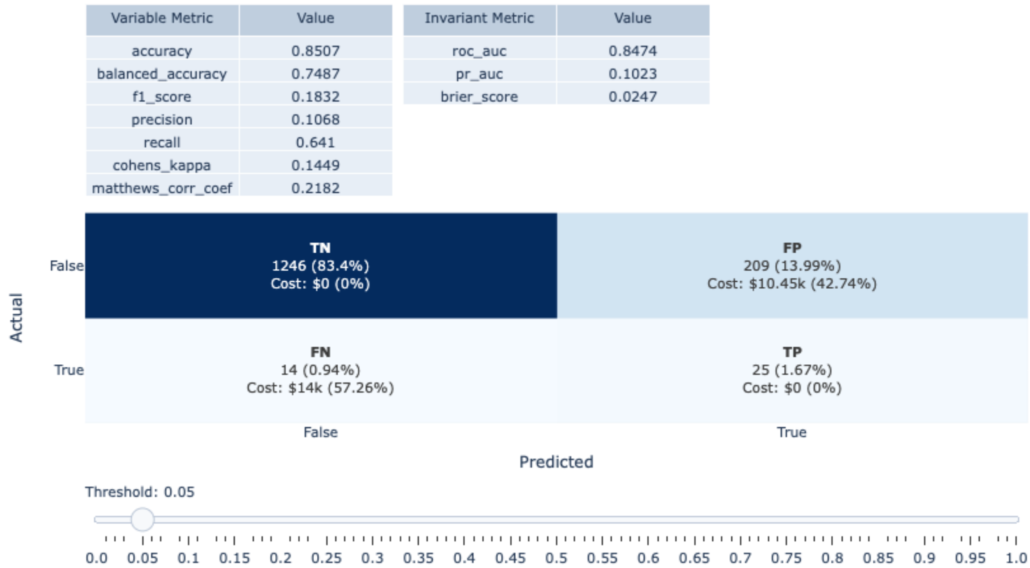


*Figure 3.* *Interactive UI with a confusion matrix, cost details, and model performance metrics.*

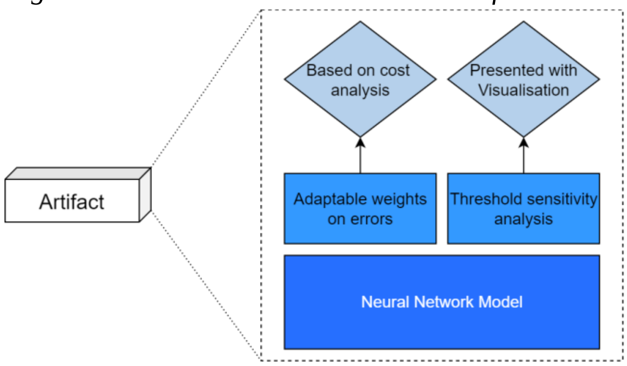


*Figure 4.* *Components of the beta version.*

## 5.5 BIE cycle 5: Evaluating the artifact

We comprehensively evaluated the artifact by conducting the DSR evaluation approach suggested by Venable et al. (2012) and Nunamaker et al. (2015). To do so, we evaluated the following: (1) the effectiveness of the artifact (POC), (2) its efficiency in achieving its stated purpose (POV), and (3) its usefulness in the actual operational context (POU).

The significantly high AUC scores listed in Table 2, especially for the random forest and neural network models, demonstrated the superiority of predictive performance compared to a random guess. This high predictive performance signified the effectiveness of the artifact in assisting decision-makers to identify failing devices. Hence, it presented POC.

When evaluating its efficiency in achieving its stated purpose, we examined the extent to which the artifact satisfied the needs and objectives of the business (see Section 5.1). The efficiency was measured by calculating the potential cost savings of implementing the artifact. Accurate prediction of the failing devices that would trigger corrective/maintenance actions yielded cost savings. In Section 5.4, we present the interactive UI on which the decision maker could alter the classifier threshold (T) and input the relevant costs of a particular scenario. Depending on the cost for each company (e.g., the value of the goods) and the classifier threshold choice, the cost savings change. Adopting the same baseline error cost assumptions (Table 3), we plotted the cost against the classifier threshold (see Figure 5). Then, by setting T = 0.05 (dark dotted line), the total cost would be $24,500. Compared with the total inactivity (i.e., no corrective action on any device; T = 1) cost of $39,000, this saved $14,500 for the 10 days (our window size; see Figure 2), which would accumulate to ~$530,000 on an annual basis, manifesting a ~37.2% cost saving. These significant cost savings provided POV of the artifact.

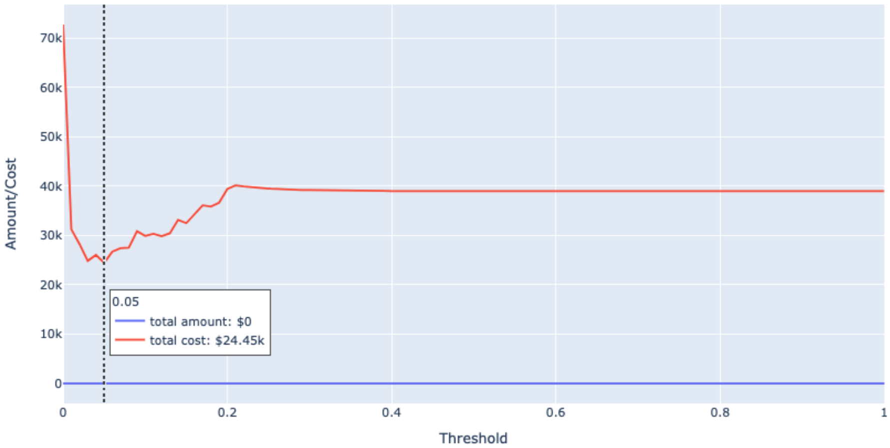


*Figure 5. Cost against classifier threshold plot.*

We investigated the usefulness of the artifact in the actual operational context by conducting field experiments and interviews with users and executives. First, we designed an experiment to measure the overfitting of our prediction model. We compared the AUC scores between the model trained on prior data and that on new unseen data and found similar AUC scores for our model, confirming low overfitting. Second, we conducted interviews with the company's device monitoring manager and VP of data and automation to explore whether the company would use and deploy the artifact in their operations. The VP of data and automation stated that "the artifact is impressive, and we will try to deploy it to make use of it in our operations." This was due to the high performance and flexibility it offered when applied to different customers and dynamics. Furthermore, he said that the model would be good for handling the company's expected scaling. Both interviewees agreed that the project had helped confirm the powerful data that the company had and said that the POC, POV, and feasibility of the project would inspire the company to accelerate its AI journey. Together, the experiment and interviews presented POU.

# 6 Reflection, Learning, and Formalization

Through multiple BIE cycles, we successfully developed an AI-based predictive maintenance system that predicts device failures and augments corrective and maintenance decisions in SCMo. The artifact was continuously reviewed and improved by maintaining a strong knowledge exchange between the domain experts and researchers who played mutually influential roles over five BIE cycles. Additionally, the reflection and learning were guided by our robust, concurrent, and authentic artifact evaluation methodologies (see Section 5.5), which enabled us to identify the design principles organically emerging from our design process while simultaneously demonstrating the user utility in addressing the practitioner problem. The designed artifact is not merely useful to our case company, but makes a contribution to the development of AI-based predictive maintenance systems that predict device failures in broader SCMo use cases (our problem-solution class). To synthesize this contribution, we reflected on our learnings during the ADR project and abstracted the technical solution as well as the related intended and unintended consequences of designing and deploying the artifact in real operational settings to a broader class. Table 4, which is placed at the end of this paper, presents an overview of the design principles (DPs), their rationalization, and their influence on the artifact.

# 7 Discussion

The purpose of this paper was to synthesize the design principles for an AI-based predictive maintenance system that predicts device failures and augments corrective/maintenance decisions in SCMo. Our study contributes to both the academic literature and managerial practice. First, we contribute to the IS artifact design literature with a set of six design principles for AI-based predictive maintenance system development for SCMo applications (Table 4). Despite the proliferation of IT and AI technologies in the SCMa and SCMo domains, there is a dearth of knowledge on how to design AI-based predictive maintenance systems in such contexts. We address this research gap by providing both a technical solution that is effective, efficient, and useful in practice and the six design principles that we identified through contemporaneous interaction between (1) design, development, and use and (2) the organizational and technological context (Orlikowski and Iacono, 2001). In this process of guided emergence, we reflected on the challenges we faced and, in our workarounds, derived generalizable findings. The design principles thus developed address key concerns of human-AI organizing, such as reconfigurability (Miah et al., 2019), tacit knowledge integration, accountability, algorithm appreciation (Feuerriegel et al., 2022), explainability (Bauer et al., 2023), cost efficiency, and system evaluation (Venable et al., 2012) within our problem class. Our evidence-based design principles exhibit how these crucial conceptualizations of human-AI symbiosis can be integrated into predictive maintenance systems in practice positioning our case as an instance. Thereby, our work finds practice-based empirical evidence for emerging conceptualizations in contemporary human-AI literature. Furthermore, our design principles were distilled from rich data sources, grounded in existing theories, and inspired by a class of problems in practice. Thus, we generalize our findings to the broader domain of AI-based predictive maintenance systems that predict device failures in SCMo.

Second, our work has several managerial implications. Most importantly, practitioners can develop customized instantiations for their own problem contexts in SCMo using our design principles to guide their design and decision making in the process (Möller et al., 2020). Our ADR process allows practitioners to effectively combine tacit domain knowledge and data science expertise when developing an AI artifact in SCMo. Furthermore, our evaluation metrics for POC, POV, and POU offer practitioners direction on what metrics they can use to prove their artifact's effectiveness, efficiency, and usefulness in countering the business problem. A positive result of such rigorous evaluation processes would give AI change agents a strong foundation to drive their cultural change and knowledge management journey toward embracing AI augmentation in organizations.

# 8 Conclusion and Future Work

Our study is not without limitations, the most prominent of which is that it is a single-company study. Such studies are influenced by organizational biases, as all the practitioners and their design, development, and evaluation of the artifact are guided by the same organizational context (culture, technologies, routines, and procedures). Moreover, we cannot guarantee that the proposed design principles are comprehensive and conclusive; as they were developed for dynamic AI technologies in a fast-growing scale-up, the proposed list of design principles may need to be changed.

Our findings and limitations leave us with several promising opportunities for future research. First, deriving design principles by developing similar artifacts for SCMo in other organizations could supplement, complement, or challenge the design principles already proposed. This would overcome the limitation of organizational bias. Second, even though our artifact exhibits superior performance, exploring different technical framings of the problem (e.g., ML-based regression models) might allow the artifact to achieve higher performance. Besides, advanced AI algorithms (e.g., deep learning and reinforcement learning) could yield higher performance and interesting implementation challenges that would be different from what we encountered. Third, not all decisions are the same, and there can be different topologies of SCMo decisions (e.g., descriptive, diagnostic, predictive, and prescriptive). Since we focused on the prediction (of device failure) typology, extending our investigation to other decision typologies could open up interesting puzzles for IS and DSR scholars.

| DP | Rationalization | Influence on artifact |
|---|---|---|
| DP1. Pick the right performance metric to optimize | Accuracy (the number of correct classifications) is a logical metric for most classification problems (Lantz, 2013). However, when data are imbalanced, a model can simply predict all data points to be a part of the majority class, thus showing higher accuracy. The ROC curve plots the TPR against the FPR at different classification thresholds, while the AUC provides an aggregate measure of performance across those different thresholds (Bradley, 1997). It prioritizes the improvement of the TPR and FPR (independently) and provides us with a performance indicator indifferent to the classification threshold (Bradley, 1997). | We optimized and compared models based on the AUC score primarily due to the class imbalance. |
| DP2. Treat errors differentially | Classification errors can be either FNs or FPs (Lantz, 2013). In our case, FNs have detrimental consequences for SCMo operations and the business in general. This demonstrates that the FPs and FNs cannot be treated identically. Depending on the problem context and problem at hand, the trade-off between the two errors needs to be managed. This can be in the form of performing the test multiple times and prioritizing minimizing one error over the other or incorporating the cost differences between FN and FP, as we did in our project (Choudhury et al., 2021). | We prioritized FNs due to their severity while maintaining the upper limit on FPs (within the feasible region). |
| DP3. Incorporate customizable parameters | Associated with DP2, the costs associated with prediction errors can be attributed to diverse sources, such as shipment, client, and external environment characteristics, which are better evaluated by the domain experts. There should be customizable parameters to incorporate the domain expertise of the decision makers into the AI artifact to achieve meaningful and robust decision augmentation. This is also closely connected to DP4. | We allowed decision makers to customize the classifier threshold and error costs. |
| DP4. Ensure human-in-the-loop | There are three key reasons for human-in-the-loop. First, due to its probabilistic and stochastic nature, AI is unpredictable and might yield unintended results (Shrestha et al., 2021). Therefore, it is important to obtain human judgment and evaluation of the problem to identify the instances in which systems might fail, assess the associated risks, and develop contingency plans. Second, it is essential that AI systems are integrated with human domain expertise (Ju, 2007; Murphy et al., 2017). The human provides domain expertise and can easily understand intangible information that the AI will not capture. Auditing an AI system can help the AI to make better decisions and prevent it from taking wrong and extreme measures (Grønsund and Aanestad, 2020). Third, over-reliance on AI could make decision makers lose their domain knowledge and accountability (Keding and Meissner, 2021). Therefore, it is wise to keep humans involved so that an organization does not completely lose its monitoring/risk-sensing skills (Alexander et al., 1997). | We integrated the domain expertise at every crucial point (interactive UIs, customizable parameters, auditing). |
| DP5. Employ interactive user interfaces | To implement DP3 and DP4, the results of AI systems should be interpretable and explainable for humans to understand the mechanisms in play. However, most AI/ML algorithms are not easily explainable (Choudhury et al., 2021), making it difficult for decision makers to comprehend how the parameters they control affect the final outcomes. One potential way to mitigate this problem is through interactive UIs coupled with visualizations (Bauer et al., 2023). Such mechanisms are useful in overcoming algorithmic aversion (Dietvorst et al., 2018), leading to algorithmic appreciation (Keding and Meissner, 2021) and trust (Burkart and Huber, 2021). While introducing interpretability, the designer should also keep in mind the potential drop in accuracy and carefully manage the interpretability–accuracy trade-off (Baryannis et al., 2019). | We developed an interactive UI to visualize important outputs of the system. |
| DP6. Scrutinize performance, value, and use | AI efforts are costly endeavors, demanding capital, a specialized workforce, and time, which are constrained organizational resources (Davenport and Zhang, 2021). Therefore, it is important to ensure that AI systems are effective, efficient, and usable to solve the chosen business problem. Choosing measurable use cases is the prerequisite of rigorous evaluations of the system (Tarafdar et al., 2019). A comprehensive evaluation scrutinizes the concept (POC), the value created (POV), and the usability in practice (POU) (Nunamaker et al., 2015; Venable et al., 2012). The measurement and demonstration of these parameters help garner stakeholder confidence and trust (Davenport and Zhang, 2021; Herath Pathirannehelage et al., 2022). | We obtained POC, POV, and POU of the system. |

*Table 4. Overview of the design principles, their rationalization, and influence on the artifact.*

## References

Abad, E., Palacio, F., Nuin, M., de Zárate, A. G., Juarros, A., Gómez, J. M., and Marco, S. (2009). "RFID smart tag for traceability and cold chain monitoring of foods: Demonstration in an intercontinental fresh fish logistic chain," *Journal of Food Engineering* 93 (4). https://doi.org/10.1016/j.jfoodeng.2009.02.004

Abbasi, A., Sarker, S., and Chiang, R. H. (2016). "Big data research in information systems: Toward an inclusive research agenda," *Journal of the Association for Information Systems* 17 (2), 3.

Abbasi, A., Zhang, Z., Zimbra, D., Chen, H., and Nunamaker Jr, J. F. (2010). "Detecting fake websites: The contribution of statistical learning theory," *MIS Quarterly* 435-461.

Ågerfalk, P., Conboy, K., Crowston, K., Eriksson Lundström, J., Jarvenpaa, S., Mikalef, P., and Ram, S. (2021). "Artificial intelligence in information systems: State of the art and research roadmap," *Communications of the Association for Information Systems*.

Ågerfalk, P. J. and Karlsson, F. (2020). "Artefactual and empirical contributions in information systems research," *European Journal of Information Systems* 29 (2), 109-113.

Alexander, P. A., Murphy, P. K., Woods, B. S., Duhon, K. E., and Parker, D. (1997). "College instruction and concomitant changes in students' knowledge, interest, and strategy use: A study of domain learning," *Contemporary Educational Psychology* 22 (2), 125-146.

Alicke, K., Dilda, V., Görner, S., Mori, L., Rebuffel, P., Reiter, S., and Samek, R. (2021). *Succeeding in the AI supply-chain revolution*. McKinsey & Company.

Altendeitering, M. and Guggenberger, T. M. (2021). "Designing data quality tools: Findings from an action design research project at Boehringer Ingelheim," in *Twenty-Ninth European Conference on Information Systems (ECIS2021)*.

Bailey, D. E., Faraj, S., Hinds, P. J., Leonardi, P. M., and Von Krogh, G. (2022). "We are all theorists of technology now: A relational perspective on emerging technology and organizing," *Organization Science* 33 (1), 1-18.

Baird, A. and Maruping, L. M. (2021). "The next generation of research on IS use: A theoretical framework of delegation to and from agentic IS artifacts," *MIS Quarterly* 45 (1).

Balakrishnan, T., Chui, M., Hall, B., and Henke, N. (2020). *The state of AI in 2020*. McKinsey & Company. URL: https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/global-survey-the-state-of-ai-in-2020 (visited on November 10, 2022).

Baryannis, G., Dani, S., and Antoniou, G. (2019). "Predicting supply chain risks using machine learning: The trade-off between performance and interpretability," *Future Generation Computer Systems* 101, 993-1004.

Baskerville, R. and Wood-Harper, A. T. (1998). "Diversity in information systems action research methods," *European Journal of Information Systems* 7 (2), 90-107.

Bauer, K., von Zahn, M., & Hinz, O. (2023). Expl (AI) ned: The impact of explainable artificial intelligence on users' information processing. *Information Systems Research*.

Berente, N., Gu, B., Recker, J., and Santhanam, R. (2021). "Managing artificial intelligence," *MIS Quarterly* 45 (3), 1433-1450.

Berkson, J. (1944). "Application of the logistic function to bio-assay," *Journal of the American Statistical Association* 39 (227), 357-365.

Bhavana, K., Nekkanti, V., and Jayapandian, N. (2019). "Internet of things enabled device fault prediction system using machine learning," in: *International Conference on Inventive Computation Technologies*. Cham, Switzerland: Springer.

Bishara, R. H. (2006). "Cold chain management: An essential component of the global pharmaceutical supply chain," *American Pharmaceutical Review* 9 (1).

Bradley, A. P. (1997). "The use of the area under the ROC curve in the evaluation of machine learning algorithms," *Pattern Recognition* 30 (7). https://doi.org/10.1016/S0031-3203(96)00142-2

Breiman, L. (2001). "Random forests," *Machine learning* 45 (1), 5-32.

Burkart, N., and Huber, M. F. (2021). “A survey on the explainability of supervised machine learning,” *Journal of Artificial Intelligence Research* 70, 245-317.

Catal, C. and Diri, B. (2009). “A systematic review of software fault prediction studies,” *Expert Systems with Applications* 36 (4), 7346-7354. https://doi.org/10.1016/j.eswa.2008.10.027

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., and Wirth, R. (2000). “CRISP-DM 1.0: Step-by-step data mining guide,” *SPSS Inc.* 9, 13.

Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. (2002). “SMOTE: Synthetic minority over-sampling technique,” *Journal of Artificial Intelligence Research* 16, 321-357.

Chen, H., Hailey, D., Wang, N., and Yu, P. (2014). “A review of data quality assessment methods for public health information systems,” *International Journal of Environmental Research and Public Health* 11 (5). https://doi.org/10.3390/ijerph110505170

Chen, K. L., Lassen, A., Li, C., and Møller, C. (2022). “Exploring the value of IoT data as an enabler of the transformation towards servitization: An action design research approach,” *European Journal of Information Systems*. https://doi.org/10.1080/0960085X.2022.2046515

Choudhury, P., Allen, R. T., and Endres, M. G. (2021). “Machine learning for pattern discovery in management research,” *Strategic Management Journal* 42 (1). https://doi.org/10.1002/smj.3215

Chui M., Hall, B., Singla, A., and Sukharevsky, A. (2021). *The state of AI in 2021.* McKinsey & Co. URL: https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/global-survey-the-state-of-ai-in-2021 (visited on January 25, 2022).

Davenport, T. H., and Zhang, R. (2021). “Achieving return on AI projects,” *MIT Sloan Management Review*. URL: https://sloanreview.mit.edu/article/achieving-return-on-ai-projects.

Dietvorst, B. J., Simmons, J. P., and Massey, C. (2018). “Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them,” *Management Science* 64 (3), 1155-1170.

Feuerriegel, S., Shrestha, Y. R., von Krogh, G., & Zhang, C. (2022). “Bringing artificial intelligence to business management,” *Nature Machine Intelligence* 4 (7), 611-613.

Fink, O. (2020). “Data-driven intelligent predictive maintenance of industrial assets,” in: *Women in industrial and systems engineering*, Cham, Switzerland: Springer.

Fosso Wamba, S., Queiroz, M. M., Guthrie, C., and Braganza, A. (2021). “Industry experiences of artificial intelligence (AI): Benefits and challenges in operations and supply chain management,” *Production Planning & Control*, 1-13.

Friedman, J. H. (2001). “Greedy function approximation: A gradient boosting machine,” *Annals of Statistics* 1189-1232.

Fuegener, A., Grahl, J., Gupta, A., and Ketter, W. (2022). “Cognitive challenges in human-artificial intelligence collaboration: Investigating the path toward productive delegation,” *Information Systems Research* 33 (2), 678-696.

Golovianko, M., Gryshko, S., Terziyan, V., and Tuunanen, T. (2022). “Responsible cognitive digital clones as decision-makers: A design science research study,” *European Journal of Information Systems* 1-23.

Grønsund, T. and Aanestad, M. (2020). “Augmenting the algorithm: Emerging human-in-the-loop work configurations,” *The Journal of Strategic Information Systems* 29 (2), 101614.

Han, J. W., Zuo, M., Zhu, W. Y., Zuo, J. H., Lü, E. L., and Yang, X. T. (2021). “A comprehensive review of cold chain logistics for fresh agricultural products: Current status, challenges, and future trends,” *Trends in Food Science & Technology* 109, 536-551.

Herath Pathirannehelage, S., Shrestha, Y. R., and Von Krogh, G. (2022). “Design principles for AI-based IS artefacts in organisational decision-making: An action design research study,” in: Yash, R. and Von Krogh, G. (eds.) *Design principles for AI-based IS artefacts in organisational decision-making: An action design research study.*

Hevner, A., March, S., Park, J., and Ram, S. (2004). “Design science in information systems research,” *MIS Quarterly* 28 (1), 75-105.

Jia, X., Feng, Q., Fan, T., and Lei, Q. (2012). “RFID technology and its applications in Internet of Things (IoT),” in: *Proceedings of the 2nd International Conference on Consumer Electronics, Communications and Networks, CECNet 2012*. https://doi.org/10.1109/CECNet.2012.6201508

Ju, B. (2007). “Does domain knowledge matter: Mapping users' expertise to their information interactions,” *Journal of the American Society for Information Science and Technology* 58 (13), 2007-2020.

Kanawaday, A. and Sane, A. (2018). “Machine learning for predictive maintenance of industrial machines using IoT sensor data,” in: *Proceedings of the IEEE International Conference on Software Engineering and Service Sciences, ICSESS*. https://doi.org/10.1109/ICSESS.2017.8342870

Keding, C. and Meissner, P. (2021). “Managerial overreliance on AI-augmented decision-making processes: How the use of AI-based advisory systems shapes choice behavior in R&D investment decisions,” *Technological Forecasting and Social Change*, 171, 120970.

Koot, M., Mes, M. R., and Iacob, M. E. (2021). “A systematic literature review of supply chain decision making supported by the Internet of Things and Big Data Analytics,” *Computers & Industrial Engineering* 154, 107076.

Lantz, B. (2013). *Machine learning with R: Learn how to use R to apply powerful machine learning methods and gain an insight into real-world applications*. in: Livery Place. Packt Publishing Ltd.

Leukel, J., González, J., and Riekert, M. (2021). “Adoption of machine learning technology for failure prediction in industrial maintenance: A systematic review,” *Journal of Manufacturing Systems* 61. https://doi.org/10.1016/j.jmsy.2021.08.012

Lyytinen, K., Nickerson, J. V., and King, J. L. (2020). “Metahuman systems = humans + machines that learn,” *Journal of Information Technology*, *36*(4), 427-445.

Mandviwalla, M. and Flanagan, R. (2021). “Small business digital transformation in the context of the pandemic,” *European Journal of Information Systems* 30 (4). https://doi.org/10.1080/0960085X.2021.1891004

Mathiassen, L. (2002). “Collaborative practice research,” *Information Technology & People* 15 (4). https://doi.org/10.1108/09593840210453115

Miah, S. J., Gammack, J. G., and McKay, J. (2019). “A metadesign theory for tailorable decision support,” *Journal of the Association for Information Systems* 20 (5), 4.

Möller, F., Guggenberger, T. M., and Otto, B. (2020). “Towards a method for design principle development in information systems,” in: *International Conference on Design Science Research in Information Systems and Technology*, Cham, Switzerland: Springer.

Mullarkey, M. T. and Hevner, A. R. (2019). “An elaborated action design research process model,” *European Journal of Information Systems* 28 (1), 6-20.

Murphy, P. K., Firetto, C. M., and Li, M. (2017). “Knowledge and the model of domain learning,” in: *The Model of Domain Learning*. Routledge.

Murray, A., Rhymer, J., and Sirmon, D. G. (2020). “Humans and technology: Forms of conjoined agency in organizations,” *Academy of Management Review* 46 (3), 552-571.

Nunamaker Jr, J. F., Briggs, R. O., Derrick, D. C., and Schwabe, G. (2015). “The last research mile: Achieving both rigor and relevance in information systems research,” *Journal of Management Information Systems* 32 (3), 10-47.

Orlikowski, W. J. and Iacono, C. S. (2001). “Research commentary: Desperately seeking the ‘IT’ in IT research—A call to theorizing the IT artifact,” *Information Systems Research* 12 (2), 121-134.

Padmanabhan, B., Sahoo, N., and Burton-Jones, A. (2022). “Machine learning in information systems research,” *Management Information Systems Quarterly* 46 (1), iii-xix.

Puranam, P. (2021). “Human-AI collaborative decision-making as an organization design problem,” *Journal of Organization Design* 10 (2), 75-80.

Quinlan, J. R. (1986). “Induction of decision trees,” *Machine Learning* 1 (1), 81-106.

Rai, A. (2017). “Editor’s comments: Diversity of design science research,” *MIS Quarterly* 41 (1), iii-xviii.

Raisch, S. and Krakowski, S. (2021). “Artificial intelligence and management: The automation-augmentation paradox,” *Academy of Management Review* 46 (1), 192-210.

Rosenblatt, F. (1958). “The perceptron: A probabilistic model for information storage and organization in the brain,” *Psychological Review* 65 (6), 386.

Sein, M. K., Henfridsson, O., Purao, S., Rossi, M., and Lindgren, R. (2011). “Action design research,” *MIS Quarterly* 37-56.

Seuring, S. and Müller, M. (2008). “From a literature review to a conceptual framework for sustainable supply chain management,” *Journal of Cleaner Production* 16 (15). https://doi.org/10.1016/j.jclepro.2008.04.020

Shih, C. W. and Wang, C. H. (2016). “Integrating wireless sensor networks with statistical quality control to develop a cold chain system in food industries,” *Computer Standards and Interfaces* 45. https://doi.org/10.1016/j.csi.2015.12.004

Shrestha, Y. R., Ben-Menahem, S. M., and Von Krogh, G. (2019). “Organizational decision-making structures in the age of artificial intelligence,” *California Management Review* 61 (4), 66-83.

Shrestha, Y. R., Krishna, V., and Von Krogh, G. (2021). “Augmenting organizational decision-making with deep learning algorithms: Principles, promises, and challenges,” *Journal of Business Research* 123, 588-603.

Sogi, S. K. and Mittal, S. K. (2020). “A time series forecasting model for equipment failure prediction using IoT sensor data,” *Advances in Mathematics: Scientific Journal* 9 (7). https://doi.org/10.37418/amsj.9.7.21

Stobierski, T. (2021). *The advantages of data-driven decision-making*. Harvard Business School Online. URL: https://online.hbs.edu/blog/post/data-driven-decision-making (visited on November 10, 2022).

Sun, X., Chakrabarty, K., Huang, R., Chen, Y., Zhao, B., Cao, H., ... and Jiang, L. (2019). “System-level hardware failure prediction using deep learning,” in: *2019 56th ACM/IEEE Design Automation Conference (DAC)*. IEEE.

Swanson, D., Goel, L., Francisco, K., and Stock, J. (2018). “An analysis of supply chain management research by topic,” *Supply Chain Management* 23 (2). https://doi.org/10.1108/SCM-05-2017-0166

Tarafdar, M., Beath, C. M., and Ross, J. W. (2019). “Using AI to enhance business operations,” *MIT Sloan Management Review* 60 (4), 37-44.

Teo, C. W., Low, K. L., Narang, V., and Thean, A. V. Y. (2019). “TCAD-enabled machine learning defect prediction to accelerate advanced semiconductor device failure analysis,” in: *2019 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD)*. IEEE.

Toorajipour, R., Sohrabpour, V., Nazarpour, A., Oghazi, P., and Fischl, M. (2021). “Artificial intelligence in supply chain management: A systematic literature review,” *Journal of Business Research* 122. https://doi.org/10.1016/j.jbusres.2020.09.009

Tsang, Y. P., Choy, K. L., Wu, C. H., Ho, G. T., Lam, C. H., and Koo, P. S. (2018). “An Internet of Things (IoT)-based risk monitoring system for managing cold supply chain risks,” *Industrial Management and Data Systems* 118 (7), 1432-1462.

Venable, J., Pries-Heje, J., and Baskerville, R. (2012). “A comprehensive framework for evaluation in design science research,” in: *International Conference on Design Science Research in Information Systems*, Berlin and Heidelberg, Germany: Springer.

Von Krogh, G. (2018). “Artificial intelligence in organizations: New opportunities for phenomenon-based theorizing,” *Academy of Management Discoveries* 4 (4), 404-409.

Wache, H. and Dinter, B. (2021). “Digital twins at the heart of smart service systems: An action design research study,” in: *Twenty-Ninth European Conference on Information Systems (ECIS2021)*.

Yang, S. H. (2014), “Internet of things,” in: *Wireless Sensor Networks*, London: Springer.