

Leuthe, Daniel; Meyer-Hollatz, Tim; Plank, Tobias; Senkmüller, Anja

**Article — Published Version**

## Towards Sustainability of AI – Identifying Design Patterns for Sustainable Machine Learning Development

Information Systems Frontiers

**Provided in Cooperation with:**

Springer Nature

*Suggested Citation:* Leuthe, Daniel; Meyer-Hollatz, Tim; Plank, Tobias; Senkmüller, Anja (2024) : Towards Sustainability of AI – Identifying Design Patterns for Sustainable Machine Learning Development, Information Systems Frontiers, ISSN 1572-9419, Springer US, New York, NY, Vol. 26, Iss. 6, pp. 2103-2145,  
<https://doi.org/10.1007/s10796-024-10526-6>

This Version is available at:

<https://hdl.handle.net/10419/319255>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



<http://creativecommons.org/licenses/by/4.0/>



# Towards Sustainability of AI – Identifying Design Patterns for Sustainable Machine Learning Development

Daniel Leuthe<sup>1,2,4</sup> · Tim Meyer-Hollatz<sup>1,3,4</sup> · Tobias Plank<sup>1,3,5</sup> · Anja Senkmüller<sup>3,5</sup>

Accepted: 30 July 2024 / Published online: 16 September 2024  
© The Author(s) 2024

## Abstract

As artificial intelligence (AI) and machine learning (ML) advance, concerns about their sustainability impact grow. The emerging field "Sustainability of AI" addresses this issue, with papers exploring distinct aspects of ML's sustainability. However, it lacks a comprehensive approach that considers all ML development phases, treats sustainability holistically, and incorporates practitioner feedback. In response, we developed the sustainable ML design pattern matrix (SML-DPM) consisting of 35 design patterns grounded in justificatory knowledge from research, refined with naturalistic insights from expert interviews and validated in three real-world case studies using a web-based instantiation. The design patterns are structured along a four-phased ML development process, the sustainability dimensions of environmental, social, and governance (ESG), and allocated to five ML stakeholder groups. It represents the first artifact to enhance each ML development phase along each ESG dimension. The SML-DPM fuels advancement by aggregating distinct research, laying the groundwork for future investigations, and providing a roadmap for sustainable ML development.

**Keywords** Artificial intelligence · Design patterns · ESG · Machine learning · Sustainability of AI

## 1 Introduction

Driven by the development of powerful hardware, the availability of extensive data sources, and the development of new algorithms, artificial intelligence (AI) and its prominent subset machine learning (ML) create value in various

domains such as education, finance, and manufacturing (Enholm et al., 2022; Kim et al., 2022). While there are ever-increasing efforts to apply AI and ML for sustainable usage, i.e., the early detection of wildfires (Wanner et al., 2020) or cancer (Schoormann et al., 2023), AI's negative impacts on resource consumption, societal injustice, or even human rights cannot be neglected anymore (Cowls et al., 2023; Dennehy et al., 2023; Koniakou, 2023). For instance, AI holds the unintended risk of reflecting the implicit social bias at the expense of equality, e.g., between genders or ethnic groups (Gupta et al., 2022; van Noorden & Perkel, 2023) or the amount of computing power needed to train current AI models, which has doubled every 3.4 months since 2012 (Amodei & Hernandez, 2018; Debus et al., 2023). The 'dark side' of AI has therefore become more apparent (Mikalef et al., 2022), leading to calls to work toward the sustainability of AI (SAI) (Schoormann et al., 2023; Schwartz et al., 2020; Tornede et al., 2022).

Analogous to the research streams green information systems (IS) and green information technology (IT) (Veit & Thatcher, 2023), SAI describes the sustainable design, development, and use of AI through its entire lifecycle (van Wynsberghe, 2021). Besides the work bundled under the term SAI, researchers analyzed adjacent topics such as

---

✉ Daniel Leuthe  
daniel.leuthe@fit.fraunhofer.de  
Tim Meyer-Hollatz  
tim.meyer-hollatz@fit.fraunhofer.de  
Tobias Plank  
tobias.plank@fim-rc.de  
Anja Senkmüller  
anja.senkmueeller@tum.de

<sup>1</sup> FIM Research Center for Information Management, Augsburg, Germany  
<sup>2</sup> University of Applied Sciences Augsburg, Augsburg, Germany  
<sup>3</sup> University of Bayreuth, Bayreuth, Germany  
<sup>4</sup> Branch Business & Information Systems Engineering of the Fraunhofer FIT, Augsburg, Germany  
<sup>5</sup> Technical University of Munich, Munich, Germany

‘responsible AI’, ‘ethical AI’, and ‘green ML/AI’ (Verdecchia et al., 2023). To consider these various topics, numerous perspectives, such as technical, social, and governance, must be considered (Kreuzberger et al., 2023; Merhi, 2023a). To enable this at an operational level, multiple stakeholders need to work together throughout the entire lifecycle to make AI and ML more sustainable (Papagiannidis et al., 2023; van Wynsberghe, 2021).

Given its interdisciplinary sociotechnical structure, the IS discipline has a vital role in accounting for these perspectives (Sarker et al., 2019). IS researchers have explored how to use information technologies and organizational methods to foster sustainability and social responsibility (Dennehy et al., 2023; Lee et al., 2012; Thomas et al., 2016). Thus, IS can contribute to integrating a sustainability perspective into AI and ML development. The previous work can be classified into three streams. First, a majority of papers have focused on solutions to reduce the energy consumption of AI development and ML models and therefore the environmental impacts (e.g., Patterson et al., 2022; Veit & Thatcher, 2023; Verdecchia et al., 2023). Second, recent publications have focused on social and ethical aspects of ML as well as increasing fairness during ML development to foster responsible ML (e.g., Dennehy et al., 2023; Ferrara, 2023; Mikalef et al., 2022). Third, an increasing number of papers are focusing on the challenges ML holds from a governance perspective (Koniakou, 2023; Papagiannidis et al., 2023; Verdecchia et al., 2023).

Overall, previous work is fragmented across several streams, leading to overlapping recommendations and difficulties, especially for practitioners, to comprehensively assess possible measures toward more sustainable ML. At the same time, there is an increasing demand as well as calls for research to shift from pure principles to comprehensive design approaches and implementable best practices for sustainable AI, for instance, to avoid involuntary exclusion or unnecessary resource consumption to counteract digital inequalities or negative environmental externalities (Dennehy et al., 2023; Pappas et al., 2023; Shneiderman, 2021; Vassilakopoulou & Hustad, 2023). Here, design patterns (DPs) have been proven to be valuable, as they capture best practices, guidelines, and recommendations and are a common tool to provide methodological support (Gamma, 1995; Goel et al., 2023). They have the advantage of being specific to solve a problem but also generic enough to address future similar problems, as they provide simple entry points and are easy to understand (Gregor et al., 2020). Thus, this paper seeks to answer the following research question (RQ):

*What are design patterns that ML development stakeholders can incorporate to increase the sustainability of the ML development process?*

In response to the RQ, we developed a comprehensive framework, namely the Sustainable Machine Learning Design Pattern Matrix (SML-DPM), that provides researchers and practitioners with recommendations to increase the sustainability of the ML development process. The SML-DPM provides 35 DPs structured along four phases of the ML development process and subdivides them into three sustainability dimensions. We follow the design science research (DSR) paradigm to develop the SML-DPM in close alignment with four literature-grounded key requirements (Hevner et al., 2004; Peffers et al., 2007). We derive the first set of DPs from 41 multivocal references. To evaluate and iterate on these DPs, we use the criteria developed by Sonnenberg and vom Brocke (2012b). Thus, we first assess their applicability and usefulness through focus groups and semi-structured interviews with subject matter experts. We then develop a web-based prototype to evaluate the intentions of users, leveraging our SML-DPM based on a case study in three real-world ML projects.

The research results make a theoretical contribution by conceptualizing SAI’s multidimensionality, which needs to be considered to increase sustainability across the whole ML development process and what SAI means in terms of implementable practices (Gill et al., 2022; Wu et al., 2022). The SML-DPM combines hitherto fragmented theoretical knowledge from separated research areas in an aggregated view and validates it with real-world ML project insights. This lays the foundation for further theorizing in the SAI field by embracing research approaches that encompass the multidimensionality of sustainability, including environmental, governance, social, technical, and human-centric perspectives. As a contribution to practice, the SML-DPM serves as a diagnostic tool for different ML development stakeholders to capture the status quo of sustainability and to develop a vision regarding the sustainability of the ML development process in their current and future ML projects. Further, the DPs and the associated web-based prototype offer an easily accessible guidance with a clear starting point for every ML stakeholder to transform the ML development processes towards more sustainability.

## 2 Theoretical Background

We structured the theoretical background into three distinct sections. First, we describe the characteristics of ML projects, including their process phases and stakeholders. Next, we discuss various sustainability frameworks that can be applied to ML development. Finally, we provide a comprehensive overview of related work.

**Table 1** Machine learning development process

		ML Development Frameworks						
		CRISP-DM (Wirth & Hipp, 2000)	KDD (Fayyad et al., 1996)	CRISP-ML (Studer et al., 2021)	AIDOP (Allen et al., 2017)	TDSP (Tabladillo, 2022)	Microsoft ML workflow (Amershi et al., 2019)	ML operation principles (Kreuzberger et al., 2023)
ML Development Phases	ML Demand Specification	Business understanding	Understanding of the application domain	Business understanding	Plan	Business understanding	Model requirements	MLOps product initiation, Requirements for feature pipeline
	Data Collection and Preparation	Data understanding, Data preparation	Data selection, preprocessing, and transformation	Data understanding, Data preparation	Data preparation, Build ground truth	Data acquisition and understanding, Feature engineering	Data collection, Data labeling, Data cleaning, Feature engineering	Feature engineering pipeline
	Modeling and Training	Modeling, Evaluation	Algorithm selection, Data mining	Modeling, Evaluation	Train model, Evaluate model	Model training, Model evaluation	Model training, Model evaluation	Experimentation
	Deployment and Monitoring	Deployment	Interpretation, Evaluation	Deployment, Monitoring and maintenance	Deploy, Monitor, Capture feedback	Deployment	Model deployment, Model monitoring	Automated ML workflow pipeline

## 2.1 Machine Learning Project and Process

AI is considered an umbrella term that includes different algorithmic approaches and methods (Russell & Norvig, 2016). The prevailing ones include ML (Ågerfalk, 2020). ML comprises capabilities to iteratively learn from training data and to improve their results, solving tasks automatically without explicitly being programmed (Collins et al., 2021). To realize ML's value, companies carry out ML projects. Generally, to accomplish a project's implementation, companies pass through the four main phases of planning, developing, deploying, and maintaining (Cooper & Zmud, 1990). While the overall discussion of projects, e.g., IT projects, is mature, the emergence of ML projects challenges established knowledge owing to the differences between ML and IT projects (Berente et al., 2021; Merhi, 2023b). To address this, the literature provides frameworks to guide the execution of ML projects. A comparison of the relevant different ML development frameworks appears in Table 1.

We reframed the four project phases (Cooper & Zmud, 1990) by including a more data-centric perspective, accounting for ML specifics such as iterative learning and by considering the ML development frameworks from Table 1. This both accounts for a comprehensive overview over the whole ML development process in four phases and enables an end-to-end consideration of the entire ML development process. First, during the planning phase, besides understanding the

organizational problems and opportunities, there needs to be a stronger focus on the identification of ML model requirements (Amershi et al., 2019; Kreuzberger et al., 2023). Second, there is a need for a data-centric phase prior to model development, as the data basis must be explicitly considered in ML projects besides the pure IT application development (Allen et al., 2017; Papagiannidis et al., 2023). Thus, the data challenges of data understanding, preparation, transformation, feature engineering, labeling, and cleaning must be considered (Tabladillo, 2022). Third, during the development phase, iterative training, experimentation, and evaluation loops must be conducted to benchmark different ML models (Fayyad et al., 1996; Tabladillo, 2022). This comprises the modeling, training, and evaluation of the ML models based on the previous phase's data (Amershi et al., 2019). Depending on the ML model evaluation results, adjustments such as hyperparameter optimization (HPO) can be made (Kreuzberger et al., 2023). Fourth, the deployment and monitoring phases should be considered together in ML projects, as changes such as data drift and concept drift must be continually monitored and could lead to redeployment (Kreuzberger et al., 2023). Consequently, this consists of deploying the ML model, transitioning the ML model into a software product, and monitoring its predictions and decisions in a real-world environment (Studer et al., 2021). This leads to the four overarching ML development phases (see, Table 1). These four phases are not purely

**Table 2** Definition of stakeholder groups within the ML development process

Stakeholder Group	Description	Kreuzberger et al. (2023)	Bhatt et al. (2020)	Yurrita et al. (2022)
Business Stakeholder (B)	They define the business problem, track the economic prospects, communicate within the organization, and interact with the end-users	X	X	
Domain Expert (D)	They provide the other stakeholders with in-depth know-how about the application domain. They assist by guiding the understanding of the data and the interactions with the system boundaries			X
ML Development (M)	This role includes data preparation, data transformation, feature engineering, model development, and model optimization	X	X	X
Software Development (S)	Their main responsibility is the transformation of the raw ML algorithm into a sustainable software product	X		X
Auditing and Testing (AT)	This role analyzes the used data and model throughout the process, focusing on the legal aspects and conformance with company regulations, and as well as the model's usability		X	X

sequential: iterations and feedback loops between the phases are both possible and necessary (Singla et al., 2018). The four phases provide a framework-agnostic analysis of the ML development processes by aggregating the ML development framework-specific phases to a higher abstraction level.

Further, different stakeholders are involved in the ML development process (Berente et al., 2021; Yurrita et al., 2022). Based on Kreuzberger et al. (2023), Bhatt et al. (2020), and Yurrita et al. (2022), we identified five stakeholder groups that intersect with the ML development process (Table 2).

The stakeholder groups we identified are based on Kreuzberger et al. (2023), yet these authors focused on ML operations and therefore defined highly specialized roles for the ML and software development phase. While these roles provide valuable granularity for ML operations, they prove overly detailed for the broader development process. To address this, we adopted the concept of “ML Development” proposed by Bhatt et al. (2020), which amalgamates various roles such as data scientist. Further, Bhatt et al. (2020) merged software engineering and architecture roles into the unified category “Software Development”. Given our aim to encompass the entirety of the ML development process, we expanded the roles to encompass the categories of “Auditing and Testing” (Bhatt et al., 2020) and “Domain Expert” (Yurrita et al., 2022). Also, we excluded the “End User” role, as detailed by Bhatt et al. (2020) and Yurrita et al. (2022), since it is implicitly incorporated through the requirements outlined by business stakeholders and forms no active part of the ML development process.

## 2.2 Sustainability Frameworks

Sustainability considerations first gained significant international traction when in 1987 the United Nations issued the Brundtland Report, which defines that the needs of the

present should be met without compromising future generations’ ability to meet their own needs (Brundtland, 1987). Over time, the term *sustainability* was further refined in different ways, with various definitions, perspectives, and concepts of sustainability (Glavič & Lukman, 2007). Among leading studies, particularly the connection between environmental and socio-economic issues is considered a multidimensional challenge (Hopwood et al., 2005; Spangenberg, 2002). For instance, environmental concerns may limit social welfare and economic growth. In light of these dimensions, we define sustainability as a multidimensional concept.

To operationalize this concept, there is a need for a framework to define its dimensions. Given the plethora of sustainability frameworks (Missimer et al., 2017), we analyzed approaches from different fields to derive a suitable guideline for our research. First, from a single entity’s perspective, Elkington (2018) coined the term “triple bottom line” by defining sustainability as economic (e.g., profit), environmental (e.g., benefits for the planet), and social (e.g., benefits for people) development. However, the concept has been criticized for its hierarchal presentation of the economic perspective alongside the ecological and the social perspectives (Isil & Hernke, 2017), and for the interdependencies among the three dimensions (Sridhar & Jones, 2013). Second, as a clear agenda is needed to concretize these goals internationally, the UN developed 17 Sustainable Development Goals (SDGs) to structure global efforts for the peace and prosperity of humanity and the planet. Third, based on regulatory requirements – such as the EU Emission Trading System, the EU’s corporate sustainability reporting directive, or China’s Green Taxonomy – effective governance and policymaking became a *sine qua non* in the sustainability discourse (European Massuga et al., 2023; Parliament, 2022). The concept of environmental, social, and governance (ESG) evolved as a central paradigm to operationalize



sustainability while maintaining a multidimensional and holistic approach, especially in corporate contexts (T.-T. Li et al., 2021; Tsang et al., 2023). We chose to use ESG, as it fits our holistic perspective on sustainability and its frequent use in the corporate environment (Drempetic et al., 2020; Sætra, 2023). Within this work, we define the ESG dimensions according to Li et al. (2021) and Ketter et al. (2020). They describe the environmental dimension as the preservation of the natural environment, including tackling pollution and climate change. The social dimension focuses on humans, both at the individual and the community levels, and embraces for instance diversity and social justice. Finally, the governance dimension includes the rules and norms that guide corporate activities, such as data protection and information mechanisms.

### 2.3 Related Work

To date, the research at the intersection of AI and sustainability can be distinguished into two fields: ‘AI for sustainability’ and ‘sustainability of AI’ (van Wynsberghe, 2021). The academic literature and practitioner discourse have focused on how to use ML and related technologies to improve sustainable development (Natarajan et al., 2022; Schoormann et al., 2023). Since ML is ubiquitously applied in many domains, the sustainability of AI has gained attention, even though this research stream developed separately, focusing on the sustainable design and use of ML (Schoormann et al., 2023).

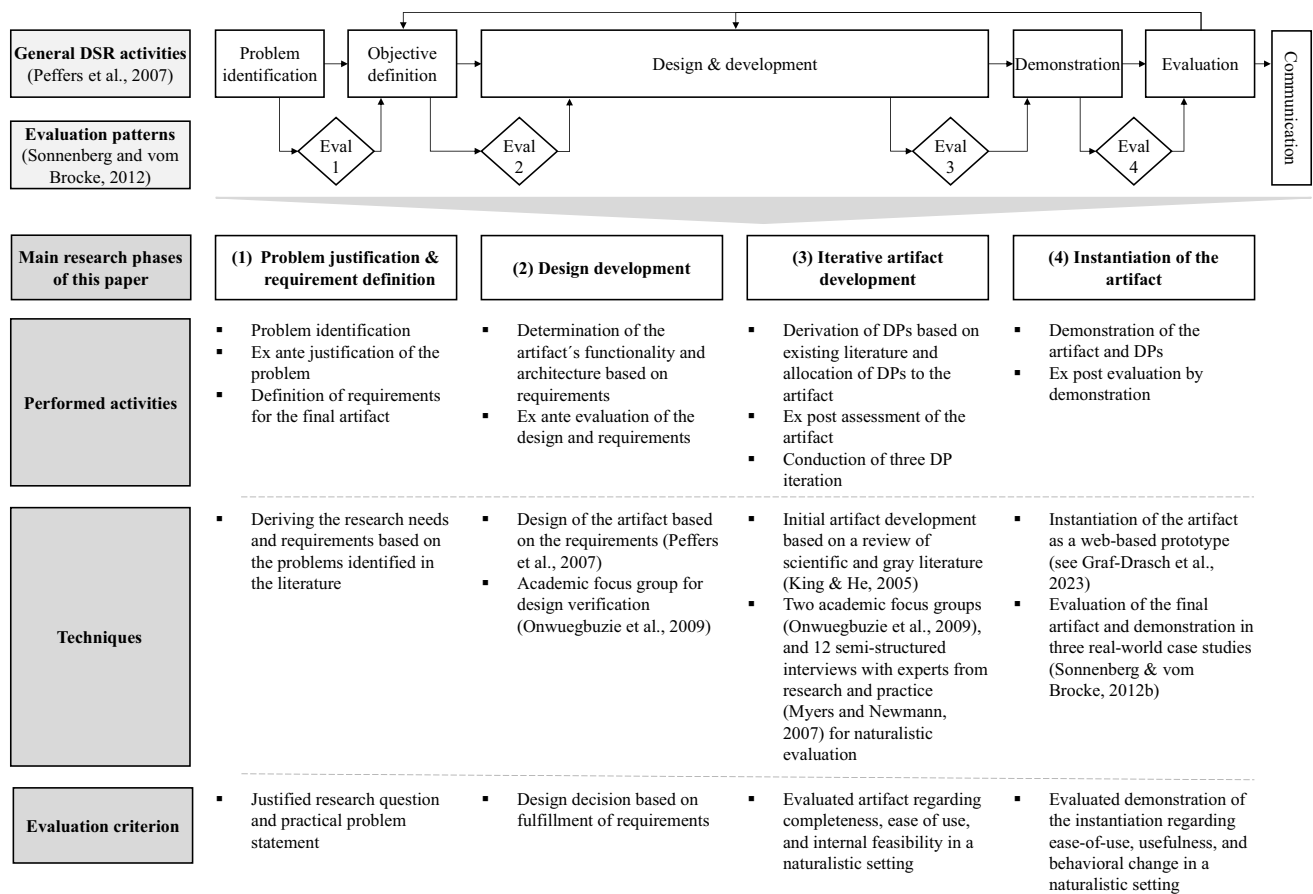
Various research papers have analyzed ML’s sustainability from multiple perspectives. Natarajan et al. (2022) analyzed the different recommendations to develop affordances for environmentally sustainable ML models. Similar papers by Schwartz et al. (2020) and Patterson et al. (2022) analyzed ML’s environmental sustainability under the term ‘Green AI/ML’. Henderson et al. (2022) and Cowls et al. (2023) provided strategies for mitigating carbon emissions and reducing ML models’ energy consumption. Schneider et al. (2019) derived principles for developing green data mining, which can be transferred to ML development. Verdecchia et al. (2023) concluded from a meta-analysis of research papers on ML development’s environmental sustainability that these studies focused mainly on the training phase. They called for new research that incorporates the gray literature and interviews with practitioners to validate previous findings (Verdecchia et al., 2023). Another separate research stream began to analyze the social implications of ML models regarding biases, fairness, and appropriate reliance (Mehrabi et al., 2022; Pagano et al., 2023; Singh et al., 2022). In this research stream, multiUsers\meyerhol\Zotero\storage\ple authors have emphasized the growing awareness to account for social risks during the use and development of AI (Dennehy et al., 2023).

In response, multiple frameworks have been designed to define rules for the socially responsible development of AI systems (c.f., Floridi et al., 2018; Montreal Declaration, 2017). For instance, Fahse et al. (2021) depicted multiple mitigation techniques for mitigating social biases, Akter et al. (2021) identified managerial capabilities to mitigate the three primary sources of algorithmic bias (i.e., data biases, method biases, and societal biases), and Friedler et al. (2019) compared different ML models on different data-sets regarding fairness. Nonetheless, there is a necessity to transition from the recognized risks and ethical principles to practical, actionable practices (Mäntymäki et al., 2022; Shneiderman, 2021). Finally, ML model governance focuses on promoting accountability mechanisms and governance structures, such as regulations and policies, to ensure that society benefits, while minimizing risks and harms (Taeiagh, 2021). For instance, Gill et al. (2022) provided an overview over the implications of ML governance on the regulatory, organizational, and process levels to drive the responsible development and uses of ML. Nonetheless, the ML governance research remains in its infancy owing to rapid ML development (Gill et al., 2022; Laato et al., 2022). Finally, Rohde et al. (2024) are among the first to synthesize the different perspectives of SAI by deriving sustainability criteria and associated sustainability indicators to evaluate the sustainability of AI systems. They structured the sustainability criteria using the triple bottom line (i.e., economic, environmental, and social) extended by an overarching governmental perspective along the ML development process. While their work offers valuable insights, they focus primarily on the sustainability assessment of existing AI systems and less on providing prescriptive action-oriented measures to make AI systems more sustainable.

Overall, the research in the sustainable development of AI and ML is fragmented across several streams, as multiple papers have focused on different areas of sustainability. While first research papers focus on the connection between sustainability dimensions and ML development, we lack a holistic approach that focuses on each ML development phase, provides clear recommendations for mitigating sustainability risks holistically, and integrates feedback from practitioners.

## 3 Research Methodology

Our methodological four-step research approach is derived from the DSR paradigm (Gregor & Hevner, 2013; Hevner et al., 2004), which aims to design artifacts to solve problems grounded in practice (vom Brocke, Winter, et al., 2020). Following Peffers et al. (2007), DSR is implemented as an iterative process that starts with a problem definition,



**Fig. 1** Our four-step research approach

defines goals, develops and refines a solution, demonstrates or applies it, evaluates whether the requirements have been met and the problem has been solved, and finally communicates the insights. While Hevner et al. (2004) and Peffers et al. (2007) focus on all activities involved in the DSR process, other works provide detailed insights into the evaluation steps (vom Brocke et al., 2020a). In this vein, Sonnenberg and vom Brocke (2012b) provide four patterns and corresponding criteria for evaluating DSR activities (i.e., Eval 1–4) and distinguish between ex-ante and ex-post evaluation. On the one hand, ex-ante evaluation patterns aim to justify the problem statement (e.g., the magnitude of the research need) and the validity of design decisions (e.g., the consistency of design requirements). On the other hand, ex-post evaluation challenges the artifact in artificial (e.g., challenging its internal feasibility) and naturalistic settings (e.g., its usefulness in a real-world demonstration) (Sonnenberg & vom Brocke, 2012a, 2012b; vom Brocke et al., 2020a). Within this work, we combine the described DSR process and the four subsequent evaluation patterns (as depicted in the upper part of Fig. 1) into four main research phases, each comprising activities and

a subsequent evaluation (as depicted in the lower part of Fig. 1). This four-step methodological approach simplifies both the research process and the structure of this paper by providing a clear and systematic framework. This reduction in complexity enhances comprehensibility and replicability, aligning with the standards set in earlier studies (e.g., Neff et al. (2014), Hausladen and Schosser (2020), Stahl et al. (2023)).

The first phase in our approach embraces the justification of the problem and the definition of requirements. In the second phase, a suitable artifact design based on the identified requirements was selected. The third phase comprised the iterative artifact development, especially regarding the derivation of the DPs and naturalistic evaluation. Finally, in the fourth phase, the artifact was naturalistically evaluated in a case study in three different projects. In the four phases, we used various research techniques. In every phase, we performed an evaluation step based on Sonnenberg and vom Brocke (2012b). Except for the problem justification (EVAL1) and demonstration (EVAL4), each evaluation step consisted of one or more interviews or focus groups (see Table 3).

**Table 3** Overview over the evaluation partners

#	Role	Stakeholder Group	Industry	Experience [years]	Size	Length [min]	Number of Participants	Evaluation Step	Design Pattern Iteration
F1	Scientific ML Community	B, M	Research & Consulting	1–4	Medium	50	22	EVAL2	-
F2	Scientific ML Community	B, M	Research & Consulting	2–5	Medium	55	54	EVAL3	Iter 1
F3	Scientific ML Community	M, S	Research & Consulting	2–5	Medium	85	8	EVAL3	Iter 1
E1	Scientific ML Consultant	M	Research & Consulting	> 3	Medium	48	1	EVAL3	Iter 2
E2	ML Consultant	M	Transportation & Mobility	> 4	Medium	55	1	EVAL3	Iter 2
E3	Data Analytics Specialist	D, A	Manufacturing	> 10	Large	60	1	EVAL3	Iter 2
E4	Data Product Owner	D, A	Technology	> 5	Large	44	1	EVAL3	Iter 2
E5	ML Transformation Project Expert	B	Telecommunication	> 2	Large	30	1	EVAL3	Iter 2
E6	Data Science Manager	B, M	Medicine	> 5	Start-Up	40	1	EVAL3	Iter 3
E7	Project Manager Data Science & ML	B, M	Consulting	> 4	Large	43	1	EVAL3	Iter 3
E8	Co-Founder	B, S	Technology	> 7	Start-Up	70	1	EVAL3	Iter 3
E9	Senior Expert AI	M	Consulting	> 10	Large	51	1	EVAL3	Iter 3
E10	Scientific ML Consultant	M	Research & Consulting	> 4	Medium	47	1	EVAL3	Iter 3
E11	ML Expert	M	Finance	> 5	Large	35	1	EVAL3	Iter 3
E12	Senior AI Strategist	B	Technology	> 6	Medium	49	1	EVAL3	Iter 3

**Problem justification & requirement definition** To understand and evaluate the importance and novelty (Sonnenberg and vom Brocke, 2012a) of challenges related to the sustainability of the ML development process, we justify the underlying problem based on the motivation and the literature provided in Section 2 (EVAL1). Thus, based on the problem justification (Sections 1 and 2), the stated RQ, and the gaps identified in the literature in the two research streams of ML and sustainability (Sections 2.1 and 2.2), we derive the three key requirements R1 to R3 for our SML-DPM. The requirements were further refined in the first focus group which led to the addition of one more requirement, R4 (see EVAL2).

**Design development** In the second phase, we developed the SML-DPM's structure and design. The first design of the artifact was theoretically grounded in the justificatory knowledge (Peffer et al., 2007) from the two research streams, sustainability and ML projects, as discussed in Sects. 2.1 and 2.2. Further, the realm of possible solutions (i.e., the solution space) was determined by the four key requirements. Based on the initial draft of the design, we conducted an artificial ex ante evaluation using a focus group. According to Belanger (2012), focus group research is suitable among

others if the participants evaluate a theoretical model or explore new topics. Since our artifact's design is comparable theoretically and the conjunction of sustainability and ML development is only just emerging, we conducted an academic focus group discussion to evaluate the first design of the artifact (EVAL2). We followed the high-level guidelines summarized by Onwuegbuzie et al. (2009). The focus group #F1 (first row in Table 3) consisted of 22 researchers working on industrial projects full-time due to an industry doctorate in data analytics and ML that lasted around 50 min. The participants are all involved in applied ML projects, such as predictive maintenance in industrial applications, punctuality prediction in public transportation, text classification of service reports, or the development of retrieval-augmented generation pipelines based on large language models for extracting market data. To facilitate the discussion, one of the co-authors acted as moderator and another as assistant moderator (Krueger, 1988). The results of the focus group session were recorded on a digital whiteboard. Krueger and Casey (2015) identified four guiding components that should lead the facilitation of a focus group. Accordingly, we first introduced the participants to the underlying research topic of SAI, our RQ, and key requirements (*the introductory*



stage). Second, we discussed their experience in research and practice regarding general and sustainability challenges in ML projects (*the transition stage*). Third, we presented and discussed our first artifact's design (*in-depth investigation*). The discussion was guided by the moderator to center around the design's completeness, understandability, and usability. Specifically, we asked the participants if the artifact's design was consistent and whether it met the key requirements. This was done to validate and, if needed, refine the compact structure in line with real-world ML projects to guarantee that the SML-DPM and its design add value to varying practical ML project settings as the participants of #F1 have extensive insights into different ML projects and industries (Sonnenberg & vom Brocke, 2012a). Finally, we summarized the key issues addressed by the participants (*closure*).

**Iterative artifact development** In the third phase, the artifact's content (i.e., the DPs) was iteratively developed in three iterations. We derived the first set of DPs by analyzing the literature (*Part I – Literature Analysis*) and synthesizing information on the sustainable ML development (*Part II – Initial Artifact Development*). Finally, we conducted an ex post evaluation of the first SML-DPM with two focus groups (*Part III – Iterative Development and Naturalistic Evaluation*).

*Part I – Literature Analysis:* To ensure the maximal bandwidth of input, we used a narrative literature review where we reviewed both the scientific and the nonscientific outlets. Given the novelty of the sustainability of ML, the numerous definitions, and the scarcity of studies at the intersection of all three ESG factors and ML (Tornede et al., 2022; Vinuesa et al., 2020; Wu et al., 2022), we chose this research approach so as to provide a holistic overview over recommendations (i.e., mitigation strategies, affordances) (King & He, 2005). Overall, this literature review type has the benefit of a particularly broad representation of reference literature (Green et al., 2006) owing to its ability to include knowledge from different perspectives on ML, data science, sustainability, and ESG. We first utilized a keyword-search in scientific databases, i.e., Science Direct, Springer Link, IEEEExplore, and AIS Electronic Library. We used a two-part search string for related research publications (Huang, et al. 2015). The first part of the string addressed ML and related mature research fields (“data analytics” OR “data mining” OR “machine learning” OR “ML” OR “artificial intelligence” OR “AI” OR “deep learning” OR “neural network”). We particularly included the terms “data analytics” and “data mining” to include this mature research field (e.g., Schneider et al., 2023) and the shared characteristics between data mining projects and ML development projects, which has been widely acknowledged by the adoption of CRISP-DM for ML research and projects (e.g., Singh

et al., 2022; Studer et al., 2021). Further, we included the ML subset “deep learning” and the technical wording “neural network” to include this stream of energy-intensive ML models (Desislavov et al., 2023). The second part sought references that provide insights into relevant sustainability topics (“sustainab\*” OR “energy” OR “environment\*” OR “social” OR “fair\*” OR “unbiased” OR “governmental” OR “trust\*” OR “responsib\*” OR “ethic\*”). The first part (up to “environment\*”) of the second search string related to the environmental dimension of sustainability, while the second part related to the governance and social dimensions. The sustainability dimensions of social and governance are often simultaneously addressed in the literature and, therefore, also in our search string (see Dennehy et al., 2023). Also, we checked for gray and industrial literature in Google Scholar, Google, arxiv.org, and OECD.AI, as well as technical reports of main ML companies such as Amazon, Microsoft, Google, and IBM. The literature identification step led us to 43 references (gray literature: 15, scientific literature: 28). Second, we applied a forward–backward search on the collected scientific papers to gain a deeper knowledge base. Simultaneously, we assessed the quality of the observed gray literature, relying on a process similar to Gramlich et al. (2023), using three criteria, selecting the gray literature based on the novelty of the contributions being referenced in the white literature, and its impacts through open access. Ultimately, these three steps (keyword search, forward–backward, quality assessment) ensured a broad literature base (41 references, gray literature: 8, scientific literature: 33) for deriving the DPs.

*Part II – Initial Artifact Development:* To extract these calls for action at the individual paper level, we screened the content for recommendations to increase the ML development process' sustainability. We allocated the findings to one of the three sustainability dimensions, one of the ML development phases, and the ML stakeholder groups. Ultimately, we combined the calls for action to derive the first set of DPs that collectively ensure sustainable ML development processes. We grounded our procedure in three meta-requirements which led the DP development process. Each DP had to be identified in two or more sources to ensure validity of the results. The first meta-requirement is accompanied by the general purpose of DPs to provide proven solutions for recurring problems. This guarantees that if the DP is identified in several sources, the DP contains established solutions for reappearing problems (Gamma, 1995; Goel et al., 2023). Further, if multiple recommendations focused on the same abstract solution while providing different instantiations, they were combined into one DP, i.e., abstraction (Gregor et al., 2020). Based on the second meta-requirement, we eliminated differences in wording among the various research streams. This level of abstraction allows to aggregate, abstract, and

generalize the codified design knowledge from the literature to be used for a class of SAI problems (e.g., only for a specific industry sector such as finance or telecommunication or a specific ML method such as computer vision or natural language processing Ayres & Sweller, 2014; Baxter et al., 2007; Schoormann et al., 2023)). Opposing to the prior meta-requirement, the final meta-requirement called for distinction between the DPs. Therefore, we included a new DP every time the authors identified a new cluster of calls which has not been included in the combination of phases and sustainability dimension. This resulting set of DPs is thus intended to guarantee that the underlying multidimensional SAI topic can be comprehended by the users. This allows a transfer of knowledge for users who have previously only dealt with one dimension and supports them in understanding the SAI problem context (Dickhaut et al., 2023; Rothe et al., 2020; vom Brocke, Winter, et al., 2020).

Those three meta-requirements left us with a first iteration of 39 DPs that can be levered to ameliorate the ML process along the ESG dimensions (see Appendix Fig. 6). To create transparency regarding the final selected studies and their assignment to the 39 DPs, we provide a detailed overview over the literature-DP allocation in Appendix Table 5.

**Part III – Iterative Development and Naturalistic Evaluation (EVAL3):** To evaluate and refine the DPs, we used two academic focus groups #F2 and #F3 (second and third row in Table 3) in the field of data analytics and ML. (for a description of the applied methodology, see the previously described phase 2).

After introducing the underlying research problem and scope in line with the RQ, we discussed their interaction with sustainability in ML development. Afterward, we presented and discussed each DP. We discussed how to ensure that each DP is easy to use, compact yet complete, and clearly assigned to one sustainable dimension, one ML development phase, and the relevant stakeholders. After going through all the DPs, we discussed whether they are mutually exclusive and collectively exhaustive. Finally, we had an open discussion on whether there were any other recommendations, best practices, or advice that were missing from the first set of DPs. Following the two focus groups, we conducted 12 interviews with subject matter experts #E1-12 (fourth to fifteenth row in Table 3) to validate the applicability and usefulness of the derived DPs. Following Myers and Newmann (2007), we used semi-structured interviews, which allow different perspectives to be absorbed while providing in-depth information (Miles & Huberman, 2009). We selected the interviewees from our professional network based on their sufficient expertise in the development of ML projects. We defined sufficient expertise as leading or co-developing more than three ML projects in an organizational setting. Further, we included a diverse set of backgrounds and current positions. Table 3 provides an overview over the interviewees

along their business area, experience with ML development, and current position. Overall, the total interview time was 572 min. The interviews were done over three weeks by two of the co-authors. Each interview had four parts. We began by introducing every interviewee to our motivation and the problem statement. We then deliberated over the prerequisites for the SML-DPM and its structure. Step 3 presented the DPs. Finally, the interviewees reviewed the application of the artifact to verify whether it met the practical requirements.

For both the focus groups and the interviews, we calculated a form of code frequency counts as the saturation metric. Saturation metrics are widely regarded as the primary criterion for evaluating results' sufficiency in qualitative research methodologies (Hennink & Kaiser, 2022). In the group of saturation metrics, code frequency counts are often used to evaluate the saturation of responses in interviews and focus groups (e.g., Ando et al., 2014; Young & Casey, 2018). Adopted to our methodological procedure, we rigorously analyzed the results from the interviews or the focus groups for emerging, changing, or diminishing DPs at each iteration. The number of each previously introduced modification, identified in every interview, was meticulously tracked, observing the point at which the emergence of new codes began to decline, indicating a trend toward saturation. We calculated the saturation – i.e., the number of changes divided by the number of DPs at the start of the iteration – per iteration, publishing the results in Appendix Table 6.

**Instantiation of the artifact** Finally, to complete the DSR process, we conducted a real-world demonstration using the SML-DPM (EVAL4) as the result of the previously described research phase three. The demonstration sought to outline how the artifact, i.e., the SML-DPM, can be used to solve the identified problem in line with our research question (Peppers et al., 2007; Sonnenberg & vom Brocke, 2012a). Even though the research community is divided on which specific goals to follow when evaluating design artifacts, the main objective is to analyze whether a design artifact holistically addresses an observed problem, which remains consistent (Prat et al., 2015). Thus, within the SML-DPM, we provided actionable DPs the ML development stakeholders can incorporate to increase the sustainability of the ML development process. To evaluate the intentions of users leveraging our SML-DPM to carry out ML projects sustainably, we conducted a case study in three ML development teams. Within the case study, we asked the participants (*real users*) to use our SML-DPM to identify areas of improvement for the development (*real task*) of their current ML projects (*real systems*) (see Sonnenberg & vom Brocke, 2012b). Case studies as a method to demonstrate the usefulness of design artifacts were proposed by Sonnenberg & vom Brocke (2012b). We selected two publicly funded ML projects and one industrial ML project as our case study's

subject of investigation. The three projects were chosen as a stratified sample (Robinson, 2014). All projects focus on productive ML development and are going through or have gone through all four ML development phases. Thus, the projects are comparable in this regard and correspond to the defined user group of the SML-DPM. Notably, these three projects had different application areas and customer groups and use different data and algorithm types. This allowed us to evaluate the SML-DPM in different contexts, so as to validate generalizability. Further details on the three case study projects (EVAL4) appear in Section 5.3.

In line with previous works by Graf-Drasch et al. (2023) and Schoormann et al. (2023), we instantiated the SML-DPM as a web-based prototype to assist the evaluation. The web-based prototype serves as a means of communicating the SML-DPM (i.e., transfer medium) and contains the same structure (i.e., the design) and content (i.e., the DPs) as the SML-DPM, enabling a straightforward assessment for the SML-DPM's target group (i.e., the ML development stakeholders). We followed Sommerville's (2011) prototype development process. In the workshops, we levered the web-based prototype to perform a fit-gap analysis. This process involved identifying the DPs that were already implemented in the participants' ongoing projects and that the participants were interested in incorporating into their workflows. For the subsequent interactive part, the participants were asked to take the perspectives from their current projects. After addressing preliminary inquiries, they were prompted to pinpoint the DPs already implemented in their projects. Adhering to the fit-gap analysis methodology, the participants were then encouraged to discern the DPs they wanted to integrate into their existing projects. The fit-gap analysis' results were then compared across the projects and were thoroughly discussed to synthesize common themes and divergences in the SML-DPM application. To amplify the comparison, we calculated the difference between the applied DPs and the planned DPs (after the workshop) as an additional metric for the SML-DPM's long-term impacts.

To ascertain the primary objectives of our investigation, we examined the ease-of-use, usefulness, and behavioral intention associated with the introduced intervention (Graf-Drasch et al., 2023; e.g., Sonnenberg & vom Brocke, 2012a; Zacharias et al., 2022). The selection of these evaluative criteria was theoretically underpinned by the Technology Acceptance Model (TAM) (Davis, 1989), which is often applied in IS research to assess the adoption and utilization of novel IT artifacts (e.g., Baroni et al., 2022; McCoy et al., 2007). In line with our research question, the first evaluation criteria – ease-of-use – focuses on evaluating the overall accessibility and simple usage of the SML-DPM, resulting in a positive attitude toward it (Davis, 1989), as the ML development stakeholders need to easily identify DPs that they can integrate into their own ML development process so that the effort is not unnecessarily increased. The second evaluation criteria – usefulness

– focuses on whether the SML-DPM creates practical added value for the ML development stakeholders and can, therefore, increase the sustainability of the ML development process. The third evaluation criteria – behavioral intention – aims to evaluate the attitude and intention of the ML development stakeholders to use the SML-DPM eventually, as more and more ML models continue to be integrated, increasing the need to take greater account of their sustainability risks (Denehy et al., 2023; van Wynaesberghe, 2021). As there is no clear agreement in the IS community on which questions to pose for each evaluation category (Prat et al., 2014), we followed Zacharias et al. (2022) and Wormeck et al. (2024) by combining questions from multiple research endeavors which include the same evaluation criterion. Hence, we derived three questions per criterion from similar DSR approaches to operationalize the TAM framework. In this evaluation, a structured survey along a five-point Likert scale was disseminated to the workshop participants. This scale required them to express their concurrence or dissent with the presented statements by choosing a value from 1 (*strongly disagree*) to 5 (*strongly agree*). An overview of the survey and the associated questions per evaluation criterion appears in Appendix Table 7. The survey data analysis represents the concluding step of the final evaluation phase in the DSR process.

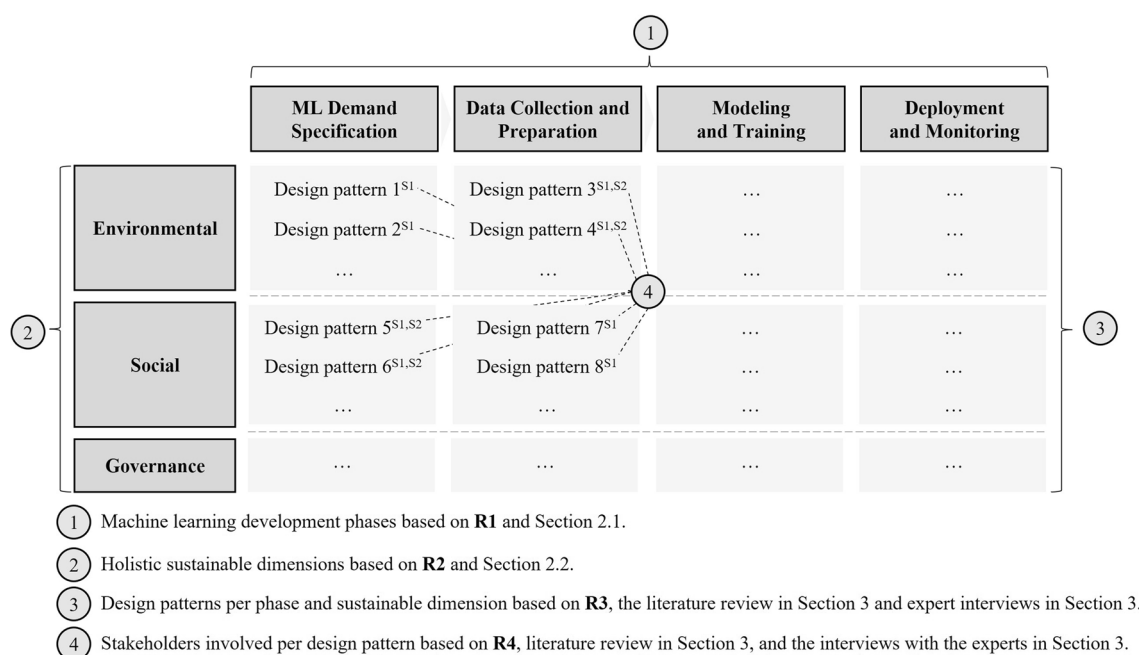
## 4 Results

We will now present our research results, following the methodological procedure outlined in Section 3.

### 4.1 Requirement Definition

Based on the identified problem (Section 1), grounded in the literature (Section 2), and justified in focus group 1 (Section 5.1), we derived four key requirements. The first two key requirements, R1 and R2, refer to the superordinate artifact, i.e., the SML-DPM, and its structure. The last two key requirements, R3 and R4, refer to the content of the artifact, i.e., the DPs.

**[R1] End-to-end consideration of the ML development process** ML projects often get stuck in an experimental pilot phase without transitioning to productive value-adding applications (Benbya et al., 2021; Merhi, 2023b). Thus, ML projects often fail to live up to their intended outcomes or are even terminated prior to completion (Westenberger et al., 2022). To successfully implement and fully grasp the real impact of ML models on sustainability, there is a need to consider all phases of the ML development process (Verdecchia et al., 2023; Wu et al., 2022). Hence, an end-to-end view of the ML development process is crucial for aligning the project to the problem's requirements, ensuring data



**Fig. 2** The design and structure of the SML-DPM

quality, selecting appropriate models, successfully deploying them, and maintaining their performance over time.

**[R2] Holistic view on sustainability** Sustainability goes beyond merely focusing on one dimension such as resource efficiency (Vinuesa et al., 2020). It involves an overarching view that considers factors such as systems thinking, a long-term perspective, social inclusivity, cross-disciplinary collaboration, and legal framework conditions, which go beyond purely environmental factors (van Wynsberghe, 2021). Therefore, sustainability must be defined holistically and must encompass more than the environmental dimension.

**[R3] Applicability of the design patterns for ML development stakeholder** The artifact must provide clear guidance on advancements toward more sustainable ML development. Hence, in line with the RQ, the derived results in the artifact need to be accessible and applicable to ML development stakeholders (Sonnenberg & vom Brocke, 2012b). Also, the DPs need to be usable for various distinct ML projects so as to ensure generality. Thus, the DPs should be ML model-agnostic. This ensures practical relevance and fidelity with the real-world problem at hand (Sonnenberg & vom Brocke, 2012a).

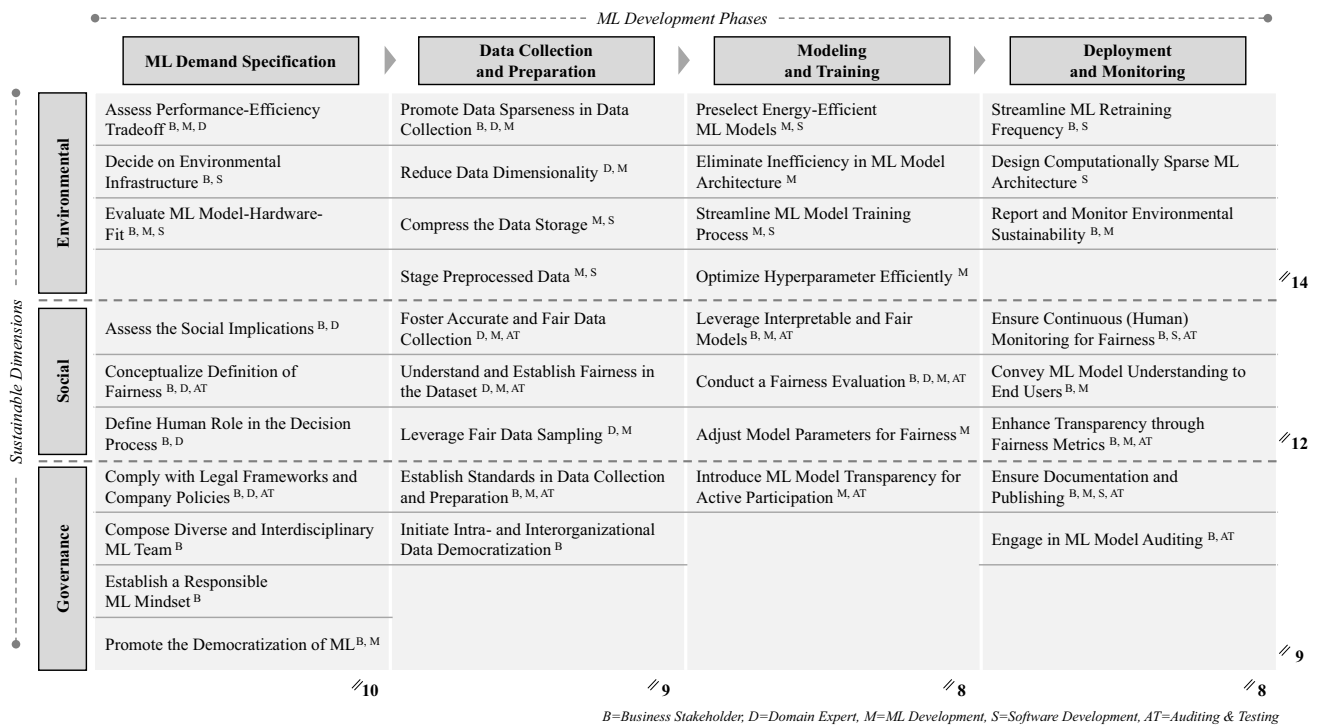
**[R4] Clear assignment of the ML development stakeholders involved** During the ML development process, stakeholders from different areas need to work together to achieve successful and sustainable development of ML (Kreuzberger

et al., 2023; Papagiannidis et al., 2023). As there are different options for DPs in each of these areas, the DPs must be clearly assigned to one or more involved stakeholders to limit the solution space of potential DPs to their roles and responsibilities. This allows the stakeholders to identify DPs they can engage with and ensures the ease of use of the final SML-DPM and the associated DPs (Sonnenberg & vom Brocke, 2012a).

## 4.2 SML-DPM's Design Description

In the following, we present the final design and structure of the SML-DPM (Fig. 2). In developing its design, we followed the four key requirements, R1 to R4. Based on the requirements R1 and R2, we chose a matrix layout that allows for a clear and comprehensive structure. First, the ML development process phases were included on the horizontal axis, enabling an end-to-end consideration of the whole ML development process (R1). For this purpose, we built on the four ML process phases: “ML Demand Specification”, “Data Collection and Preparation”, “Modeling and Training,” and “Deployment and Monitoring” (Section 2.1). Second, we incorporated the sustainability dimensions on the vertical axis (R2). To provide a holistic view on sustainability, including the ESG factors, we drew guidance from the ESG dimensions (Section 2.2). Third, the resulting SML-DPM's design allows clearly assignable DPs for ML development stakeholders, as each DP can be explicitly positioned in one ML process phase and one





**Fig. 3** The SML-DPM

sustainability dimension (R3). We opted for this explicit and non-overlapping one-to-one allocation, which makes it clear when which DP is relevant and clearly applicable in one phase and dimension. Fourth, we included a clear assignment of each DP to one or more ML development stakeholders by inserting them in superscript after each DP (R4).

Further, each DP must follow a standardized tripartite structure to ensure generalizability in line with the meta-requirements defined in section “Part II – Initial Artifact Development” in the research methodology to derive the DPs: 1) An action title must be defined according to a uniform structure regarding wording by using a verb and the object of investigation of the DP to briefly describe what is to be done to solve a general problem in a particular context (see the second meta-requirement). This compact and uniform structure of the action title across all DPs allows a quick and standardized entry point to the topic of SAI (Gamma, 1995; Gregor et al., 2020). 2) A DP must contain a theoretical description supported by the literature to provide justificatory knowledge (Jones & Gregor, 2007). On the one hand, this theoretical description clearly states what the DP is about and what can be done with it to increase the sustainability of the ML development process (Gregor et al., 2020). On the other hand, support by literature means that the functionality and meaningfulness of each DP has already been explicitly described in two or more papers, e.g., in the form of a quantitative

case study, a theoretical derivation, or a qualitative survey (see the first and third meta-requirement). This allows us to combine and harmonize the previous work, which is fragmented across several streams, on one level. 3) Each DP must be underpinned by a practical example and feedback from experts so as to ensure practical relevance and fidelity with the real-world problem at hand (Sonnenberg & vom Brocke, 2012a). This practical application-oriented focus enables the shift from pure theoretical principles to implementable best practices that practitioners have validated and can, therefore, be applied in the ML development process.

### 4.3 Design Patterns Description

The SML-DPM (Fig. 3) is divided into the ESG dimensions on the vertical axis and the ML development phases on the horizontal axis. The environmental dimension encompasses 14 DPs, the social dimension 12 DPs, and the governmental dimension 9 DPs. In the following three subsections, each DP is introduced based on the tripartite structure described in the previous chapter. Therefore, the action title is provided first, followed by the theoretical description. We further provide the practical justificatory knowledge from the interviews with domain experts for each DP. Due to the length, this is listed in Appendix Table 8.



### 4.3.1 Environmental Dimension

The environmental dimension consists of 14 DPs. The first phase *ML Demand Specification* is subdivided into three DPs:

The first DP “**Assess Performance-Efficiency Trade-off<sup>B, M, D</sup>**”, centers on the equilibrium between the performance (i.e., how well an ML model can accomplish a specific task required for an ML model to generate business value) and the diminished energy efficiency of more sophisticated ML models or hyperparameter configurations (Naser, 2023). Among others, Brownlee et al. (2021) have shown that a drop in accuracy of 1.1% can lead to energy savings of up to 77%. Estimating the benefits of additional performance upfront vs. the environmental cost when the ML model is trained and deployed can support the decision-making process on whether higher model accuracy justifies higher energy costs (A. Kumar, 2022; Schwartz et al., 2020).

Second, the DP “**Decide on Environmental Infrastructure<sup>B, S</sup>**” focuses on the infrastructure selection to reduce the carbon footprint per computing unit (Martínez-Fernández et al., 2023; Schneider et al., 2019). Practitioners must evaluate whether the computing power should be provided on-premise or in the cloud. Here, shifting workloads to regions supplied with renewable energy and carbon-efficient energy grids leads to a strong decline in carbon emissions (Henderson et al., 2022). Further, this lays the foundation for aligning the energy-intensive ML model training with the availability of renewable energy. Workloads should be scheduled flexibly according to times of renewable energy supply (Schneider et al., 2019).

Third, energy demand is largely determined by the fit between the ML model and the hardware used (Patterson et al., 2022), which is described in “**Evaluate ML Model-Hardware-Fit<sup>B, M, S</sup>**”. Li et al. (2016) elaborated that the combination of hardware setup and ML model influences the overall energy demand (D. Li et al., 2016). Further, several authors have analyzed tailored ML hardware architectures, substantially increasing the energy efficiency (Chen et al., 2017; Esser et al., 2015). Thus, the fit between the hardware and ML model will guide the decision whether to train and deploy on-premise, adjust the infrastructure, or outsource to other providers (Wu et al., 2022).

For the second phase, *Data Collection and Preparation*, we derived four DPs:

First, “**Promote Data Sparseness in Data Collection<sup>B, D, M</sup>**” describes the tradeoff between the collection of more data points (e.g., through additional sensors or external server calls), thereby increasing the CO<sub>2-eq</sub> footprint and the performance increases associated with this data point (Schneider et al., 2019). This can be achieved by gauging the performance increase of each data point before designing the acquisition (Yu, 2014). On a technical level, the sparseness

can be embraced by using efficient data collection algorithms (Rohankar et al., 2015; Xiang et al., 2013).

“**Reduce Data Dimensionality<sup>D, M</sup>**” embraces a set of techniques to reduce the energy impacts of data storage and processing by mapping inputs from higher to lesser dimensions without losing important information (Chhikara et al., 2022; Reddy et al., 2020). Yu (2014) suggested assessing the quantity of data required for the desired performance level. Further, aggregating or dropping attributes can decrease the amount of data. Similarly, it may appear reasonable to investigate the effects of larger measuring intervals (for time series data) or smaller sample sizes (for cross-sectional data) to downsize the data (Reddy et al., 2020; Schneider et al., 2019).

The importance of data retrieval and storage performance is growing, as many organizations shift their data to cloud-based storage. At the same time, the volume of generated data continuously increases (Kuschewski et al., 2023). Therefore, the DP “**Compress the Data Storage<sup>M, S</sup>**” proposes reducing the required storage and network bandwidth and thus increasing energy efficiency (Schneider et al., 2019). For instance, general-purpose compression algorithms such as Lempel–Ziv allow faster access to and manipulation of compressed data, which is especially useful if data are transferred across networks or are infrequently accessed (Schneider et al., 2019; Stolikj et al., 2012). Also, altering data formats per variable can further compress the demand for storage (Schneider et al., 2019).

Finally, the environmental efficiency can be improved by the DP of “**Stage Preprocessed Data<sup>M, S</sup>**”. Staging of preprocessed data reduces the need for recalculations (Vassiliadis, 2009). Schneider et al. (2019) suggested using intermediary stages of the processed data – such as features stores – to facilitate rapid modeling, as operations only need to be executed once. Further, the research has focused on the intelligent calculation of the time point to re-extract and re-transform the data when working with changing datasets (Vassiliadis, 2009).

The third phase, *Modeling and Training*, consists of four DPs:

The DP “**Preselect Energy-Efficient ML Models<sup>M, S</sup>**” focuses on selecting ML models from the perspective of an ML model’s lifetime carbon footprint in relation to its performance (Henderson et al., 2022; Strubell et al., 2019). It is advised to consider simpler ML models such as boosted trees instead of deep neural networks, pre-trained ML models, or transfer learning for ML models (Henderson et al., 2022). Estimates such as the floating-point operations of an ML model can guide the decision process (Schwartz et al., 2020).

Subsequently, the DP “**Eliminate Inefficiency in ML Model Architecture<sup>M</sup>**” focuses on reducing the energy consumption in the ML model architecture by optimizing

energy-intensive parts (Lee et al., 2023; Microsoft, 2023a). One example of model optimization in the context of artificial neural networks is utilizing optimized open-source code. For instance, pre-trained initializations can lead to more energy-efficient convergence (Xu, 2022). Kumar et al. (2020) suggested using profiling software (e.g., Java Energy Profiler and Optimizer) to get real-time suggestions for energy-saving adjustments.

Previous studies provided evidence that emissions from ML training can be significantly reduced when using servers in selected geographic regions at specific times (Dodge et al., 2022; Xu, 2022). Therefore, the DP “**Streamline ML Model Training Process<sup>M, S</sup>**” describes the optimization of the ML training setup to allow flexible training schedules and to lever renewable energy. In-depth analyses of different techniques can be found in Xu (2022) and Radovanovic et al. (2023).

Finally, the DP “**Optimize Hyperparameter Efficiently<sup>M</sup>**” focuses on the distinct energy consumption levels shown by different HPO techniques (Guido et al., 2022). For instance, Yarally et al. (2023) recommend avoiding random search for HPO. Further, warm-starting or zero-shot ML, a mechanism that integrates knowledge from previous executions into the current, improves the search process (Tornede et al., 2022; Wang et al., 2021; Yarally et al., 2023). Finally, pruning needs idiosyncratic consideration to improve HPO’s performance (Akiba et al., 2019).

The last phase, *Deployment and Monitoring*, contains three DPs that should be considered:

“**Streamline ML Retraining Frequency<sup>B, S</sup>**” describes optimizing the ML model retraining cycles to avoid unneeded and energy-expensive retraining (Microsoft, 2023b; Natarajan et al., 2022). To achieve this, practitioners must answer how often an ML model should be retrained (Schwartz et al., 2020). The two predominant approaches display the performance-efficiency trade-off. On the one hand, retraining models in fixed time intervals or based on conditions such as observed data drift are less accurate, but also less energy-intensive. On the other hand, constant ML model retraining is more energy-intensive but also more accurate (Microsoft, 2023b).

“**Design Computationally Sparse ML Architecture<sup>S</sup>**”, focuses on reducing the environmental costs associated with the inference phase of ML models. Considering storage, Donovan (2020) suggested analyzing the ML architecture regarding a) how long data must be stored, as storage uses much energy, and b) where data will be stored. For instance, for large datasets, on-premises storage may be more efficient (Donovan, 2020). Further, the inference type must be chosen, i.e., batch or real-time inference. The latter requires continuous server uptime and, therefore, a higher energy demand (Natarajan et al., 2022). Despite computational limitations, multiple authors have advised using edge devices owing to lower energy consumption and latencies (Zhu et al., 2022).

The DP “**Report and Monitor Environmental Sustainability<sup>B, M</sup>**” advises publishing and monitoring the power consumption, carbon emissions, training time, and hardware setup. Algorithmic and hardware advances have led to new generations of ML models with higher accuracy yet substantial energy consumption (Strubell et al., 2019). At the same time, ML researchers and organizations often omit the reporting of environmental-related metrics (Henderson et al., 2022). With advances in tracking and calculating energy demand and carbon emissions, it is easier to publish and monitor these metrics (Anthony et al., 2020; Budenny et al., 2022; Strubell et al., 2019).

#### 4.3.2 Social Dimension

The social dimension consists of 12 DPs throughout the four ML development phases. We define social sustainability as an ML model’s fairness. *Fairness* and *bias* are highly discussed in the literature and are used interchangeably (Mehrabian et al., 2022). To foster consistent wording, we use *fairness* (Kleinberg et al., 2017). We define fairness along the four elements of Colquitt (2001), who subdivide perceived fairness into distributive justice, procedural justice, interpersonal justice, and informational justice. Each DP focuses on one or more elements of this definition of perceived fairness. The initial phase, *ML Demand Specification*, contains three DPs and lays the foundation for a fair ML model:

ML models face a multitude of social concerns (i.e., fairness) while simultaneously providing potentials to deliver positive social impacts (Ayling & Chapman, 2022; Tomašev et al., 2020). Therefore, “**Assess the Social Implications<sup>B, D</sup>**” describes the examination of social objectives and potential social risks concerning the ML model and its system boundaries. On the one hand, this addresses contingent consequences that may lead to a socially unfair outcome (Blackman, 2020; van Giffen et al., 2022). Further, an open exchange about the competing objectives of all stakeholder groups helps one to find an overall compromise between fairness, accuracy, transparency, accountability, explainability, privacy, and security (Singh et al., 2022). On the other hand, the potentials for improving the social good should also be integrated into the decision process (Tomašev et al., 2020).

As highlighted in the introduction to this subsection, a holistic definition of fairness that covers each element is not possible (Kleinberg et al., 2017). Thus, the DP “**Conceptualize Definition of Fairness<sup>B, D, AT</sup>**” focuses on the selection or development of a conception of fairness for the problem at hand. Whenever possible, existing definitions and metrics of fairness should be favored (Friedler et al., 2019). To achieve this, it may be helpful to discuss the application context with domain experts and business stakeholders to determine the key features to incorporate in the following model training phase (Bellamy et al., 2019; van Giffen et al., 2022).

**“Define Human Role in the Decision Process<sup>B, D</sup>”** describes the interactions between humans and ML algorithms (van Giffen et al., 2022). As ML models are indeterministic, they require careful user-model interface design (Amershi et al., 2019). Amershi et al. (2019) proposed guidelines for ML system design that ML developers can follow, while Fabri et al. (2023) developed archetypes of human-ML interaction that can supplement the process.

In the *Data Collection and Preparation* phase, three DPs help achieve a socially sustainable ML model:

The DP **“Foster Accurate and Fair Data Collection<sup>D, M, AT</sup>”** bundles mitigation techniques in the data collection step to enhance the dataset’s fairness throughout the collection process. Several authors have shown that we must be aware of biases in the underlying data (Ferrara, 2023; Greshgorn, 2018; Holstein et al., 2019). For instance, in cancer care, the goal is to improve prevention. Thus, fair datasets may include additional factors such as ethnicity, or disability, so as to better reflect real-world circumstances (Dankwa-Mullan & Weeraratne, 2022).

**“Understand and Establish Fairness in the Dataset<sup>D, M, AT</sup>”** describes the data analysis steps to identify social unfairness. Before ML developers can avoid potential obstacles to fairness in a dataset, they must first understand where unfairness can occur (Holstein et al., 2019; Tang et al., 2023). Understanding a dataset and identifying possible areas of unfairness require data plotting, exchanges with domain experts, and even the design of proxy variables to identify relationships and reasons for social biases (Ferrara, 2023; van Giffen et al., 2022). Gu et al. (2021) recommend using interactive tools for data analysis, as they provide a better understanding of data. After understanding sources of unfairness, ML developers can initiate mitigation strategies (van Giffen et al., 2022).

Finally, **“Leverage Fair Data Sampling<sup>D, M</sup>”** describes the application of mitigation techniques in the data sampling steps (Fahse et al., 2021; Friedler et al., 2019). Several ML frameworks enable users to mitigate biases by pre-processing datasets (e.g., AI Fairness 360) (Bellamy et al., 2019). Proposed techniques include oversampling, undersampling, stratified folds, and synthetic data generation (Ferrara, 2023).

The *Modeling and Training* phase has three DPs that serve as levers for producing socially sustainable ML models:

First, the DP **“Leverage Interpretable and Fair Models<sup>B, M, AT</sup>”** describes the prioritization of interpretable and fair ML models over ‘black box’ models whenever possible. Interpretable ML models enable individuals who lack a comprehensive statistical background to understand decisions, detect errors, and bolster the due diligence process (Wang et al., 2023). Several studies have shown that interpretable ML models can perform approximately as well as black box models while providing additional benefits (Nori

et al., 2019; Wang et al., 2023). Further, models that were designed fairly can directly improve decision-making from a social perspective (van Giffen et al., 2022).

Second, **“Conduct a Fairness Evaluation<sup>B, D, M, AT</sup>”**, represents a DP that seeks to do fairness-driven ML model evaluations. This is crucial to ensure that fairness is mitigated throughout the training phase. Pagano et al. (2023), Bellamy et al. (2019), and Weerts et al. (2023) proposed using tools such as AIF360, TensorFlow Responsible AI, and Aeiquitas, which help developers to identify fairness issues early on.

Third, **“Adjust Model Parameters for Fairness<sup>M</sup>”** focuses on integrating fairness mitigation techniques (e.g., layers, loss functions) into ML models. This starts with verifying equalized odds in the ML model to guarantee the uniformity of false positives and negatives across all groups (van Giffen et al., 2022). If a need for adaptation is discovered, changes must be made to the ML model, from an adapted optimization to introducing an adversarial classifier alongside the regular model (Pagano et al., 2023; Zhang et al., 2018).

The last phase, *Deployment and Monitoring*, contains three DPs:

The first DP **“Ensure Continuous (Human) Monitoring for Fairness<sup>B, S, AT</sup>”** entails the ongoing monitoring of ML model predictions and decisions in the real-world environment (Fahse et al., 2021). This can be achieved by establishing a continual process to evaluate the ML model for fairness of predictions as new data are integrated and the ML model is retrained (Burkhardt et al., 2019). Furthermore, one can consider interrogating ML model decisions for plausibility by humans in fixed intervals (van Giffen et al., 2022).

Second, **“Convey ML Model Understanding to End Users<sup>B, M</sup>”** accounts for the perception of complexity and risk in the ML model prediction due to their non-deterministic functionality (i.e., varying text outputs of large language models based on similar input prompts due to different random number generators) (Baier et al., 2019; Westerberger et al., 2022). Especially since ML models are often user-facing and hence have wide-ranging social implications, the results of the ML models need to be transparent or at least understandable for the end user (Singh et al., 2022; van Giffen et al., 2022).

Finally, **“Enhance Transparency through Fairness Metrics<sup>B, M, AT</sup>”** describes the improvement of ML model transparency by calculating, analyzing, and publishing fairness metrics. Therefore, Pagano et al. (2023) proposed the collection of metrics such as equality of opportunity, demographic parity, and individual differential fairness. Further, introducing a multidifferential fairness auditor helps to analyze the results of classifiers regarding different groups with similar features (Gitiaux & Rangwala, 2019). Thus, this DP creates more transparency around an ML model’s decisions.

### 4.3.3 Governance Dimension

The governance dimension has nine DPs, of which four form the starting point for ensuring an ML model with a solid governance framework for the *ML Demand Specification* phase:

First, the DP **“Comply with Legal Frameworks and Company Policies<sup>B, D, AT</sup>”** emphasizes the importance of the early evaluation of legal frameworks and company policies. Generally, regulation and transparency needs increase ML projects’ governance requirements, which lead to higher costs and potential legal or ethical issues (Laato et al., 2022). Thus, gathering information and examining applicable laws and regulations for the ML usage case is a crucial first step (Gill et al., 2022). Establishing an ML governance team as an ethical review board that oversees all ML projects in an organization is advisable so as to facilitate knowledge-sharing (Floridi et al., 2018).

Second, **“Compose Diverse and Interdisciplinary ML Team<sup>B</sup>”** describes the required composition of the ML project team, as diverse and interdisciplinary teams foster creativity and mitigate biases (Burgdorf et al., 2022). Organizations must keep the ML development closely aligned to ethics and must reflect on critical voices through conversations (Barocas & Boyd, 2017). Diverse teams are characterized by different experiences as well as social and domain-specific insights. This enables the development of both innovative approaches necessary for solving challenging problems and for mitigating social risks (Burkhardt, 2019; Johnson et al., 2021). Thus, business stakeholders must account for ML team composition early in the ML demand specification (Barocas & Boyd, 2017).

Third, **“Establish a Responsible ML Mindset<sup>B</sup>”** describes a company’s inherited mindset among the ML stakeholders, especially the business stakeholders, regarding social values such as inclusion and equality in the ML development process. This can be enabled by transferring the company values to ML design in a way that values sustainability (Burkhardt, 2019). Concretely, Smith and Rustagi (2020) suggested establishing responsible ML practices as a key performance indicator and integrating them into the firm’s objectives and in individual performance interviews to drive awareness.

Fourth, **„Promote ML the democratization of ML<sup>B, M</sup>”** describes the development of an ML-skilled workforce (Ng et al., 2021) to enhance the accessibility of ML (Sundberg & Holmström, 2023). Increased ML literacy leads to more employees being able to participate in developing ML models and levers more diverse research approaches (van Giffen & Ludwig, 2023). One frequently used framework is the four Cs of AI, in which low-level and easy-to-understand information about the ML model’s concept, context, capability, and creativity is provided to the stakeholders (Talagala, 2021).

In the second phase, *Data Collection and Preparation*, two DPs should be considered from a governance perspective:

The DP, **“Establish Standards in Data Collection and Preparation<sup>B, M, AT</sup>”**, fosters clear internal guidelines for data access, generation, and collection (Dankwa-Mullan & Weeraratne, 2022). While missing standards expose sensitive data, with serious financial and reputational consequences, standards enable the provision of clean, fair, and socially safe data (Gill et al., 2022). Thus, organizations must establish data collection and preparation standards such as meta-data catalogs, data lineage, and data ownership at a governance level (Cowls et al., 2023).

**“Initiate Intra- and Interorganizational Data Democratization<sup>B</sup>”** should be facilitated, as ML models are developed based on domain-specific knowledge. Therefore, the access and competencies to understand data are necessary for successful ML implementation (van Giffen & Ludwig, 2023). Democratization can occur inside and outside an organization. For the former, companies need to issue policies and foster data exchanges in the organization to facilitate the generation of business value (Harvard Business Review Analytics Service, 2020). The latter refers to open data exchange. By participating in data exchange for non-critical data, an organization can contribute to and benefit from interorganizational ML research (De Saulles, 2020; Elgarah et al., 2005).

Considering the phase *Modeling and Training*, one DP applies at the governance level.

Contemporary ML models are often black-box models that are difficult to interpret regarding input and output data and the data processing in the ML model (Gao & Guan, 2023). To achieve company-wide adoption and participation, ML models must be relatively interpretable (Grennan et al., 2022). Interpretability refers to the extent to which a person can understand the reasoning behind a decision (Biran & Cotton, 2017; Miller, 2017). The DP **“Introduce ML Model Transparency for Active Participation<sup>M, AT</sup>”** encourages ML developers to apply interpretability methods during the development phase to embrace the discussion. These methods range from simple methods, such as an overview over the created input features, to allocate importance to global post hoc methods.

Considering the *Deployment and Monitoring*, there are two DPs:

First, the DP **“Ensure Documentation and Publishing<sup>B, M, S, AT</sup>”** helps organizations to scale their ML efforts by documentation, publishing, versioning, and metadata management of ML artifacts (e.g., code, training data) (Visengeriyeva et al., 2023). Besides development documentation, it is important to aggregate model characteristics using toolkits such as datasheets, model cards, and model registries (Mitchell et al., 2019). Thus, it is essential to maintain documentation



throughout the entire ML development process to ensure reproducibility and usability.

Second, the higher the risk of potentially harmful decisions, the more companies should engage in auditing ML models prior to deployment (Schulam & Saria, 2019). Therefore, “**Engage in ML Model Auditing**<sup>B,AT</sup>” describes the formal audit process of ML models. “Algorithm Auditing is the research and practice of assessing, mitigating, and assuring an algorithm’s safety, legality, and ethics” (Koshiyama et al., 2021, p. 2). To mitigate the risk of inadequate auditing, Laato et al. (2022) proposed directly connecting the deployment testing to organizational audit goals, ensuring ongoing auditing.

#### 4.4 SML-DPM as a Web-Based Prototype

After proposing generalized design knowledge, it needs to be brought to life (Jones & Gregor, 2007). Therefore, we developed a web-based prototype that facilitates the communication of the SML-DPM and the associated DPs to practitioners and researchers. We followed Sommerville’s (2011) prototype development process.

The prototype’s objective, as derived from our research question, is to educate practitioners about DPs so as to facilitate sustainable ML development projects. The prototype is designed to enable both practitioners and researchers to easily identify relevant DPs and to incorporate them into their ML projects, enhancing their sustainability. Thus, simplifying the dialogue about the sustainability of machine learning (I – Prototype objectives). Thus, the prototype should motivate its users to prioritize sustainability in their ML projects by using easy and comprehensible language and should encourage and simplify conversations between researchers and practitioners on sustainable ML development. It also provides detailed insights into the various DPs, ensuring that users can fully understand and apply them in their projects. Further, it should maintain the same flexibility level as the SML-DPM regarding the studied dimensions, allowing for a wide array of applications and adaptations (II – Prototype functionality). We illustrate the prototype across three categories of pages in Fig. 4.

The prototype was implemented in Vitepress,<sup>1</sup> an extendable Markdown-centered static site generator. To facilitate the communication between researchers and practitioners and allow the extension of the DPs, we developed the website openly on Codeberg<sup>2</sup> and GitHub<sup>3</sup> (III – Prototype development). Finally, the evaluation of our prototype was thoroughly conducted in EVAL4 (IV – Prototype evaluation).

<sup>1</sup> <https://vitepress.dev/>

<sup>2</sup> <https://codeberg.org/havspect/sml-dpm>

<sup>3</sup> <https://github.com/havspect/sml-dpm>

## 5 Evaluation

Assessing a developed solution’s effectiveness, appropriateness, and usefulness is a critical aspect of DSR (Sonnenberg & vom Brocke, 2012a). To demonstrate the proposed artifact’s novelty, relevance, and utility, we carried out four distinct evaluative activities, the methodology of which was detailed in Section 3. In the following, we will present the findings from the four evaluative activities (EVAL1 to EVAL4).

### 5.1 Eval 1: Ex Ante Artificial Evaluation—Problem Justification

Having discussed our research’s significance in Sects. 1 and 2, we justified the underlying problem based on the motivation and the literature presented in those two sections (EVAL1). First, the average number of ML projects carried out in companies has been anticipated to double approximately every year (Gartner, 2019). Accompanying, as elucidated in Section 1, ML holds the unintended risk of reflecting the implicit social bias at the expense of equality (Gupta et al., 2022; van Noorden & Perkel, 2023). Further, the amount of computing power needed to train current AI models has doubled every 3.4 months since 2012 (Amodei & Hernandez, 2018; Debus et al., 2023). As a result, resource use has increased exponentially, while the awareness of sustainability has also increased, for instance, 76% of green AI papers were published after 2020 (Merhi, 2023a; van Wynsberghe, 2021; Verdecchia et al., 2023). Second, only a small fraction of studies provide solutions or tools to foster sustainable ML development, and most focus only on ML during training (Verdecchia et al., 2023). While first research papers focus on the connection between sustainability dimensions and ML development (cf. Section 2.3), a comprehensive approach that addresses each phase of ML development, offers clear recommendations for mitigating sustainability risks in an integrated manner, and incorporates feedback from practitioners is missing. Hence, there is a need for an applicable framework in practice to render the overarching ML development process more sustainable.

### 5.2 Eval 2: Ex Ante Naturalistic Evaluation—Artifact Design

EVAL 2 was performed prior to the development of the artifact’s content (i.e., the DPs). As elucidated in Section 3, we conducted an academic focus group discussion to evaluate the first design of the artifact. Table 4 summarizes participant statements regarding and beyond the fulfillment of SML-DPM’s design decisions against the first three key requirements, R1 to R3. Comments with similar content were merged and sorted to the top.



1

Towards Sustainability of AI

## Towards Sustainability of AI The Sustainable ML Design Pattern Matrix (SML-DPM)

Enhance the sustainability of ML development processes across environmental, social, and governance (ESG) dimensions.

Design Patterns



**End-to-end consideration of the ML development process**  
Agnostic ML development process that enables the use in every organization.



**Holistic view on sustainability**  
Sustainability is more than just ecological or social. It is the combination that matters.



**Directly applicable for ML development stakeholders**  
The design patterns should motivate stakeholders to develop sustainable ML algorithms.



**Clear assignment of the ML development stakeholders**  
Know the responsibilities to act accordingly.



2

Towards Sustainability of AI

Sustainable development of ML

Introduction

The paper

Sustainability dimensions

Ecological

Social

Governance

ML development phases

ML Demand Specification

Data Collection and Preparation

Modeling and Training

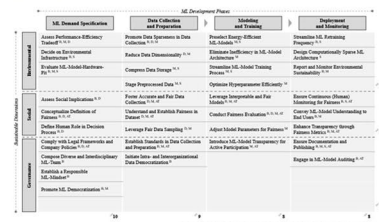
Deployment and Monitoring

### Introduction

"What AI companies are doing is like releasing race cars without breaks." Mike Häkkinen

As artificial intelligence (AI) and its subset machine learning (ML) advance, concerns about their sustainability impact grow. The emerging field "Sustainability of AI" addresses this issue, with papers exploring distinct aspects of sustainability. However, it lacks a comprehensive approach that considers all ML development phases, treats sustainability holistically, and incorporates practitioner feedback. In response, using Design Science Research, we developed the sustainable ML design pattern matrix (SML-DPM) consisting of 35 design patterns. The design patterns are structured along a four-phased ML development process, the sustainability dimensions of environmental, social, and governance (ESG), and allocated to five ML stakeholder groups. Through expert evaluation, we validated the SML-DPM's applicability and usefulness. It represents the first artifact to enhance each ML development phase along the ESG dimensions. The SML-DPM fuels advancement by aggregating distinct research, laying the groundwork for future investigations, and providing a roadmap for sustainable ML development.

The resulting matrix:



3

## Report and Monitor Environmental Sustainability

### Overview

Sustainability Dimension	Ecological
ML Development Phase	Deployment and Monitoring
ML Development Stakeholders	Business Stakeholder, ML Development

### Description

The DP "Report and Monitor Environmental Sustainability" advises publishing and monitoring power consumption, carbon emissions, training time, and hardware setup. Algorithmic and hardware advances have led to new generations of ML models with higher accuracy and simultaneously substantial energy consumption (Strubell et al., 2019). At the same time, ML researchers and organizations often omit the reporting of environmental-related metrics (Henderson et al., 2022). With advances in tracking and calculating energy demand and carbon emissions, publishing and monitoring those metrics is easier (Anthony et al., 2020; Budenny et al., 2022; Strubell et al., 2019).

### Real-world example

Initiatives like [codecarbon](#) and [carbontracker](#) offer frameworks and tools to measure and report the sustainability of models. For instance, [codecarbon](#) calculates the CO2 of machine learning applications by multiplying the hardware energy consumption with the regional carbon intensity. While the approach inherits some limitations, it gives the first tendencies for the induced carbon emissions.

#### 1. Install the package

```
conda install -c conda-forge codecarbon
```

#### 2. Initiate the tool

```
codecarbon init
```

#### 3. Save the obtained experiment\_id

```
[codecarbon]
log_level = DEBUG
save_to_api = True
experiment_id = 2bcbcb8b-850d-4692-af0d-76f6f36d79b2
```

#### 4. Monitor your ML code

```
from codecarbon import track_emissions
@track_emissions()
def your_function_to_track():
    # your code
```

#### 5. Analyse the results on the website (Online Mode) or the log-files (Offline Mode).

For more information about the usage of [codecarbon](#) we encourage you to go to their website.

### Sources

- Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and Policy Considerations for Deep Learning in NLP. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 3645–3650. <https://doi.org/10.18653/v1/P19-1355>
- Henderson, P., Hu, J., Romoff, J., Brunsell, E., Jurafsky, D., & Pineau, J. (2022). Towards the Systematic Reporting of the Energy and Carbon Footprints of Machine Learning. The Journal of Machine Learning Research, 112. <https://doi.org/10.48550/ARXIV.2007.05651>
- Anthony, L. F. W., Kanding, B., & Selvan, R. (2020). Carbontracker: Tracking and Predicting the Carbon Footprint of Training Deep Learning Models. ICM Workshop on Challenges in Deploying and Monitoring Machine Learning Systems. <https://doi.org/10.48550/ARXIV.2007.03051>
- Budenny, S. A., Lazarev, V. D., Zakharenko, N. N., Korovin, A. N., Plosskaya, O. A., Dimitrov, D. V., Akhripkin, V. S., Pavlov, I. V., Oseledets, I. V., Barsola, I. S., Egorov, I. V., Kosterina, A. A., & Zhukov, L. E. (2022). eco2AI: Carbon Emissions Tracking of Machine Learning Models as the First Step Towards Sustainable AI. Doklady Mathematics, 106(51), S118–S128. <https://doi.org/10.1134/S1064562422060230>

**Fig. 4** Web application of the SML-DPM: 1) landing page, 2) introduction to the overall SML-DPM, 3) exemplary representation of the DPs

The participants confirmed that our outlined problem setting is relevant and that providing clear operational recommendations or actions measures for ML projects to increase their sustainability is crucial for both academia and practice. They found the SML-DPM's design to be understandable and highlighted its simplicity as it follows a structured approach by proving uniformly structured DPs in a comprehensive matrix format. Notably, the participants found an end-to-end view of the ML development process useful, as there is often a strong focus on the purely technical ML

model development and training in ML projects. Nonetheless, there was criticism regarding the multitude of DPs in our artifact. Based on this and on proposed mitigation strategies, we included the fourth requirement, R4. The clear assignment of the stakeholder groups to the DPs allows to identify DPs they can engage with based on their area of expertise.

In sum, the participants supported the design of the SML-DPM and its underlying concept. They agreed that the SML-DPM covers all relevant phases of the ML

**Table 4** Qualitative comments on the design of the SML-DPM

Positive	Negative
<ul style="list-style-type: none"> <li>Structuring action measures (i.e., the DPs) toward more sustainable ML development in a compact matrix form is helpful, as it provides a quick overview</li> <li>The SML-DPM solves a real problem that is relevant or is becoming increasingly relevant owing to the current hype around AI in academia and practice</li> <li>The end-to-end consideration of the whole ML development process is necessary, since ML development teams often primarily focus only on the pure <i>modeling and training</i> phase</li> </ul>	<ul style="list-style-type: none"> <li>An overview over action measures may seem overwhelming at first glance</li> <li>A large number of different DPs could lead to stakeholders in an ML project not feeling responsible for them, because they are outside their area of expertise</li> </ul>

development process in a comprehensive way (R1). Further, they acknowledged that the SML-DPM provides a holistic view on sustainability and that it focuses on more than the often-predominant environmental dimension (R2). Finally, they also confirmed that our intended structure of the DPs allows them to be applicable to different ML developments and projects, which ensures generalizability (R3). Also considering R4, we inferred that the SML-DPM is the very first approach to comprehensively address the defined key requirements, which further supports the research need and design decisions of the SML-DPM.

### 5.3 Eval 3: Ex Post Naturalistic Evaluation—Design Pattern

The ex post evaluation of the DPs was structured in three iterations. Each iteration led to a set of modifications to the DPs. For a transparent, detailed overview over all the adjustments in each iteration, see Appendix Fig. 6. We will now describe the adjustments per iteration and will conclude with DP-agnostic findings from the interviews.

The **first iteration** – consisting of two focus groups – led to three major adjustments, resulting in 36 DPs. First, we revised and standardized the wording of all DPs to enable a direct understanding of what is meant by each. For instance, we changed the wording of “Focus Efforts on the Most Energy-Consuming phases” to “Optimize the ML Model”. Second, we merged four DPs into two so that they became mutually exclusive, i.e., we merged “Measure the Energy Intensity” and “Quantify the Carbon Footprint”. Third, we moved three patterns into different ML development phases. Finally, several stakeholder allocations have been changed (Appendix Fig. 6).

Based on the first five interviews, we conducted a **second iteration** of the DPs, because we received similar inputs on a few DPs regarding their practical application, leading to four adjustments (Appendix Fig. 6). First, we introduced the two DPs of “Promote Data Collection Sparseness” and “Reduce ML Retraining” since these are crucial in the ML development process. Second, we shifted the DP of “Develop Corporate ML Literacy” to the phase of ML

demand specification within the governance dimension and changed its wording to “Promote ML Democratization”. Third, we did three merges to improve the DPs’ unambiguity. The second iteration resulted in 35 DPs.

Based on the last seven interviews, we conducted a **third iteration** with only two minor adjustments (Appendix Fig. 6). First, we merged “Define Guidelines to Scrutinize Model Predictions” into the DP of “Ensure Continuous (Human) Monitoring for Fairness”. Second, we slightly adjusted some DPs’ wording to ensure internal consistency across all of them by applying a uniform canonical structure (Sonnenberg & vom Brocke, 2012a). After gaining no more adjustments from three subsequent interviews, we concluded the SML-DPM development.

A particular focus in the SML-DPM development process was evaluating its practical value based on interviews with experts. Besides the insights we included in the description of the DPs, four global, pattern-agnostic insights were revealed, which we will now elucidate.

**The relationship between today’s application of design patterns and increases in revenue** First, several interviews highlighted the relationships between revenue-increasing patterns and application frequency in real-world settings. The context in E4 was particularly descriptive. Here, the interviewee described the importance of the social dimension for his business by highlighting the relationships between trust, fairness, and revenue, which E4 described as “[...] *in our business, trust is generated by the preservation of privacy, which is closely related to a model’s fairness. More trust can be seen in more customers and therefore more revenue.*” The observation was grounded by E5, who described the regulatory constraints in his organization’s area of expertise and deduced that disregarding users’ privacy can lower both trust and revenue. He also stated that the DP “Assess Social Implications” is crucial when developing ML systems in areas with strong exposure to privacy-related data. Besides the social domain, this relationship was also described in other dimensions. Here, E3 and E8 emphasized the relationships between the widespread application of environmental DPs and the associated increases in revenue.

**Environmental sustainability in ML implies cost reductions** The environmental sustainability dimension directly impacts the resulting costs, as highlighted by almost all the experts. Therefore, some underlying DPs are already considered in the current ML development, such as “Assess Performance-Efficiency Tradeoff”, “Reduce Data Dimensionality”, “Promote Data Collection Sparseness”, or “Streamline ML Retraining Frequency”. For example, the reduction of data dimensionality has a direct impact. For instance, reducing the data dimensionality directly impacts on the “[...] amount of memory required to hold the training data for ML model training [...]” (E4) and, thus, directly impacts server costs. Further, promoting data collection sparseness in line with externally acquired data is reasonable “[...] because every external API call for new geospatial data causes us costs for the call itself as well as additional storage costs” (E8) or “[...] each additional data point requires new sensors, hardware to upload data, or additional data pipelines, which induce costs regarding the economic and ecological dimension” (E4). Finally, a decision about the appropriate ML retraining frequency “[...] strongly impacts on cloud server uptime” (E3) and therefore deployment costs. In contrast, in the scientific domain, the DPs in the environmental dimension were less relevant since the costs were of secondary importance (E1, E2, E10).

**Context-dependency and its focal points for sustainability** Based on the interviews, we observed a first tendency for the DPs to be strongly context dependent. On the one hand, implementing individual DPs in the three sustainability dimensions varied based on the industry context (E3, E4, E5, E7). Energy-intensive and CO<sub>2</sub>-eq-intensive companies (e.g., the manufacturing sector) are primarily concerned with DPs from the environmental dimensions, partly owing to stricter reporting requirements. E3 described this as an “[...] increasing change in awareness due to sustainability reporting [...],” which is now mandatory. Further, this industry is now experiencing strong pressure on margins, which means that the cost savings resulting from the environmental DPs are coming into focus. On the other hand, industries that process more personal data (e.g., finance and telecommunication) and thus have a stronger social impacts focus, primarily on DPs from the social and governance dimensions, mainly because they are subject to strong legal requirements: “We need to be aware of the legal and social consequences of a data breach before we start any data analytics project and whether we want to do it at all.” (E5). We also noticed a stronger shift toward “ML Demand Specification” and the “Deployment and Monitoring” phases, as laws must first be reviewed and the “[...] social as well as personal effects [...]” (E8) must be monitored closely.

**The bigger, the better does not hold true for sustainability in ML development** To better understand this insight, we divided the explanation into a data stream and an ML model stream. First, within the data stream, we can observe a positive impact on model quality and user acceptance of DPs along all three ESG dimensions that have a data-centric focus, e.g., “Reduce Data Dimensionality”, “Understand Fairness in Dataset”, or ‘Initiate Intra- and Interorganizational Data Democratization’. E9 stated that “[...] fewer data with no or almost no missing data [...] result in less noisy data, which are better suited for training our algorithms.” Similarly, “[...] elaborating in-depth on the underlying data lead us to better understand the class distribution and possible bias in the data. This allows us to sample meaningful training data” (E10). Second, in the ML model stream, larger, energy-intensive, and resource-intensive ML models based on complex technical infrastructure are often too complex for multicriteria decisions. Thus, “we prefer simple models such as decision trees because it’s easier to understand the input and output” (E6). In sum, the often-underlying paradigm of ‘the bigger, the better’ in ML development does not hold true for sustainable ML development.

#### 5.4 Eval 4: Ex Post Naturalistic Evaluation—SML-DPM

To evaluate users’ intentions to lever the SML-DPM, we conducted a case study in three ML development teams using the web-based instantiation. We will now describe each case (*Alpha*, *Beta*, and *Gamma*) and then provide insights into the fit-gap analysis based on the 35 DPs for each case. A detailed overview of the status of each DP – i.e., the DP has been considered, will not be considered, or will be considered in the future – in each case appears in Appendix Table 9. Finally, we will describe the evaluation’s results regarding the ease-of-use, usefulness of the SML-DPM and the behavioral intentions toward it.

**Case Alpha: Explainable ML product quality prediction in manufacturing** The ML team consisted of two people from the stakeholder groups B, M, and S. The project focuses on predictive quality algorithms for the printed circuit board assembly in two electronics manufacturers. It aims to develop explainable ML-based predictive quality algorithms to enable early intervention in the production process of printed circuit board assembly. To achieve this, a consolidated data warehouse was implemented to store heterogeneous data sources from various production and test lines, resulting in more than two billion data points per year. Different tabular ML models are being developed and benchmarked under considerations of explainability approaches such as Shapely Additive Explanations. The model performance will be continually monitored and evaluated by

production specialists to enable timely adjustments of process parameters and to minimize rejects.

In sum, 14 DPs were considered in the project, and 12 DPs will be considered in the future. In the project, the team was confronted with a multi-year dataset of high-frequency production data. Thus, the team initially prioritized DPs aimed at enhancing performance, reducing data points, and evaluating the ML model-hardware fit. However, during the workshop, they recognized opportunities to lower greenhouse gas (GHG) emissions by compressing existing data and leveraging optimized algorithms. Since they were among the first teams to work with these data, they realized the potential to set data collection and preparation standards by modeling the data structure according to clear rules, i.e., multilayer schema and standardized data models. Moreover, given the data's sensitivity, encompassing shift details and proprietary production knowledge, the team applied multiple DPs to ensure adherence to legal standards, regulatory compliance, and interpretable models. In this vein, the team plans to “Convey ML Model Understanding to End Users” based on the interpretable models. Finally, particularly in the governance dimension, two DPs stood out. First, the team early settled on “Initiate Intra- and Interorganizational Data Democratization” by anonymizing data to exchange results. Second, based on the DP “Promote the democratization of ML”, the team conducts half-day workshops to increase the acceptance of ML.

**Case Beta: ML-based sustainability optimizations in the printing industry** This ML team consisted of three people from the stakeholder groups M, S, and AT. Overall, the project develops ML services for a sustainable printing industry by reducing resource consumption through the detection of anomalies. First, a solution space of 16 ML use cases was derived, focusing on reducing resource consumption in industrial newspaper printing processes to identify conspicuous consumption and abnormal deviations. The team had to merge energy and mechanical process data in a managed cloud database and conducted subsequent time series data aggregation to reduce the amount of data. Based on the aggregated time-series dataset, the team will implement different ML models and will deploy the best-performing models on-site at a project partner.

In sum, the project incorporated 10 DPs and intends to explore 10 additional DPs in future work. During the workshop, it was highlighted that the participants had previously concentrated primarily on the technical development of ML prototypes, with less attention to their long-term sustainability. Nonetheless, the integration of certain DPs, particularly those relating to the environmental dimension, has begun. Monitoring GHG emissions has been a key focus of the technical lead in pinpointing opportunities for enhancement in subsequent projects. During the workshop, the team

collectively expressed surprise at the multitude of potential DPs in the social dimension. There was unanimity about the fact that they had overlooked this aspect in the past. In light of the proposed DPs, the team concurred on the necessity to evaluate their work's social implications, and they committed to helping end-users to better understand the technology.

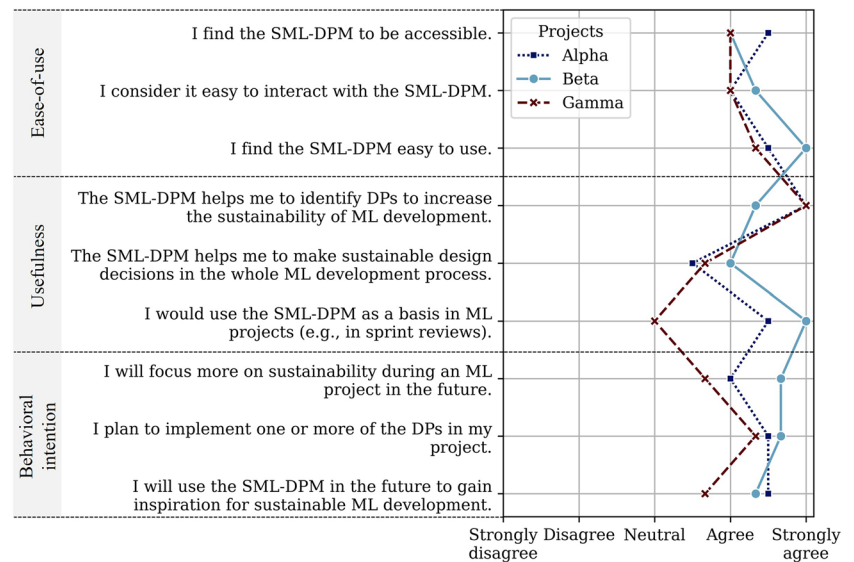
**Case Gamma: ML automation in the financial sector** This ML team consisted of three people from the stakeholder groups D, M, and S. The project aims to develop smart data and ML solutions to automate the classification of erroneous customer documents in financial services. Most of the developed use cases provide internal improvements rather than external facing innovations. During the “ML Demand Specification”, it was identified that manual checking of customer documents in rules-based processes is time-consuming and error-prone. To address this issue, an extended data lake was implemented to store historical and new customer documents. Different ML models were developed and compared to classify different document types. The best-performing model was deployed to automate the classification process. The team focused on the deployment architecture and on evaluating the ML models' impacts on human experts.

In sum, 14 DPs are already considered in the project and 6 DPs will be considered in the future. While the team was faced with sensible financial information and decisions, they only rarely focused on the social impact of their work and as such the DPs linked to the social domain. The omission of the social DPs originated from the strong regulations within the financial sector, which left the team with a limited scope of action. To comply with the regulatory requirements, the ML team has already employed nearly all DPs in the governance dimension and felt to sufficiently comply with the social dimension. However, after identifying new DPs, the team jointly agreed to focus more on the inclusion of interpretable and fair models as well as to enhance the transparency through fairness metrics. Further, the team highlighted different maturity levels within each DP. For example, the DP ‘Compress Data Storage’ was identified by the software developer as the DP with the greatest leverage due to the rudimentary implementation to date.

**Overall evaluation and implications** Figure 5 offers a comprehensive breakdown of each evaluation metric, the underlying three questions, and the project's results.

The feedback on the SML-DPM was predominantly positive across all three metrics. Participants in the case study were particularly favorable to the statements pertaining to changes in behavioral intentions. We also recognized opportunities to enhance the artifact's usefulness. Specifically, the case *Gamma* emphasized the need for more examples to improve the usefulness in developer meetings. Therefore, we included the section on practical examples in the



**Fig. 5** Evaluation results of the SML-DPM

prototype. The results in the behavioral intention category underlined our research's idea that a set of comprehensible and actionable DPs can lead to a shift towards more sustainable ML development. Overall, the number of DPs that the teams plan to incorporate in the development process was 1.8 higher than the number of employed DPs. In conjunction with the strong evaluation of the participants' behavioral intention to continue using the artifact and to focus on ML development projects' sustainability, we could confirm that the SML-DPM is applicable in real-world environments. The case studies showed that applying the SML-DPM can increase sustainability awareness in ML projects owing to its compact, multidimensional design while simultaneously highlighting the potential to integrate additional DPs.

From analyzing the case study results (i.e., selection of DPs and workshop recordings), we inferred the three following observations. First, the four overarching themes identified from the expert interviews were affirmed by the case studies. For instance, the relationship between cost-saving potential, especially for environmental DPs, and application was shown. Second, based on the results from the case studies, we identified first insights into recurring relations between different DPs. For instance, in case *Gamma*, the team highlighted the relation between some governance and ML demand specification DPs and DPs in the first two phases of the social dimension. In case *Beta*, the team highlighted the relations within the environmental dimension by outlining the need to estimate the required performance beforehand (i.e., "Assess Performance-Efficiency Tradeoff") to make suitable choices in the Modelling and Training phase. Based on the case results, the results from the interviews, and logical reasoning within the authors' team, we could identify two types of relationships. On the one hand, DPs can require other DPs to be

performed beforehand. On the other hand, DPs can improve other DPs but are not mandatory beforehand. An example of the former is the "Define Human Role in Decision Process" which is required to perform fairness evaluations. For the latter, the DP "Design Sustainable ML Architecture" enacts an example since it is significantly improved by several environmental DPs throughout the development process. Furthermore, the relations highlighted that the dimensions of social and environmental have little overlap. Contrary, the dimensions of social and governance have many relations. Besides the complex interactions, only a few DPs were mentioned as requirements for another DP. For instance, in case *Alpha* and *Gamma* both project teams started to enhance the access to and understanding of data to enable the promotion of ML democratization. Overall, we could derive multiple key<sup>4</sup> DPs that were identified as relevant among each case study, albeit sometimes implemented and sometimes on the implementation roadmap. In the social dimension, two DPs were either implemented by all case teams or soon to be implemented. In the first phase (i.e., ML Demand Specification) the definition of the human role in the decision process was a recurring theme that has already been implemented by all teams. The second DP that has been identified as a key DP is the use of fair and interpretable models. Within the discussions, the case teams agreed that both DPs represent easy-to-implement DPs and, therefore, enacted as starting points for socially sustainable ML. In the environmental dimension, a panoply of DPs was identified as key DPs among each case study. The teams uniformly describe the relationship between

<sup>4</sup> DPs were identified as 'key' when the DP was already implemented or on the implementation roadmap for each of the three case studies.



economic savings, budget limitations, and time restrictions as reasons for the implementation of the environmental DPs. For a comprehensive overview of the key DPs, we refer to Appendix Table 9. Finally, in the governance dimension, the DPs “Comply with Legal Frameworks and Company Policies”, “Establish Standards in Data Collection and Preparation”, and “Ensure Documentation and Publishing” were identified as key DPs. However, there was no clear link between implementation or implementation intention in each case study. Third, within the three case studies, we could observe external influencing factors – i.e., mainly legal and reporting requirements, regulatory compliance, or shareholder expectations—which make the implementation of certain DPs less relevant. For instance, in case *Gamma*, the strong regulations in the financial sector, such as the Banking Supervisory Requirements for IT (Leuthe et al., 2024), led to multiple guidelines for collecting and storing customer data, which made some social DPs obsolete. Additionally, in case *Alpha*, as the company is privately owned and therefore sustainability reporting obligations are only necessary to a limited extent, the DP “Report and Monitor Environmental Sustainability” is currently less relevant. Hence, those two observations directly relate to the third insight identified in the third evaluation, highlighting the context dependency for sustainability of the DPs.

In conclusion, we can confirm that the SML-DPM can improve the sustainability of ML development processes. Each case study team acknowledged that they want to include more sustainability practices in the ML development process after working with the SML-DPM, as shown in the strong evaluations in the behavioral intention category. At the same time, they mentioned some potential ways to improve the applicability and ease-of-use of the artifact. Especially in case *Gamma*, the project team – driven by numerous regulations in their agile development flow – kept reiterating the importance of an effortless and gentle integration into existing development workflows. Therefore, further iterations of the artifact should focus on its usefulness in sprint reviews to make the adoption of SML-DPM in recurring meetings easier. Nevertheless, the three case studies underscored that the SML-DPM can facilitate the sustainable development of ML throughout the ML development process, underscoring that the SML-DPM solves our research question.

## 6 Discussion

In the following discussion, we will delve into the contributions of our study (Section 6.1), explore its theoretical implications (Section 6.2), and consider the practical implications for the field (Section 6.3).

### 6.1 Contribution

With the rapid advancement of ML development, there are growing concerns about sustainability risks (Fahse et al., 2021; Gill et al., 2022; Henderson et al., 2022). To mitigate these risks and enable the different ML development stakeholders to take action, there is a need for more sustainable ML development (Schoormann et al., 2023; Wu et al., 2022). Although efforts in SAI have recently increased in the literature and in practice (Luccioni et al., 2022; Schoormann et al., 2023; Schwartz et al., 2020; Tornede et al., 2022), they remain fragmented and lack practical perspectives as well as comprehensive design approaches (e.g., Dennehy et al., 2023; Gill et al., 2022; Verdecchia et al., 2023). To fill this research gap, we set out to answer the research question: *What are design patterns that ML development stakeholders can incorporate to increase the sustainability of the ML development process?* In this vein, DPs are based on the principles of providing tangible proven solutions to recurring problems — such as ensuring the sustainability of ML — by codifying complex knowledge in an applicable and accessible way in contrast to rather abstract design principles (Dennehy et al., 2023; Gamma, 1995; Gregor et al., 2020). From a practical perspective, DPs are a solution to tackle new and recurrent challenges, such as increasing the sustainability of the ML development process. From a theoretical perspective, DPs enable to structure and unify concrete solutions for a specific challenge at hand by abstracting the design knowledge to the same level and deriving structuring elements such as dimensions or focus areas (Dickhaut et al., 2023; vom Brocke, Winter, et al., 2020). Thus, we derived the SML-DPM, which embraces 35 DPs for increased sustainability in the ML development process. To design a valuable artifact, we built on two research streams. First, we orientated to the ESG concept. Second, we inferred the ML lifecycle by relying on the four ML process phases and attributed the DPs to clearly defined ML development stakeholders. Overall, The SML-DPM was developed along a four-step DSR approach. To ensure the SML-DPM’s practical relevance, it was developed in close alignment with four practical requirements, an artificial evaluation in a focus group, and naturalistic evaluations in multiple focus groups and interviews with experts (Gregor & Hevner, 2013). Moreover, based on a demonstration in three case studies, the SML-DPM’s ease-of-use, usefulness, and behavioral intention in line with the TAM were evaluated (Davis, 1989; Sonnenberg & vom Brocke, 2012a). This led to three main contributions:

*First, the SML-DPM bridges the gap between the ESG sustainability concept and the end-to-end ML development process.* Previously, the research into sustainable

ML projects was fragmented across various disciplines, with, e.g., social fairness studied independently of environmental sustainability (Gupta et al., 2022; van Giffen et al., 2022; Veit & Thatcher, 2023). The SML-DPM unifies these research dimensions into a single artifact, reflecting sustainability's multifacetedness, encompassing environmental, social, ethical, and governance aspects (Pappas et al., 2023). The ESG dimension's structure the artifact on the y-axis, while the end-to-end ML development phases structure it on the x-axis. This accounts for phase-specific sustainability concerns, as all the phases of an ML project must be considered if one is to achieve sustainable ML (Papagiannidis et al., 2023). This contributes van Wynsberghe (2021), who defined SAI as "a movement to foster change in the entire lifecycle of AI [...]". Thus, the artifact adds to the integration of sustainability and digital transformation, echoing the research agendas of Vassilakopoulou and Hustad (2023) as well as Mikalef et al. (2022).

*Second, we provide the 35 DP with justificatory knowledge from expert insights.* We synthesized the extant SAI knowledge through a systematic literature review enriched with focus group discussions and interviews with experts. The resulting and validated DPs are specific to each cell (i.e., a combination of the sustainability dimension and the ML development phase) and are clearly assigned to one or more ML development stakeholders. This became particularly clear in the two naturalistic evaluations (EVAL3 and EVAL4) as the DPs must be clearly identifiable for each stakeholder and suitable for the current ML development phase of the project. Yet, to enact their potential instead of providing ineffective intelligence, the DPs need to be precisely described, integrable in existing workflows, padded with examples, and straightforwardly communicated (see EVAL4). Overall, the application of DPs in the realm of sustainability and ML highlighted the suitability of DPs to aggregate and convey solutions to recurring problems. Furthermore, the DPs, which provide an easy entry point to SAI through their standardized tripartite structure (i.e., action title, theoretical description, and practical justificatory knowledge), enable one to increase the sustainability of the ML development process, which is a relevant building block for an overarching responsible digital transformation that is ethical and sustainable (Pappas et al., 2023; Veit & Thatcher, 2023). This contributes to the call for action on how organizations can optimize their digital transformation projects' sustainability impacts, considering their entire lifecycle (Pappas et al., 2023). Here, especially the DPs in the social and governance dimension make it possible to counteract the digital divide phenomenon as reducing digital inequalities is critical for a sustainable digital transformation (Vassilakopoulou & Hustad, 2023).

For instance, the DPs provide clear mitigation strategies for the three major sources of algorithmic bias (Akte et al., 2021): 1) data bias addressed by "Leverage Fair Data Sampling" and "Understand and Establish Fairness in Dataset", 2) method bias addressed by "Leverage Fair and Interpretable Models" and "Adjust Model Parameters for Fairness", and 3) societal bias addressed by "Establish Standards in Data Collection and Preparation" and "Compose Diverse and Interdisciplinary ML-Team".

*Third, we contribute by providing extensive naturalistic insights into the SML-DPM's application based on its web-based prototype.* On the one hand, we derived four global, pattern-agnostic insights (i.e. "the bigger the better does not behold true for sustainability in ML"). These insights allow further theorizing on the interaction between sustainability and ML. On the other hand, in the three ML case studies (EVAL4) based on "Case Alpha: Explainable ML product quality prediction in manufacturing", "Case Beta: ML-based sustainability optimizations in the printing industry", and "Case Gamma: ML automation in the financial sector", we could highlight the importance of covering all the sustainability dimensions (Case Beta), each development phase (Case Alpha and Gamma), and the stakeholders (Case Beta and Gamma). In this vein, the SML-DPM has been proven to show new and applicable DPs that can be used in the ML development process. Above all, the SML-DPM can specifically stimulate discussion in ML teams about which practices are necessary to increase the sustainability of the ML development process owing to its compact multidimensional structure, which is necessary for SAI. To support the selection and prioritization process, we further extracted preliminary findings about the relations between different DPs. The web-based prototype contributes to stimulating the discussion by providing an accessible path for researchers and practitioners to incorporate sustainable practices. Thereby following the calls of Shneiderman (2021) and Dennehy et al. (2023) to provide actionable research.

## 6.2 Theoretical Implications

There are two primary theoretical implications of our study and its resulting SML-DPM, as they bring together, structure, and enrich existing knowledge on how to increase the sustainability of the ML development process and lay the foundation for further theorizing in the SAI field.

*First, our work has opened a new discussion on how to structure SAI and, subsequently, what SAI comprises regarding clear and implementable practices.* We specifically investigated the relationship between the end-to-end ML development phases and the three sustainability dimensions of environmental, social, and governance.

Thus, the results shed light on the end-to-end process view of ML by opening a discussion about the different phases of ML development and the unique sustainability challenges faced in each of these (Papagiannidis et al., 2023). We hereby extended the quest to include an additional perspective on resource allocation in ML development projects (Papagiannidis et al., 2023; Pappas et al., 2023). This is particularly important for the value-creation of ML projects, since it provides a clear picture of how different phases and resources are levered to create business value while considering sustainability (Enholm et al., 2022; Vial et al., 2023). We thereby extend previous works by Rohde et al. (2024), who evaluated sustainability along the phases of organizational ML embedding with a more granular perspective. Regarding the sustainability dimension, the research has predominantly focused on one of the three dimensions (c.f. Fahse et al., 2021; Schneider et al., 2023; Verdecchia et al., 2023). However, Veit and Thatcher (2023), among others, have highlighted the importance of a joint consideration of the dimensions when focusing on sustainability in IS. We have responded to this call by enabling researchers to build on our dimensions to develop artifacts that contribute holistically to SAI, such as SAI archetypes and development paths or maturity levels of the DPs. Moreover, we extend this structure (i.e., the relationship between the ML development phases and the sustainability dimensions) by including the different ML development stakeholders and linking them to actionable practices (e.g., the DPs) to address challenges in the SAI field. This dimensional link is noteworthy, since it focuses on a human-centric perspective, which overall constitutes the IS discipline (Vössing et al., 2022). Thus, multiple researchers from different disciplines (e.g., computer science, law, ethics, and IS) can join the discussion about mitigating the sustainability risks associated with the ‘dark side of AI’. A discussion that has intensified over the past few years in academia (Rohde et al., 2021; Schoormann et al., 2023; Verdecchia et al., 2023), while a clear output-driven discussion toward implementable practices has only emerged in recent examples (Dennehy et al., 2023; Patterson et al., 2022; Polyviou & Zamani, 2023; Shneiderman, 2021). Notably, Rohde et al. (2024) made the first endeavor in this realm by providing assessment indicators for sustainable ML development along the EESG (i.e., environmental, ecological, social, and governance) dimensions and phases of organizational ML embedding. While their managerial focus (i.e., by focusing on a broad development process and indicators) provides a valuable assessment tool, our work provides researchers with a specific framework for the actionable sustainability improvement of ML projects. Researchers may contribute to this discussion by extending our efforts and deriving specific

consequences of action per stakeholder or by developing a more nuanced understanding of each stakeholder’s roles in relation to sustainability. Further, researchers could build upon the preliminary relationships between the DPs in each phase and dimension to provide holistic governance frameworks that promote sustainability holistically. Overall, the SML-DPM contributes to the nascent IS field that focuses on SAI by providing a multidimensionally structured framework. Hence, our results emphasize the importance of discussing SAI by shifting from a one-sided to a multifaced perspective.

*Second, by presenting the 35 DPs and validating them with subject-matter experts, we have responded to calls for research into merging hitherto fragmented theoretical knowledge and validating it with practitioner views, facilitating theorizing toward sustainable AI* (Veit & Thatcher, 2023; Verdecchia et al., 2023). The derived 35 DPs in their entirety can serve as an impetus for a nascent design theory in the field of SAI (Jones & Gregor, 2007). Jones and Gregor (2007) describe eight components that make up a design theory, including principles of form and function and justificatory knowledge. Thus, the DPs could be levered as a basis for the principles of form and function as those should serve as a blueprint for conducting sustainable ML development projects. In this vein, related works on nascent design theories subdivided those principles into the hierarchical structure of design requirements, design principles, and lastly, design features as their smallest unit (e.g., Dickhaut et al. (2023), Herm et al. (2022), Jonas et al. (2023)). As such design features represent measures for action, our extensive number of 35 DPs can be used as a basis for those. Thereafter, these design features are aggregated into a distinct number of design principles to meet the design requirements raised. For instance, an exemplary design principle of “Provide explanations” to meet the design requirements “Increase end user trust” and “Increase ML system accessibility” could contain the design features “Leverage Interpretable and Fair Models”, “Convey ML Model Understanding to End Users”, and “Introduce ML Model Transparency for Active Participation” resulting from our DPs (Herm et al., 2022). Additionally, the insights derived from the expert interviews and the three case studies may act as foundations for the justificatory knowledge. Such a nascent SAI design theory can further guide the improved understanding and creation of solution-oriented guidelines, key components, and theoretical knowledge on how to design sustainable AI systems along their entire lifecycle, i.e., prescription for SAI’s design and action (Gregor & Hevner, 2013; Jones & Gregor, 2007). Furthermore, the use of DPs to provide reusable development and managerial patterns adds to a growing stream of research which seeks to provide actionable practices for researchers and

practitioners while maintaining the generalizability to derive further theoretical leaps (see Papagiannidis et al., 2023; Lu et al., 2024). While the aforementioned stream is predominantly driven by structured literature reviews, we add another perspective to incorporate practical insights in the development by following the DSR process. The collection of DPs in the SML-DPM facilitates the discourse in ML research to identify new artifacts to improve the ML development process' sustainability. Moreover, we went one step further than the SML-DPM, establishing four overarching insights that highlight the relationships between different sustainability dimensions and the DPs in real-world settings, providing input for justificatory knowledge in the realm of a nascent design theory (Jones & Gregor, 2007). The first two insights (i.e., "the relationship between today's application of design patterns and revenue advances" and "environmental sustainability in ML implies cost reductions") connect to the technical and economic performance in terms of revenue and cost with a sustainability dimension. This extends insights from previous research in the green IS/IT domain about the relation between green IT/IS and economic benefits to the development of ML models (Veit & Thatcher, 2023). Further research could systematically analyze the relationships between cost reductions, revenue increases, and the proposed DPs to increase sustainability. Additionally, based on the expository instantiation of the SML-DPM as a web-based prototype to make the artifact more tangible and to conduct a case study in three ML development teams, we are able to observe increased behavioral intentions of the ML development stakeholders to include further, albeit different (Appendix Table 9), DPs toward enhancing the sustainability of their ML developments in their respective industries. In this vein, the third insight (i.e., "context-dependency and their focal points for sustainability") emphasizes the importance of context and organizational goals in ML sustainability research. We thereby enrich the research of Merhi (2023a), who identified organizational culture (i.e., values and implemented practices) as one of the enablers of responsible AI development with a practical validation. Consequently, organizations cannot employ all presented DPs to enact sustainable AI development because the resources of organizations are limited, and the change towards sustainable AI development has only recently started. Yet, the presentation of our DPs in interviews and the three cases highlighted the potential to extend current application levels. For instance, research on social DPs could leverage the strong relationship between the application of DPs and the potential for cost-savings, which was affirmed in the three case studies, to increase the application of social DPs. Additionally, the application could be increased by accessibly presenting social

DPs for technical-oriented project teams (see EVAL4). The fourth insight (i.e., "the bigger the better does not behold true for sustainability in ML development") highlights the conflict of the economic, environmental, and social dimensions, as the current race towards bigger and better ML models counteracts the sustainability dimensions (Margherita & Braccini, 2023; Pappas et al., 2023). A development which has, e.g., recently spiked in large language model development to provide compact and sustainable responsible models (Banks & Warkentin, 2024). Hence, those naturalistic insights allow researchers to develop measurement instruments to help practitioners assess the progress and, therefore, measure the behavioral intentions in their efforts. All in all, the SML-DPM adds to the SAI knowledge base by consolidating knowledge from previously separate research streams (e.g., 'green AI', 'sustainability of AI', 'fairness in AI', and 'responsible AI') enriched with practical insights and lays the foundation for further theorizing and understanding of SAI endeavors.

### 6.3 Practical Implications

From a practical perspective, the SML-DPM holds two primary implications for the decision-makers and the stakeholder groups in the ML development process (e.g., business stakeholders, domain experts, software developers) that especially became evident in our interviews and the three case studies.

*First, the different stakeholders can lever the SML-DPM to capture the status quo and to develop a vision regarding the sustainability of the ML development process.* In practice, ML projects are typically complex and versatile. Thereby, the four concise ML development phases on the SML-DPM's x-axis specify the key steps when planning and executing ML projects from a higher-level perspective, ensuring applicability independently of the chosen ML development process. The subdivision on the y-axis clearly depicts which DP promotes which of the three sustainability dimensions, facilitating a rigorous consideration of all pertinent sustainability dimensions by practitioners. Thus, the SML-DPM provides the different stakeholders with a coherent and conclusive picture of the DPs for increasing the sustainability of the ML development process and mitigating sustainability risk associated with the 'dark side of AI' (Mikalef et al., 2022; Schoormann et al., 2023). This will become increasingly important in the upcoming years, as, on the one hand, more and more ML projects are conducted in organizations (Gartner, 2019; Merhi, 2023b), and, on the other hand, sustainability reporting will be demanded more strongly or even become mandatory, aiming at improving transparency about corporate sustainability performance (Truant



et al., 2023). Hence, especially given the predominant position of the ESG framework in sustainability reporting due to, e.g., the EU's corporate sustainability reporting directive (European Parliament, 2022), our SML-DPM allows practitioners to align their sustainability reporting duties with their efforts to increase the sustainability the ML development process. In particular, in the chapters on digitalization projects in the corporate ESG sustainability reports, the SML-DPM can be used as input for structuring the reported ML sustainability initiatives. Thus, this simplifies the embedding of SAI practices in organizations. As part of this organizational embedding, the SML-DPM can also be used to facilitate meaningful strategic discussions among the various ML stakeholders, e.g., in recurring meetings such as sprint reviews in agile ML projects. Those discussions can, specifically in conjunction with the list of sustainability ML indicators from Rohde et al. (2024), serve as a foundation for a self-assessment. As shown in the three case studies, practitioners can use the SML-DPM as part of a fit-gap analysis to systematically review the sustainability measures of their ML projects. The resulting insights allow them to derive a desired target state within their efforts toward SAI. Consequently, the SML-DPM acts as a diagnostic tool to gain insights into blind spots, capture the status quo, and derive an SAI vision. Further, providing a web-based instantiation of the SML-DPM ensures the widespread communication of the results.

*Second, the SML-DPM guides the different stakeholders in implementing DPs for the sustainable development of ML.* At a higher level, the SML-DPM provides a holistic overview structured along practically known sustainability dimensions and clearly assigned to the stakeholders (A. Kumar, 2022; T.-T. Li et al., 2021). In detail, the DPs provide methodological support and act as a simple point of entry, as they are easy to understand. Thus, ML development stakeholders can use the SML-DPM to identify DPs that fit their role (e.g., business stakeholder, domain expert), the current project phase (e.g., “Modeling and Training”), and the sustainability focus (e.g., environmental). In the event that the DPs are applied and communicated to the end users, this can lead to an increase in trust in both the ML development process and in the resulting ML system. This is crucial for harvesting the potential benefits from AI and for fostering trust and resilience at a societal level (Dennehy et al., 2023; Dubber et al., 2020). This adds to recent academic discourse on shifting from pure principles to implementable recommendations and practices (Dennehy et al., 2023). By using our SML-DPM, researchers and practitioners can integrate the DPs into their development workflows to ensure the sustainability of their ML development. Here, the web-based instantiation provides an easy point of entry that

can facilitate the discussing and implementing the DPs. In this sense, Schwartz et al. (2020) emphasized the importance of integrating sustainable practices even when only researching for more performant ML models. Further, it is not enough to only apply the DPs once when designing and deploying an ML system; rather, there is a need for continual consideration as for instance new data emerge or environmental conditions change (Papagiannidis et al., 2023). Nonetheless, integrating additional DPs poses the challenge of increased complexity in the development process. How to balance the sustainability gains and the introduced complexity is a tough task, yet it is something that organizations willing to focus on increasing sustainability ought to consider. For instance, HuggingFace offers the possibility to publish the amount of CO<sub>2</sub> emitted for publicly available ML models and provides a native solution to finding low-emission ML models. Indeed, while much more needs to be done, it is promising that pioneering firms such as HuggingFace are taking important steps that will bolster the sustainability of AI, irrespective of a company's size (Gupta et al., 2022; Polyviou & Zamani, 2023).

## 6.4 Limitations and Further Research

Our research has limitations regarding the SML-DPM, the derived DPs, the evaluation, and associated findings; these offer avenues for further research.

First, the wide-ranging motivation and holistic problem definition (i.e., three sustainability dimensions and four ML process phases) limits the scope for the collection of DPs that are both extensive enough that there is enough literature and specific enough to derive applicable and domain-agnostic DPs. Thus, we have not provided practitioners with ready-to-use solutions, but abstract DPs that must be translated into engineering practices. Second, one limitation relates to the SML-DPM's structure, which was influenced by the design decisions taken and the four key requirements. The SML-DPM is designed to achieve internal consistency and comprehensiveness by aligning with established ESG sustainability dimensions and the four ML process phases. This design approach primarily follows a deductive perspective, focusing on established ESG dimensions rather than the triple bottom line. However, future work could explore the development of decision principles from an inductive perspective, in which measures toward SAI are derived from empirical data. Third, while we provide the first indications regarding the relations between different DPs, the complex intervention should be further analyzed in subsequent works. One avenue for future research is the use of a quantitative approach to account for potential interactions among DPs and should highlight positive and



negative reinforcements. This could enable the deviation of generalizable knowledge about the implementation of sustainable ML and foster discussions. For instance, further research could analyze how social and environmental DPs could be intervened to bolster both dimensions simultaneously. Fourth, while expert interviews and case studies are a proven approach to explore an emerging phenomenon, individuals' perspectives are highly subjective. Although we consulted experts from different organizational contexts (e.g., industry, size), we cannot guarantee that we have covered all the relevant perspectives on the DPs, since ML is a constantly evolving topic (Dennehy et al., 2023). Here, future research can use a confirmatory study (e.g., a Delphi study) to substantiate our DPs, and can refine or extend them. Also, further studies can extend our case study insights by performing longitudinal case studies to provide insights on the uses of our DPs in real-world scenarios and provide practical implementation guidelines. Especially the extension of the DPs with design features or practical implementations could significantly enhance their usability. Furthermore, the evaluation of DPs in relation to key performance indicators could strengthen the DPs.

## 7 Conclusion

As ML models continue to be integrated more rapidly, the associated sustainability risks are only slowly becoming recognized (Cowls et al., 2023; van Wylsberghe, 2021). Yet, as more and more ML models are deployed, giving them an ever-greater influence, it is important to do so with a focus on sustainability to reduce its potential negative impact on environmental sustainability and social fairness (Gupta et al., 2022; Papagiannidis et al., 2023; Schoormann et al., 2023). To address this issue, previous work in the field of SAI must be consolidated, made operational, and enriched by practical perspectives, enabling to increase the sustainability of the ML development process (Dennehy et al., 2023; Shneiderman, 2021; Verdecchia et al., 2023). Therefore, we developed the

SML-DPM, a holistic framework providing practitioners from research and practice with guidelines to develop sustainable ML projects. Our framework provides 35 DPs along the entire ML development process (i.e., “ML Demand Specification”, “Data Collection and Preparation”, “Modeling and Training”, and “Deployment and Monitoring”) segmented within the ESG dimensions (i.e., “environmental”, “social”, and “governance”) and attributes them to five ML development groups (e.g., “Domain Expert”). The SML-DPM was developed along a four-step research approach based on the DSR-paradigm (Hevner et al., 2004; Peffers et al., 2007) and the evaluation patterns of Sonnenberg and vom Brocke (2012b) in close alignment with four literature-grounded key requirements. Four distinct evaluation activities (i.e., Eval 1–4) were conducted using naturalistic evaluations through focus groups and semi-structured interviews with subject matter experts, followed by an evaluation in three real-world case studies (e.g., “Case Beta: ML-based sustainability optimizations in the printing industry”) using a web-based instantiation to prove the SML-DPMs applicability, completeness, and usefulness. This process ensures the satisfaction of all four design requirements and makes the SML-DPM a foundation for future research whose relevance will drastically increase as it bridges the gap between the ESG sustainability concept (relevant for organizations) and the end-to-end ML development process (relevant for the value-creation of ML projects). Notwithstanding the limitations of the SML-DPM based on its deductive design decisions and the compact representation of the DPs, we are confident that our study on the ML development process' sustainability provides researchers and practitioners with a novel overview and a systematic understanding of the sustainable development of ML. We expect the results to serve as both a foundation and stimulation for fellow researchers to continue the growing scientific discussion in the SAI field. We consider this work to be a cornerstone of advanced research into the sustainable development of ML and SAI as a whole in future work.

## Appendix

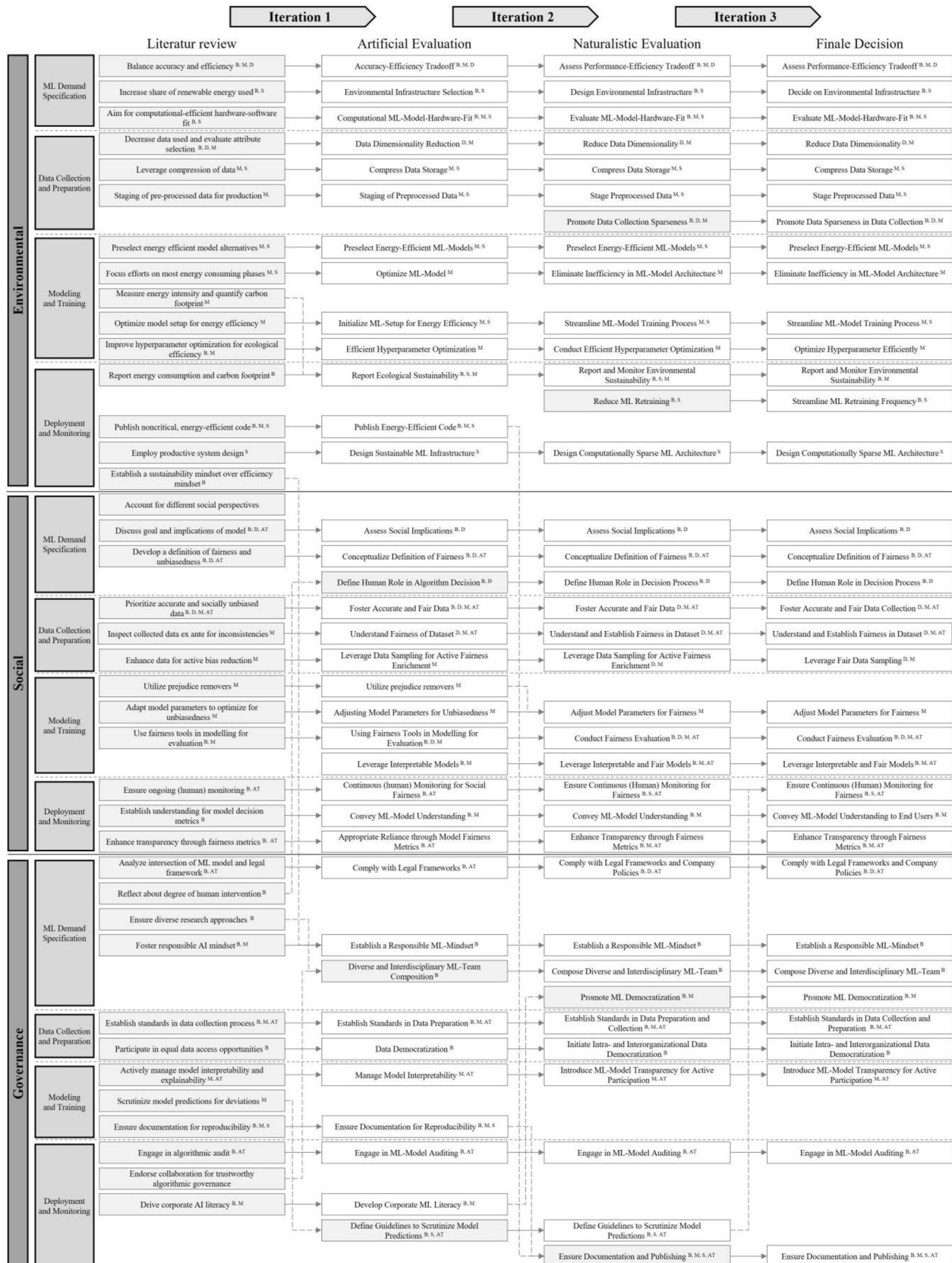


Fig. 6 Iterative development of the DPs

[illegible][illegible]

[illegible]

	Iteration 1	Iteration 2	Iteration 3
#No. of new DPs	0	2	0
#No. of DP modifications	16	13	3
#No. of DP eliminations or mergers	3	1	1
#No. of DPs at the start of the iteration	39	36	35
Saturation	0.5128	0.5596	0.8857

Evaluation Metrics	Question	Source
Ease-of-use	I find the SML-DPM to be accessible	(Adipat et al., 2011; Zacharias et al., 2022)
	I consider it easy to interact with the SML-DPM	
	I find the SML-DPM easy to use	
Usefulness	The SML-DPM helps me to identify DPs to increase the sustainability of ML development	(Davis, 1989; Meske & Bunde, 2022)
	The SML-DPM helps me to make sustainable design decisions in the whole ML development process	
	I would use the SML-DPM as a basis in ML projects (e.g., in sprint reviews)	
Behavioral intention	I will focus more on sustainability during an ML project in the future	(Iivari et al., 2021)
	I plan to implement one or more of the DPs in my project	
	I will use the SML-DPM in the future to gain inspiration for sustainable ML development	



**Table 8** Justificatory knowledge for the DPs

Environmental	Phase	Design Patterns	Justificatory knowledge
Environmental	ML Demand Specification	Assess Performance-Efficiency Tradeoff	"It is imperative for us to establish a predetermined level of performance for the model to establish realistic expectations for the business and avoid unnecessarily enhancing the require computational power." (E5)
		Decide on Environmental Infrastructure	"When purchasing computational power from major cloud providers for a new project, we now pay greater attention to ensuring that the data center has a low carbon footprint." (E5)
		Evaluate ML-Model-Hardware-Fit	"Many large companies are now offering specialized CPUs, TPUs, and GPUs. Here we look at which new developments are really relevant for our current project." (E1)
		Reduce Data Dimensionality	"Classical reduction methods like PCA are recommended at my organization because they have a significant impact on the training time of the models." (E8)
		Compress Data Storage	"We are already integrating data compression algorithms into our development projects, because besides memory reduction they also lead to faster upload and download times and thus significantly reduce the development time for large data sets." (E2)
Modeling and Training	Data Collection and Preparation	Stage Preprocessed Data	"We had a use case with voice data and had to re-run the feature pipeline every weekend, which took several hours because we couldn't cache the data." (E3)
		Promote Data Collection Sparseness	"Especially for new products, each additional data point requires new sensors, hardware to upload data, or additional data pipelines which induces costs regarding the economic and ecological dimension." (E4)
		Preselect Energy-Efficient ML-Models	"More and more SaaS solutions already offer energy-efficient models. This makes it easier for us to integrate them operationally into our systems, since we do not need models with highly complexity." (E3)
		Eliminate Inefficiency in ML-Model Architecture	"Especially for models with long runtimes per prediction, we already pay attention to the runtimes of the individual layers during development." (E1)
		Streamline ML-Model Training Process	"The incentivization of smaller cloud bills often leads us to the adoption of training schedules and thus leveraging the power of renewable energies." (E10)
Deployment	Optimize Hyperparameter Efficiently	Optimize Hyperparameter Efficiently	"We try to do everything via Genetic Optimization Algorithms because we get much faster to useful hyperparameters and avoid any kind of random grid search in our projects." (E11)
		Streamline ML Retraining Frequency	"We altered our procedure to formulate questions (e.g., 'Which performance decrease starts a new retraining?') before deploying the solution to enable customers to enable customers to adjust to changing predictions." (E3)
		Design Computationally Sparse ML Architecture	"After analyzing the system requirements and energy consumption, we decided to move our ML infrastructure to the edge. Therefore, generating substantial savings in terms of energy consumption, costs and latency." (E4)
		Report and Monitor Environmental Sustainability	"If developer publish the carbon footprint of their models, for instance at Hugging-Face Hub, we integrate those values into our selection process." (E8)

Table 8 (continued)

Phase	Design Patterns	Justificatory knowledge
Social	ML Demand Specification	<p>“Especially for autonomous models like ‘Reinforcement Learning’ it is important to assess the integration between model and environment to assess the implications of bad decisions.” (E4)</p> <p>“Omitting sensitive groups from the fairness definition can directly influence the predictions and therefore the objective level of fairness for these groups.” (E9)</p> <p>“At some point in the ideation, we have to answer ‘who is accountable for the decisions of the model?’ and ‘how can this person or group influence and understand the decisions?’” (E11)</p>
	Data Collection and Preparation	<p>“We are developing models for multiple sides. Therefore, it is important to aggregate a dataset which treats every side in the same frequency.” (E5)</p> <p>“If we know the distribution of the dataset before developing a model (for example by using explorative data analytics), we can develop significantly better models” (E5)</p>
	Modeling and Training	<p>“We always use more sophisticated sampling techniques, like stratified splits, when working with imbalanced datasets” (E8)</p> <p>“When developing new models, we strongly recommend our ML developer to include layers to purge bias [...]” (E8)</p> <p>“For each data-centric model, we have to examine whether some features led to unfairness regarding certain groups.” (E11)</p> <p>“[...] especially since the difference between the performance of huge models and interpretable models became so small and lead to understandable results, we often include interpretable models in our model selection.” (E10)</p>
	Deployment	<p>“Human Monitoring is an expensive monitoring step which we want to omit. However, due to its importance we must engage in it and see no desirable supplement without keeping the human in the loop.” (E5)</p> <p>“It is from utmost importance that the user understands the associated risks and limitations of the model to facilitate a responsible usage of the ML-Model.” (E4)</p> <p>“Before deploying our models, we have to write extensive reports to explain the functionality and fairness of the model to ensure the correctness of our models in relations to laws.” (E11)</p>

Table 8 (continued)

Governance	Phase	Design Patterns	Justificatory knowledge
ML Demand Specification	ML Demand Specification	Comply with Legal Frameworks and Company Policies	“Because of our highly regulated environment, we start every project by examining and setting the boundaries for using customer data.” (E5)
		Establish a Responsible ML-Mindset	“We had to learn the hard way that irresponsible usage of ML can lead to distrust at our customers. Especially, the usage of questionable datasets can harm your business when working with large cooperation.” (E8)
		Compose Diverse and Interdisciplinary ML-Team	“The right composition of ML teams is discussed at a very early stage [...]. At the same time, it must be considered whether there must always be all roles for each team or whether in simple cases no data scientist is needed and the business stakeholder(s) can choose a low-code approach themselves?” (E3)
		Promote ML Democratization	“In order for the requirements for an ML model to be defined in a meaningful and realistic way, the stakeholders must understand how an ML model works and what it can do. This is precisely why the training of employees is significant for the development.” (E5)
Data Collection and Preparation	Data Collection and Preparation	Establish Standards in Data Preparation and Collection	“Due to the large number of different data pools and definitions, we have introduced a data catalog in which metadata is recorded for each data point, thus structuring and standardizing the data analysis process more clearly.” (E8)
		Initiate Intra- and Interorganizational Data Democratization	“In workshops several of our employees learned techniques associated with ‘Data Storytelling’ to learn the skills to derive data-driven conclusions and participate in data-driven discussions.” (E4)
Modeling and Training	Modeling and Training	Introduce Model Transparency for Active Participation	“Especially in business areas where privacy concerning data of users are used, the integration of domain experts becomes significant. To foster an exchange and participation we encourage developers to use explainable AI technologies.” (E5)
		Engage in ML-Model Auditing	“In highly regulated markets, like our market, it is unavoidable to engage in ML Auditing to actively proof the lawfulness of our algorithms. (E11)
Deployment	Deployment	Ensure Documentation and Publishing	“Internal we try to document as much as we can. Facing to our customers it is a difficult decision since too much information can lead to product piracy.” (E5)

**Table 9** Classification of the DPs based on the three case study projects

			Classification of the DPs		
Sustainability Dimension	ML Development Phase	Design Pattern	Project Alpha	Project Beta	Project Gamma
Environmental	ML Demand Specification	Assess Performance-Efficiency Tradeoff	→	√	-
		Decide on Environmental Infrastructure	-	→	-
		Evaluate ML-Model-Hardware-Fit	√	-	→
	Data Collection and Preparation	Promote Data Sparseness in Data Collection	√	-	→
		Reduce Data Dimensionality	→	√	→
		Compress Data Storage	→	√	√
		Stage Preprocessed Data	√	→	→
		Modeling and Training	Preselect Energy-Efficient ML Models	√	√
	Eliminate Inefficiency in ML-Model Architecture		-	-	√
	Streamline ML-Model Training Process		→	→	→
	Optimize Hyperparameter Efficiently		√	→	→
	Deployment and Monitoring	Streamline ML Retraining Frequency	→	→	√
		Design Computationally Sparse ML Architecture	√	-	-
		Report and Monitor Environmental Sustainability	-	→	-
		Social	ML Demand Specification	Assess Social Implications	√
Conceptualize Definition of Fairness	-			-	-
Define Human Role in Decision Process	→			→	→
Data Collection and Preparation	Foster Accurate and Fair Data Collection		√	-	-
	Understand and Establish Fairness in Dataset		-	-	-
	Leverage Fair Data Sampling		→	-	-
Modeling and Training	Leverage Fair and Interpretable Models		√	√	√
	Conduct Fairness Evaluation		-	-	-
	Adjust Model Parameters for Fairness		→	-	-
Deployment and Monitoring	Ensure Continuous (Human) Monitoring for Fairness		→	√	-
	Convey ML-Model Understanding to End Users		→	→	-
	Enhance Transparency through Fairness Metrics		-	-	√
Governance	ML Demand Specification	Comply with Legal Frameworks and Company Policies	√	√	→
		Compose Diverse and Interdisciplinary ML-Team	→	√	→
		Establish a Responsible ML-Mindset	-	-	→
		Promote ML Democratization	√	-	→
	Data Collection and Preparation	Establish Standards in Data Collection and Preparation	→	√	→
		Initiate Intra- and Interorganizational Data Democratization	√	-	√
	Modeling and Training	Introduce ML-Model Transparency for Active Participation	√	→	-
	Deployment and Monitoring	Ensure Documentation and Publishing	√	√	→
		Engage in ML-Model Auditing	-	-	→
	Sum (√)			14	10
Sum (-)			9	15	14
Sum (→)			12	10	14

√/Design pattern is already considered

→Design pattern will be considered in the future

-Design pattern will not be considered



**Authors' Contributions** Daniel Leuthe: Conceptualization, Investigation, Data Curation, Supervision, Methodology, Visualization, Writing—Original draft, Writing—Reviewing & Editing. Tim Meyer-Holatz: Conceptualization, Investigation, Data Curation, Supervision, Methodology, Visualization, Writing—Original draft, Writing—Reviewing & Editing. Tobias Plank: Investigation, Data Curation, Methodology, Visualization, Writing—Original draft. Anja Senkmüller: Investigation, Data Curation, Methodology, Visualization, Writing—Original draft.

**Funding** Open Access funding enabled and organized by Projekt DEAL.

**Data Availability** The data that support the findings of this paper are available upon request.

## Declarations

**Ethics Approval and Consent to Participate** Not applicable.

**Competing Interests** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Adipat, Z., & Zhou. (2011). The Effects of Tree-View Based Presentation Adaptation on Mobile Web Browsing. *MIS Quarterly*, 35(1), 99. <https://doi.org/10.2307/23043491>
- Ågerfalk, P. J. (2020). Artificial intelligence as digital agency. *European Journal of Information Systems*, 29(1), 1–8. <https://doi.org/10.1080/0960085X.2020.1721947>
- Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. (2019). Optuna: A Next-generation Hyperparameter Optimization Framework. *KDD '19: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2623–2631. <https://doi.org/10.1145/3292500.3330701>
- Akter, S., McCarthy, G., Sajib, S., Michael, K., Dwivedi, Y. K., D'Ambra, J., & Shen, K. N. (2021). Algorithmic bias in data-driven innovation in the age of AI. *International Journal of Information Management*, 60. <https://doi.org/10.1016/j.ijinfomgt.2021.102387>
- Allen, J., Freed, A., & Chandrasekaran, S. (2017). *Adapt DevOps to cognitive and artificial intelligence systems*. <https://developer.ibm.com/articles/cc-devops-artificial-intelligence-cognitive/>. Accessed 3 May 2023.
- Amershi, S., Begel, A., Bird, C., DeLine, R., Gall, H., Kamar, E., Nagappan, N., Nushi, B., & Zimmermann, T. (2019). Software Engineering for Machine Learning: A Case Study. *2019 IEEE/ACM 41st International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP)*, 291–300. <https://doi.org/10.1109/ICSE-SEIP.2019.00042>
- Amodei, D., & Hernandez, D. (2018). *AI and compute*. OpenAI. <https://openai.com/research/ai-and-compute>. Accessed 28 Feb 2024.
- Ando, H., Cousins, R., & Young, C. (2014). Achieving saturation in thematic analysis: Development and refinement of a codebook. *Comprehensive Psychology*, 3, 03.CP.3.4. <https://doi.org/10.2466/03.CP.3.4>
- Anthony, L. F. W., Kanding, B., & Selvan, R. (2020). Carbontracker: Tracking and Predicting the Carbon Footprint of Training Deep Learning Models. *ICML Workshop on Challenges in Deploying and Monitoring Machine Learning Systems*. <https://doi.org/10.48550/ARXIV.2007.03051>
- Ayling, J., & Chapman, A. (2022). Putting AI ethics to work: Are the tools fit for purpose? *AI and Ethics*, 2(3), 405–429. <https://doi.org/10.1007/s43681-021-00084-x>
- Ayres, P., & Sweller, J. (2014). The Split-Attention Principle in Multimedia Learning. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (2nd ed., pp. 206–226). Cambridge University Press. <https://doi.org/10.1017/CBO9781139547369.011>
- Baier, L., Jöhren, F., & Seebacher, S. (2019). Challenges in the deployment and operation of machine learning in practice. *Proceedings of the 27th European Conference on Information Systems (ECIS)*. <https://doi.org/10.5445/IR/1000095028>
- Banks, J., & Warkentin, T. (2024, February 21). *Gemma: Introducing new state-of-the-art open models*. <https://blog.google/technology/developers/gemma-open-models/>. Accessed 5 Mar 2024.
- Barocas, S., & Boyd, D. (2017). Engaging the ethics of data science in practice. *Communications of the ACM*, 60(11), 23–25. <https://doi.org/10.1145/3144172>
- Baroni, I., Re Calegari, G., Scandolari, D., & Celino, I. (2022). AI-TAM: A model to investigate user acceptance and collaborative intention inhuman-in-the-loop AI applications. *Human-Computer Interaction*, 9(1), 1–21. <https://doi.org/10.15346/hc.v9i1.134>
- Baxter, D., Gao, J., Case, K., Harding, J., Young, B., Cochrane, S., & Dani, S. (2007). An engineering design knowledge reuse methodology using process modelling. *Research in Engineering Design*, 18(1), 37–48. <https://doi.org/10.1007/s00163-007-0028-8>
- Belanger, F. (2012). Theorizing in Information Systems Research Using Focus Groups. *Australasian Journal of Information Systems*, 17(2), 2. <https://doi.org/10.3127/ajis.v17i2.695>
- Bellamy, R. K. E., Dey, K., Hind, M., Hoffman, S. C., Houde, S., Kannan, K., Lohia, P., Martino, J., Mehta, S., Mojsilovic, A., Nagar, S., Ramamurthy, K. N., Richards, J., Saha, D., Sattigeri, P., Singh, M., Varshney, K. R., & Zhang, Y. (2019). AI Fairness 360: An extensible toolkit for detecting and mitigating algorithmic bias. *IBM Journal of Research and Development*, 63(4/5), 4:1-4:15. <https://doi.org/10.1147/JRD.2019.2942287>
- Benbya, H., Pachidi, S., & Jarvenpaa, S. L. (2021). Special Issue Editorial: Artificial Intelligence in Organizations: Implications for Information Systems Research. *Journal of the Association for Information Systems*, 22(2), 281–303. <https://doi.org/10.17705/1jais.00662>
- Berente, N., Gu, B., Recker, J., & Santhanam, R. (2021). Managing Artificial Intelligence. *MIS Quarterly*, 45, 1433–1450.
- Bhatt, U., Xiang, A., Sharma, S., Weller, A., Taly, A., Jia, Y., Ghosh, J., Puri, R., Moura, J. M. F., & Eckersley, P. (2020). Explainable machine learning in deployment. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 648–657. <https://doi.org/10.1145/3351095.3375624>

- Biran, O., & Cotton, C. (2017). Explanation and justification in machine learning: A survey. *IJCAI-17 Workshop on Explainable AI (XAI)*, 8(1) 8–13.
- Blackman, R. (2020, October 15). A Practical Guide to Building Ethical AI. *Harvard Business Review*. <https://hbr.org/2020/10/a-practical-guide-to-building-ethical-ai>. Accessed 17 Feb 2024.
- vom Brocke, J., Winter, R., Hevner, A., & Maedche, A. (2020). Special Issue Editorial – Accumulation and Evolution of Design Knowledge in Design Science Research: A Journey Through Time and Space. *Journal of the Association for Information Systems*, 21(3), 520–544. <https://doi.org/10.17705/1jais.00611>
- Brownlee, A. E. I., Adair, J., Haraldsson, S. O., & Jabbo, J. (2021). Exploring the Accuracy – Energy Trade-off in Machine Learning. *IEEE/ACM International Workshop on Genetic Improvement (GI)*, 2021, 11–18. <https://doi.org/10.1109/GI52543.2021.00011>
- Brundtland, G. H. (1987). Our Common Future—Call for Action. *Environmental Conservation*, 14(4), 291–294. <https://doi.org/10.1017/S0376892900016805>
- Budennyy, S. A., Lazarev, V. D., Zakharenko, N. N., Korovin, A. N., Plosskaya, O. A., Dimitrov, D. V., Akhripkin, V. S., Pavlov, I. V., Oseledets, I. V., Barsola, I. S., Egorov, I. V., Kosterina, A. A., & Zhukov, L. E. (2022). Eco2AI: Carbon Emissions Tracking of Machine Learning Models as the First Step Towards Sustainable AI. *Doklady Mathematics*, 106(S1), 118–128. <https://doi.org/10.1134/S1064562422060230>
- Burgdorf, K., Rostamzadeh, N., Srinivasan, R., & Lena, J. (2022). Looking at Creative ML Blindspots with a Sociological Lens (arXiv:2205.13683). <https://doi.org/10.48550/arXiv.2205.13683>
- Burkhardt. (2019). Leading your organization to responsible AI | McKinsey. <https://www.mckinsey.com/capabilities/quantumblack/our-insights/leading-your-organization-to-responsible-ai>
- Chen, Y.-H., Krishna, T., Emer, J. S., & Sze, V. (2017). Eyeriss: An Energy-Efficient Reconfigurable Accelerator for Deep Convolutional Neural Networks. *IEEE Journal of Solid-State Circuits*, 52(1), 127–138. <https://doi.org/10.1109/JSSC.2016.2616357>
- Chhikara, P., Jain, N., Tekchandani, R., & Kumar, N. (2022). Data dimensionality reduction techniques for Industry 4.0: Research results, challenges, and future research directions. *Software Practice and Experience*, 52(3), 658–688. <https://doi.org/10.1002/spe.2876>
- Collins, C., Dennehy, D., Conboy, K., & Mikalef, P. (2021). Artificial intelligence in information systems research: A systematic literature review and research agenda. *International Journal of Information Management*, 60, 102383. <https://doi.org/10.1016/j.ijinfomgt.2021.102383>
- Colquitt, J. A. (2001). On the dimensionality of organizational justice: A construct validation of a measure. *Journal of Applied Psychology*, 86(3), 386–400. <https://doi.org/10.1037/0021-9010.86.3.386>
- Cooper, R. B., & Zmud, R. W. (1990). Information Technology Implementation Research: A Technological Diffusion Approach. *Management Science*, 36(2), 123–139. <https://doi.org/10.1287/mnsc.36.2.123>
- Cowls, J., Tsamados, A., Taddeo, M., & Floridi, L. (2023). The AI gambit: Leveraging artificial intelligence to combat climate change—opportunities, challenges, and recommendations. *AI & SOCIETY*, 38(1), 283–307. <https://doi.org/10.1007/s00146-021-01294-x>
- Dankwa-Mullan, I., & Weeraratne, D. (2022). Artificial Intelligence and Machine Learning Technologies in Cancer Care: Addressing Disparities, Bias, and Data Diversity. *Cancer Discovery*, 12(6), 1423–1427. <https://doi.org/10.1158/2159-8290.CD-22-0373>
- Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3), 319. <https://doi.org/10.2307/249008>
- Debus, C., Piraud, M., Streit, A., Theis, F., & Götz, M. (2023). Reporting electricity consumption is essential for sustainable AI. *Nature Machine Intelligence*, 5(11), 1176–1178. <https://doi.org/10.1038/s42256-023-00750-1>
- Montreal Declaration. (2017). *Montreal Declaration for a Responsible Development of Artificial Intelligence*. [https://declarationmontreal-ia-responsible.com/wp-content/uploads/2023/04/UdeM\\_Decl-IA-Resp\\_LA-Declaration-ENG\\_WEB\\_09-07-19.pdf](https://declarationmontreal-ia-responsible.com/wp-content/uploads/2023/04/UdeM_Decl-IA-Resp_LA-Declaration-ENG_WEB_09-07-19.pdf). Accessed 21 Sept 2023.
- Dennehy, D., Griva, A., Pouloudi, N., Dwivedi, Y. K., Mäntymäki, M., & Pappas, I. O. (2023). Artificial Intelligence (AI) and Information Systems: Perspectives to Responsible AI. *Information Systems Frontiers*, 25(1), 1–7. <https://doi.org/10.1007/s10796-022-10365-3>
- Desislavov, R., Martínez-Plumed, F., & Hernández-Orallo, J. (2023). Trends in AI inference energy consumption: Beyond the performance-vs-parameter laws of deep learning. *Sustainable Computing: Informatics and Systems*, 38, 100857. <https://doi.org/10.1016/j.suscom.2023.100857>
- Dickhaut, E., Janson, A., Söllner, M., & Leimeister, J. M. (2023). Lawfulness by design – development and evaluation of lawful design patterns to consider legal requirements. *European Journal of Information Systems*, 1–28. <https://doi.org/10.1080/0960085X.2023.2174050>
- Dodge, J., Prewitt, T., Tachet des Combes, R., Odmark, E., Schwartz, R., Strubell, E., Luccioni, A. S., Smith, N. A., DeCario, N., & Buchanan, W. (2022). Measuring the Carbon Intensity of AI in Cloud Instances. *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, 1877–1894. <https://doi.org/10.1145/3531146.3533234>
- Donovan, R. (2020). How to put machine learning models into production. Stack Overflow Blog. <https://stackoverflow.blog/2020/10/12/how-to-put-machine-learning-models-into-production/>. Accessed 19 Jun 2023.
- Drempetic, S., Klein, C., & Zwergel, B. (2020). The Influence of Firm Size on the ESG Score: Corporate Sustainability Ratings Under Review. *Journal of Business Ethics*, 167(2), 333–360. <https://doi.org/10.1007/s10551-019-04164-1>
- Dubber, M. D., Pasquale, F., & Das, S. (Eds.). (2020). *The Oxford Handbook of Ethics of AI* (1st ed.). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780190067397.001.0001>
- Elgarah, W., Falaleeva, N., Saunders, C. C., Ilie, V., Shim, J. T., Courtney, J., & F. (2005). Data exchange in interorganizational relationships: Review through multiple conceptual lenses. *ACM SIGMIS Database: The DATABASE for Advances in Information Systems*, 36(1), 8–29. <https://doi.org/10.1145/1047070.1047073>
- Elkington. (2018). 25 Years Ago I Coined the Phrase ‘Triple Bottom Line.’ Here’s Why It’s Time to Rethink It. <https://hbsp.harvard.edu/product/H04E7P-PDF-ENG>. Accessed 23 Jun 2023.
- Enholm, I. M., Papagiannidis, E., Mikalef, P., & Krogstie, J. (2022). Artificial Intelligence and Business Value: A Literature Review. *Information Systems Frontiers*, 24(5), 1709–1734. <https://doi.org/10.1007/s10796-021-10186-w>
- Esser, S. K., Appuswamy, R., Merolla, P., Arthur, J. V., & Modha, D. S. (2015). Backpropagation for energy-efficient neuromorphic computing. In C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, & R. Garnett (Eds.), *Advances in neural information processing systems* (28). Curran Associates, Inc.
- Fabri, L., Häckel, B., Oberländer, A. M., Rieg, M., & Stohr, A. (2023). Disentangling Human-AI Hybrids: Conceptualizing the Interworking of Humans and AI-Enabled Systems. *Business & Information Systems Engineering*. <https://doi.org/10.1007/s12599-023-00810-1>
- Fahse, T., Huber, V., & Van Giffen, B. (2021). Managing Bias in Machine Learning Projects. *Innovation through*

- Information Systems, 47, 94–109. [https://doi.org/10.1007/978-3-030-86797-3\\_7](https://doi.org/10.1007/978-3-030-86797-3_7)
- Fayyad, U., Haussler, D., & Stolorz, P. (1996). KDD for science data analysis: Issues and examples. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining* (pp. 50–56). AAAI Press
- Ferrara, E. (2023). *Fairness And Bias in Artificial Intelligence: A Brief Survey of Sources, Impacts, And Mitigation Strategies* (arXiv:2304.07683). arXiv.
- Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds and Machines*, 28(4), 689–707. <https://doi.org/10.1007/s11023-018-9482-5>
- Friedler, S. A., Scheidegger, C., Venkatasubramanian, S., Choudhary, S., Hamilton, E. P., & Roth, D. (2019). A comparative study of fairness-enhancing interventions in machine learning. *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 329–338. <https://doi.org/10.1145/3287560.3287589>
- Gamma, E. (Ed.). (1995). *Design patterns: Elements of reusable object-oriented software*. Addison-Wesley.
- Gao, L., & Guan, L. (2023). *Interpretability of machine learning: Recent advances and future prospects* (arXiv:2305.00537). arXiv. Accessed 10 Sept 2023.
- Gartner. (2019). *Gartner Survey Reveals Leading Organizations Expect to Double the Number of AI Projects In Place Within the Next Year*. Gartner. <https://www.gartner.com/en/newsroom/press-releases/2019-07-15-gartner-survey-reveals-leading-organizations-expect-t>. Accessed 11 Sept 2023.
- Gill, N., Mathur, A., & Conde, M. V. (2022). A Brief Overview of AI Governance for Responsible Machine Learning Systems. *Workshop on Trustworthy and Socially Responsible Machine Learning, NeurIPS 2022*. <https://doi.org/10.48550/arXiv.2211.13130>
- Gitiaux, X., & Rangwala, H. (2019). *Multi-differential fairness auditor for black box classifiers* (arXiv:1903.07609). arXiv.
- Glavič, P., & Lukman, R. (2007). Review of sustainability terms and their definitions. *Journal of Cleaner Production*, 15(18), 1875–1885. <https://doi.org/10.1016/j.jclepro.2006.12.006>
- Goel, K., Fehrer, T., Röglinger, M., & Wynn, M. T. (2023). Not Here, But There: Human Resource Allocation Patterns. In C. Di Francescomarino, A. Burattin, C. Janiesch, & S. Sadiq (Eds.), *Business Process Management* (Vol. 14159, pp. 377–394). Springer Nature Switzerland. [https://doi.org/10.1007/978-3-031-41620-0\\_22](https://doi.org/10.1007/978-3-031-41620-0_22)
- Graf-Drasch, V., Keller, R., Meindl, O., & Röhrich, F. (2023). The Design of Citizen-Centric Green IS in Sustainable Smart Districts. *Business & Information Systems Engineering*, 65(5), 521–538. <https://doi.org/10.1007/s12599-023-00821-y>
- Gramlich, V., Guggenberger, T., Principato, M., Schellinger, B., & Urbach, N. (2023). A multivocal literature review of decentralized finance: Current knowledge and future research avenues. *Electronic Markets*, 33(1), 11. <https://doi.org/10.1007/s12525-023-00637-4>
- Green, B., Johnson, C., & Adams, A. (2006). Writing narrative literature reviews for peer-reviewed journals: Secrets of the trade. *Journal of Chiropractic Medicine*, 5(3), 101–117. [https://doi.org/10.1016/S0899-3467\(07\)60142-6](https://doi.org/10.1016/S0899-3467(07)60142-6)
- Gregor, S., & Hevner, A. R. (2013). Positioning and Presenting Design Science Research for Maximum Impact. *MIS Quarterly*, 37(2), 337–355. <https://doi.org/10.25300/MISQ/2013/37.2.01>
- Gregor, S., Kruse, L., & Seidel, S. (2020). Research Perspectives: The Anatomy of a Design Principle. *Journal of the Association for Information Systems*, 21, 1622–1652. <https://doi.org/10.17705/1jais.00649>
- Grennan, L., Kremer, A., Singla, A., & Zipparo, P. (2022). *Explainable AI: Getting it right in business*. <https://www.mckinsey.com/capabilities/quantumblack/our-insights/why-businesses-need-explainable-ai-and-how-to-deliver-it>. Accessed 10 Sept 2023.
- Greshgorn, D. (2018). *If AI is going to be the world's doctor, it needs better textbooks* [Newspage]. <https://qz.com/1367177/if-ai-is-going-to-be-the-worlds-doctor-it-needs-better-textbooks>. Accessed 6 Sept 2023.
- Gu, Z., Yan, J. N., & Rzeszutowski, J. M. (2021). Understanding User Sensemaking in Machine Learning Fairness Assessment Systems. *Proceedings of the Web Conference, 2021*, 658–668. <https://doi.org/10.1145/3442381.3450092>
- Guido, R., Groccia, M. C., & Conforti, D. (2022). A hyperparameter tuning approach for cost-sensitive support vector machine classifiers. *Soft Computing*. <https://doi.org/10.1007/s00500-022-06768-8>
- Gupta, M., Parra, C. M., & Dennehy, D. (2022). Questioning Racial and Gender Bias in AI-based Recommendations: Do Espoused National Cultural Values Matter? *Information Systems Frontiers*, 24(5), 1465–1481. <https://doi.org/10.1007/s10796-021-10156-2>
- Harvard Business Review Analytics Service. (2020). *Turning data into unmatched business value*. <https://services.google.com/fh/files/blogs/hbr-turn-data-into-business-value-report.pdf>. Accessed 6 Sept 2023.
- Hausladen, I., & Schosser, M. (2020). Towards a maturity model for big data analytics in airline network planning. *Journal of Air Transport Management*, 82, 101721. <https://doi.org/10.1016/j.jairtraman.2019.101721>
- Henderson, P., Hu, J., Romoff, J., Brunskill, E., Jurafsky, D., & Pineau, J. (2022). Towards the Systematic Reporting of the Energy and Carbon Footprints of Machine Learning. *The Journal of Machine Learning Research*, 1(2). <https://doi.org/10.48550/arXiv.2002.05651>
- Hennink, M., & Kaiser, B. N. (2022). Sample sizes for saturation in qualitative research: A systematic review of empirical tests. *Social Science & Medicine*, 292, 114523. <https://doi.org/10.1016/j.socscimed.2021.114523>
- Herm, L.-V., Steinbach, T., Wanner, J., & Janiesch, C. (2022). A nascent design theory for explainable intelligent systems. *Electronic Markets*, 32(4), 2185–2205. <https://doi.org/10.1007/s12525-022-00606-3>
- Hevner, M., & Park, & Ram. (2004). Design Science in Information Systems Research. *MIS Quarterly*, 28(1), 75. <https://doi.org/10.2307/25148625>
- Holstein, K., Wortman Vaughan, J., Daumé, H., Dudik, M., & Wallach, H. (2019). Improving Fairness in Machine Learning Systems: What Do Industry Practitioners Need? *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–16. <https://doi.org/10.1145/3290605.3300830>
- Hopwood, B., Mellor, M., & O'Brien, G. (2005). Sustainable development: Mapping different approaches. *Sustainable Development*, 13(1), 38–52. <https://doi.org/10.1002/sd.244>
- Huang, J., Kauffman, R. J., & Ma, D. (2015). Pricing strategy for cloud computing: Damaged services perspective. *Decision Support Systems*, 78, 80–92. <https://doi.org/10.1016/j.dss.2014.11.001>
- Iivari, J., Hansen, R. P., & Haj-Bolouri, A. (2021). A proposal for minimum reusability evaluation of design principles. *European Journal of Information Systems*, 30(3), 286–303. <https://doi.org/10.1080/0960085X.2020.1793697>
- Isil, O., & Hernke, M. T. (2017). The Triple Bottom Line: A Critical Review from a Transdisciplinary Perspective. *Business Strategy and the Environment*, 26(8), 1235–1251. <https://doi.org/10.1002/bse.1982>
- Johnson, M., Albizri, A., & Harfouche, A. (2021). Responsible Artificial Intelligence in Healthcare: Predicting and Preventing



- Insurance Claim Denials for Economic and Social Wellbeing. *Information Systems Frontiers*. <https://doi.org/10.1007/s10796-021-10137-5>
- Jonas, C., Lockl, J., Röglinger, M., & Weidlich, R. (2023). Designing a wearable IoT-based bladder level monitoring system for neurogenic bladder patients. *European Journal of Information Systems*, 1–23. <https://doi.org/10.1080/0960085X.2023.2283173>
- Jones, D., & Gregor, S. (2007). The Anatomy of a Design Theory. *Journal of the Association for Information Systems*, 8(5), 312–335. <https://doi.org/10.17705/1jais.00129>
- Ketter, W., Padmanabhan, B., Pant, G., & Raghu, T. S. (2020). Special Issue Editorial: Addressing Societal Challenges through Analytics: An ESG ICE Framework and Research Agenda. *Journal of the Association for Information Systems*, 21(5), 1115–1127. <https://doi.org/10.17705/1jais.00631>
- Kim, S. W., Kong, J. H., Lee, S. W., & Lee, S. (2022). Recent Advances of Artificial Intelligence in Manufacturing Industrial Sectors: A Review. *International Journal of Precision Engineering and Manufacturing*, 23(1), 111–129. <https://doi.org/10.1007/s12541-021-00600-3>
- King, W. R., & He, J. (2005). Understanding the Role and Methods of Meta-Analysis in IS Research. *Communications of the Association for Information Systems*, 16(1). <https://doi.org/10.17705/ICAIS.01632>
- Kleinberg, J., Mullainathan, S., & Raghavan, M. (2017). Inherent Trade-Offs in the Fair Determination of Risk Scores. *8th Innovations in Theoretical Computer Science Conference (ITCS 2017)*. <https://doi.org/10.4230/LIPICS.ITCS.2017.43>
- Koniakou, V. (2023). From the “rush to ethics” to the “race for governance” in Artificial Intelligence. *Information Systems Frontiers*, 25(1), 71–102. <https://doi.org/10.1007/s10796-022-10300-6>
- Koshiyama, A., Kazim, E., Treleaven, P., Rai, P., Szpruch, L., Pavey, G., Ahamat, G., Leutner, F., Goebel, R., Knight, A., Adams, J., Hitrova, C., Barnett, J., Nachev, P., Barber, D., Chamorro-Premuzic, T., Klemmer, K., Gregorovic, M., Khan, S., & Lomas, E. (2021). Towards Algorithm Auditing: A Survey on Managing Legal, Ethical and Technological Risks of AI, ML and Associated Algorithms. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3778998>
- Kreuzberger, D., Kühl, N., & Hirschl, S. (2023). Machine Learning Operations (MLOps): Overview, Definition, and Architecture. *IEEE Access*, 11, 31866–31879. <https://doi.org/10.1109/ACCESS.2023.3262138>
- Krueger, R. A., & Casey, M. A. (2015). *Focus groups: A practical guide for applied research* (5th ed.). SAGE.
- Krueger, R. A. (1988). *Focus groups: A practical guide for applied research*. (p. 197). Sage Publications, Inc.
- Kumar, M., Zhang, X., Liu, L., Wang, Y., & Shi, W. (2020). Energy-Efficient Machine Learning on the Edges. *IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW)*, 2020, 912–921. <https://doi.org/10.1109/IPDPSW50202.2020.00153>
- Kumar, A. (2022). ESG & AI / Machine Learning Use Cases. *Data Analytics*. <https://vitalflux.com/esg-ai-machine-learning-use-cases/>
- Kuschewski, M., Sauerwein, D., Alhomssi, A., & Leis, V. (2023). BtrBlocks: Efficient Columnar Compression for Data Lakes. *Proceedings of the ACM on Management of Data*, 1(2), 1–26. <https://doi.org/10.1145/3589263>
- Laato, S., Birkstedt, T., Mäntymäki, M., Minkinen, M., & Mikkonen, T. (2022). AI governance in the system development life cycle: Insights on responsible machine learning engineering. *Proceedings of the 1st International Conference on AI Engineering: Software Engineering for AI*, 113–123. <https://doi.org/10.1145/3522664.3528598>
- Lee, J. J., Park, S. -H., & Eo, J. (2012). Assessing and managing an organization's green IT maturity (Vol. 11, Iss. 3, p. 3). *MIS Quarterly Executive*.
- Lee, J., Mukhanov, L., Molahosseini, A. S., Minhas, U., Hua, Y., Del Rincon, J. M., Dichev, K., Hong, C.-H., & Vandierendonck, H. (2023). Resource-Efficient Convolutional Networks: A Survey on Model-, Arithmetic-, and Implementation-Level Techniques. *ACM Computing Surveys*, 55, 1–36. <https://doi.org/10.1145/3587095>
- Leuthe, D., Weiß, F., Dersch, J., & Bitzer, M. (2024). Towards secure cloud-computing in FinTechs – An Artefact for prioritizing information security measures. In *Proceedings of the 57th Hawaii International Conference on System Sciences*.
- Li, T.-T., Wang, K., Sueyoshi, T., & Wang, D. D. (2021). ESG: Research Progress and Future Prospects. *Sustainability*, 13(21), 11663. <https://doi.org/10.3390/su132111663>
- Li, D., Chen, X., Becchi, M., & Zong, Z. (2016). Evaluating the Energy Efficiency of Deep Convolutional Neural Networks on CPUs and GPUs. *2016 IEEE International Conferences on Big Data and Cloud Computing, Social Computing and Networking, Sustainable Computing and Communications*, 477–484. <https://doi.org/10.1109/BDCloud-SocialCom-SustainCom.2016.76>
- Lu, Q., Zhu, L., Xu, X., Whittle, J., Zowghi, D., & Jacquet, A. (2024). Responsible AI Pattern Catalogue: A Collection of Best Practices for AI Governance and Engineering. *ACM Computing Surveys*, 56(7), 1–35. <https://doi.org/10.1145/3626234>
- Luccioni, S., Mueller, Z., & Raw, N. (2022). *CO2 Emissions and the Hugging Face Hub: Leading the Charge*. <https://huggingface.co/blog/carbon-emissions-on-the-hub>. Accessed 10 Sept 2023.
- Mäntymäki, M., Minkinen, M., Birkstedt, T., & Viljanen, M. (2022). Defining organizational AI governance. *AI and Ethics*, 2(4), 603–609. <https://doi.org/10.1007/s43681-022-00143-x>
- Margherita, E. G., & Braccini, A. M. (2023). Industry 4.0 Technologies in Flexible Manufacturing for Sustainable Organizational Value: Reflections from a Multiple Case Study of Italian Manufacturers. *Information Systems Frontiers*, 25(3), 995–1016. <https://doi.org/10.1007/s10796-020-10047-y>
- Martínez-Fernández, S., Franch, X., & Durán, F. (2023). *Towards green AI-based software systems: An architecture-centric approach (GAISSA)* (arXiv:2307.09964). arXiv. <http://arxiv.org/abs/2307.09964>
- Massuga, F., Larson, M. A., Kuhl, M. R., & Doliveira, S. L. D. (2023). The influence of global governance on the sustainable performance of countries. *Environment, Development and Sustainability*. <https://doi.org/10.1007/s10668-023-03827-4>
- McCoy, S., Galletta, D. F., & King, W. R. (2007). Applying TAM across cultures: The need for caution. *European Journal of Information Systems*, 16(1), 81–90. <https://doi.org/10.1057/palgrave.ejis.3000659>
- Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2022). A Survey on Bias and Fairness in Machine Learning. *ACM Computing Surveys*, 54(6). <https://doi.org/10.1145/3457607>
- Merhi, M. I. (2023a). An Assessment of the Barriers Impacting Responsible Artificial Intelligence. *Information Systems Frontiers*, 25(3), 1147–1160. <https://doi.org/10.1007/s10796-022-10276-3>
- Merhi, M. I. (2023b). An evaluation of the critical success factors impacting artificial intelligence implementation. *International Journal of Information Management*, 69, 102545. <https://doi.org/10.1016/j.ijinfomgt.2022.102545>
- Meske, C., & Bunde, E. (2022). Design Principles for User Interfaces in AI-Based Decision Support Systems: The Case of Explainable Hate Speech Detection. *Information Systems Frontiers*. <https://doi.org/10.1007/s10796-021-10234-5>



- Microsoft. (2023a). *Code With Engineering Playbook* [Book]. <https://microsoft.github.io/code-with-engineering-playbook>. Accessed 2 Sept 2023.
- Microsoft. (2023b). *Machine learning inference during deployment—Cloud Adoption Framework*. <https://learn.microsoft.com/en-us/azure/cloud-adoption-framework/innovate/best-practices/ml-deployment-inference>. Accessed 2 Sept 2023.
- Mikalef, P., Conboy, K., Lundström, J. E., & Popović, A. (2022). Thinking responsibly about responsible AI and ‘the dark side’ of AI. *European Journal of Information Systems*, 31(3), 257–268. <https://doi.org/10.1080/0960085X.2022.2026621>
- Miles, M. B., & Huberman, A. M. (2009). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Sage.
- Miller, T. (2017). *Explanation in Artificial Intelligence: Insights from the Social Sciences*. <https://doi.org/10.48550/ARXIV.1706.07269>
- Missimer, M., Robèrt, K.-H., & Broman, G. (2017). A strategic approach to social sustainability – Part 2: A principle-based definition. *Journal of Cleaner Production*, 140, 42–52. <https://doi.org/10.1016/j.jclepro.2016.04.059>
- Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., & Gebru, T. (2019). Model Cards for Model Reporting. *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 220–229. <https://doi.org/10.1145/3287560.3287596>
- Myers, M. D., & Newman, M. (2007). The qualitative interview in IS research: Examining the craft. *Information and Organization*, 17(1), 2–26. <https://doi.org/10.1016/j.infoandorg.2006.11.001>
- Naser, M. Z. (2023). Do We Need Exotic Models? Engineering Metrics to Enable Green Machine Learning from Tackling Accuracy-Energy Trade-offs. *Journal of Cleaner Production*, 382, 135334. <https://doi.org/10.1016/j.jclepro.2022.135334>
- Natarajan, H. K., de Paula, D., Dremel, C., & Uebernickel, P. (2022). *A theoretical review on ai affordances for sustainability* (Vol. 13). AMCIS 2022 Proceedings.
- Neff, A. A., Hamel, F., Herz, TPh., Uebernickel, F., Brenner, W., & Vom Brocke, J. (2014). Developing a maturity model for service systems in heavy equipment manufacturing enterprises. *Information & Management*, 51(7), 895–911. <https://doi.org/10.1016/j.im.2014.05.001>
- Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021). Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence*, 2, 100041. <https://doi.org/10.1016/j.caeai.2021.100041>
- Nori, H., Jenkins, S., Koch, P., & Caruana, R. (2019). *InterpretML: A Unified Framework for Machine Learning Interpretability* (arXiv:1909.09223). arXiv.
- Onwuegbuzie, A. J., Dickinson, W. B., Leech, N. L., & Zoran, A. G. (2009). A Qualitative Framework for Collecting and Analyzing Data in Focus Group Research. *International Journal of Qualitative Methods*, 8(3), 1–21. <https://doi.org/10.1177/160940690900800301>
- Pagano, T. P., Loureiro, R. B., Lisboa, F. V. N., Peixoto, R. M., Guimarães, G. A. S., Cruz, G. O. R., Araujo, M. M., Santos, L. L., Cruz, M. A. S., Oliveira, E. L. S., Winkler, I., & Nascimento, E. G. S. (2023). Bias and Unfairness in Machine Learning Models: A Systematic Review on Datasets, Tools, Fairness Metrics, and Identification and Mitigation Methods. *Big Data and Cognitive Computing*, 7(1), 15. <https://doi.org/10.3390/bdcc7010015>
- Papagiannidis, E., Enholm, I. M., Dremel, C., Mikalef, P., & Krogstie, J. (2023). Toward AI Governance: Identifying Best Practices and Potential Barriers and Outcomes. *Information Systems Frontiers*, 25(1), 123–141. <https://doi.org/10.1007/s10796-022-10251-y>
- Pappas, I. O., Mikalef, P., Dwivedi, Y. K., Jaccheri, L., & Krogstie, J. (2023). Responsible Digital Transformation for a Sustainable Society. *Information Systems Frontiers*, 25(3), 945–953. <https://doi.org/10.1007/s10796-023-10406-5>
- European Parliament. (2022). Corporate Sustainability Reporting Directive. *Directive (EU) 2022/2464 of the European Parliament and of the Council*. <https://eur-lex.europa.eu/eli/dir/2022/2464/oj>. Accessed 24 May 2024.
- Patterson, D., Gonzalez, J., Holzle, U., Le, Q., Liang, C., Munguia, L.-M., Rothchild, D., So, D. R., Texier, M., & Dean, J. (2022). The Carbon Footprint of Machine Learning Training Will Plateau. *Then Shrink*. *Computer*, 55(7), 18–28. <https://doi.org/10.1109/MC.2022.3148714>
- Peffer, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems*, 24(3), 45–77. <https://doi.org/10.2753/MIS0742-1222240302>
- Polyviou, A., & Zamani, E. D. (2023). Are we Nearly There Yet? A Desires & Realities Framework for Europe’s AI Strategy. *Information Systems Frontiers*, 25(1), 143–159. <https://doi.org/10.1007/s10796-022-10285-2>
- Prat, N., Comyn-Wattiau, I., & Akoka, J. (2014). Artifact evaluation in information systems design-science research—a holistic view. *Pacific Asia Conference on Information Systems (PACIS)*, 23, 1–16.
- Prat, N., Comyn-Wattiau, I., & Akoka, J. (2015). A Taxonomy of Evaluation Methods for Information Systems Artifacts. *Journal of Management Information Systems*, 32(3), 229–267. <https://doi.org/10.1080/07421222.2015.1099390>
- Radovanović, A., Koningstein, R., Schneider, I., Chen, B., Duarte, A., Roy, B., Xiao, D., Haridasan, M., Hung, P., Care, N., Talukdar, S., Mullen, E., Smith, K., Cottman, M., & Cirne, W. (2023). Carbon-Aware Computing for Datacenters. *IEEE Transactions on Power Systems*, 38(2), 1270–1280. <https://doi.org/10.1109/TPWRS.2022.3173250>
- Reddy, G. T., Reddy, M. P. K., Lakshmana, K., Kaluri, R., Rajput, D. S., Srivastava, G., & Baker, T. (2020). Analysis of Dimensionality Reduction Techniques on Big Data. *IEEE Access*, 8, 54776–54788. <https://doi.org/10.1109/ACCESS.2020.2980942>
- Robinson, O. C. (2014). Sampling in Interview-Based Qualitative Research: A Theoretical and Practical Guide. *Qualitative Research in Psychology*, 11(1), 25–41. <https://doi.org/10.1080/14780887.2013.801543>
- Rohankar, R., Katti, C. P., & Kumar, S. (2015). Comparison of Energy Efficient Data Collection Techniques in Wireless Sensor Network. *Procedia Computer Science*, 57, 146–151. <https://doi.org/10.1016/j.procs.2015.07.399>
- Rohde, F., Wagner, J., Reinhard, P., Petschow, U., Meyer, A., Voß, M., & Mollen, A. (2021). Nachhaltigkeitskriterien für künstliche Intelligenz—Entwicklung eines Kriterien- und Indikatorensets für die Nachhaltigkeitsbewertung von KI-Systemen entlang des Lebenszyklus. *IÖW-Schriftenreihe*, 220, 21.
- Rohde, F., Wagner, J., Meyer, A., Reinhard, P., Voss, M., Petschow, U., & Mollen, A. (2024). Broadening the perspective for sustainable artificial intelligence: Sustainability criteria and indicators for Artificial Intelligence systems. *Current Opinion in Environmental Sustainability*, 66, 101411. <https://doi.org/10.1016/j.cosust.2023.101411>
- Rothe, H., Wessel, L., & Barquet, A. (2020). Accumulating Design Knowledge: A Mechanisms-Based Approach. *Journal of the Association for Information Systems*, 21(3), 771–810. <https://doi.org/10.17705/1jais.00619>
- Russell, S. J., & Norvig, P. (2016). *Artificial intelligence: A modern approach* (3 Global). Pearson.
- Sætra, H. S. (2023). The AI ESG protocol: Evaluating and disclosing the environment, social, and governance implications of artificial

- intelligence capabilities, assets, and activities. *Sustainable Development*, 31(2), 1027–1037. <https://doi.org/10.1002/sd.2438>
- Sarker, S., Chatterjee, S., Xiao, X., & Elbanna, A. (2019). The Socio-technical Axis of Cohesion for the IS Discipline: Its Historical Legacy and its Continued Relevance. *MIS Quarterly*, 43(3), 695–719.
- De Saulles, M. (2020). *Data Liquidity: Data Exchange Platforms as Drivers of Innovation*. <https://doi.org/10.13140/RG.2.2.20887.93603>
- Schneider, J., Seidel, S., Basalla, M., & vom Brocke, J. (2023). Reuse, Reduce, Support: Design Principles for Green Data Mining. *Business & Information Systems Engineering*, 65(1), 65–83. <https://doi.org/10.1007/s12599-022-00780-w>
- Schneider, J., Basalla, M., & Seidel, S. (2019). Principles of Green Data Mining. *Proceedings of the 52nd Hawaii International Conference on System Sciences*. Hawaii International Conference on System Sciences. <https://doi.org/10.24251/HICSS.2019.250>
- Schoormann, T., Strobel, G., Möller, F., Petrik, D., & Zschech, P. (2023). Artificial Intelligence for Sustainability—A Systematic Review of Information Systems Literature. *Communications of the Association for Information Systems*, 52, 199–237. <https://doi.org/10.17705/1CAIS.05209>
- Schulam, P., & Saria, S. (2019). Can you trust this prediction? Auditing pointwise reliability after learning. In *The 22nd international conference on artificial intelligence and statistics* (Vol. 89, pp. 1022–1031). PMLR.
- Schwartz, R., Dodge, J., Smith, N. A., & Etzioni, O. (2020). Green AI. *Communications of the ACM*, 63(12), 54–63. <https://doi.org/10.1145/3381831>
- Shneiderman, B. (2021). Responsible AI: Bridging from ethics to practice. *Communications of the ACM*, 64(8), 32–35. <https://doi.org/10.1145/3445973>
- Singh, V., Singh, A., & Joshi, K. (2022). Fair CRISP-DM: Embedding fairness in machine learning (ML) development life cycle. In *Proceedings of the 55th Hawaii International Conference on System Sciences*.
- Singla, K., Bose, J., & Naik, C. (2018). Analysis of Software Engineering for Agile Machine Learning Projects. *2018 15th IEEE India Council International Conference (INDICON)*, 1–5. <https://doi.org/10.1109/INDICON45594.2018.8987154>
- Smith, G., & Rustagi, I. (2020). *Mitigating Bias in Artificial Intelligence*. Berkeley Haas. <https://haas.berkeley.edu/equity/industry/playbooks/mitigating-bias-in-ai/>. Accessed 2 Jul 2023.
- Sommerville, I. (2018). *Software engineering* (10th ed.). Pearson. C:\Users\meyerhol\Zotero\storage\ID9V7VKJW\Sommerville\_2018\_Software Engineering.pdf
- Sonnenberg, C., & vom Brocke, J. (2012a). Evaluation Patterns for Design Science Research Artefacts. In M. Helfert & B. Donnellan (Eds.), *Practical Aspects of Design Science* (Vol. 286, pp. 71–83). Springer Berlin Heidelberg. [https://doi.org/10.1007/978-3-642-33681-2\\_7](https://doi.org/10.1007/978-3-642-33681-2_7)
- Sonnenberg, C., & vom Brocke, J. (2012b). Evaluations in the Science of the Artificial – Reconsidering the Build-Evaluate Pattern in Design Science Research. In K. Peffers, M. Rothenberger, & B. Kuechler (Eds.), *Design Science Research in Information Systems. Advances in Theory and Practice* (Vol. 7286, pp. 381–397). Springer Berlin Heidelberg. [https://doi.org/10.1007/978-3-642-29863-9\\_28](https://doi.org/10.1007/978-3-642-29863-9_28)
- Spangenberg, J. H. (2002). Environmental space and the prism of sustainability: Frameworks for indicators measuring sustainable development. *Ecological Indicators*, 2(3), 295–309. [https://doi.org/10.1016/S1470-160X\(02\)00065-1](https://doi.org/10.1016/S1470-160X(02)00065-1)
- Sridhar, K., & Jones, G. (2013). The three fundamental criticisms of the Triple Bottom Line approach: An empirical study to link sustainability reports in companies based in the Asia-Pacific region and TBL shortcomings. *Asian Journal of Business Ethics*, 2(1), 91–111. <https://doi.org/10.1007/s13520-012-0019-3>
- Stahl, B., Häckel, B., Leuthe, D., & Ritter, C. (2023). Data or Business First?—Manufacturers’ Transformation Toward Data-driven Business Models. *Schmalenbach Journal of Business Research*, 75(3), 303–343. <https://doi.org/10.1007/s41471-023-00154-2>
- Stolijk, M., Cuijpers, P. L. J., & Lukkien, J. J. (2012). Energy-aware Reprogramming of Sensor Networks Using Incremental Update and Compression. *Procedia Computer Science*, 10, 179–187. <https://doi.org/10.1016/j.procs.2012.06.026>
- Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and Policy Considerations for Deep Learning in NLP. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 3645–3650. <https://doi.org/10.18653/v1/P19-1355>
- Studer, S., Bui, T. B., Drescher, C., Hanuschkin, A., Winkler, L., Peters, S., & Müller, K.-R. (2021). Towards CRISP-ML(Q): A Machine Learning Process Model with Quality Assurance Methodology. *Machine Learning and Knowledge Extraction*, 3(2), 392–413. <https://doi.org/10.3390/make3020020>
- Sundberg, L., & Holmström, J. (2023). Democratizing artificial intelligence: How no-code AI can leverage machine learning operations. *Business Horizons*, S0007681323000502. <https://doi.org/10.1016/j.bushor.2023.04.003>
- Tabladillo, M. (2022). *The Team Data Science Process lifecycle*. <https://learn.microsoft.com/en-us/azure/architecture/data-science-process/lifecycle>. Accessed 23 Sept 2023.
- Taeihagh, A. (2021). Governance of artificial intelligence. *Policy and Society*, 40(2), 137–157. <https://doi.org/10.1080/14494035.2021.1928377>
- Talagala, N. (2021). *The Four Cs Of AI Literacy: Building The Workforce Of The Future*. Forbes. <https://www.forbes.com/sites/nishatalagala/2021/04/04/the-four-cs-of-ai-literacy-building-the-workforce-of-the-future/>. Accessed 19 Jun 2023.
- Tang, Z., Zhang, J., & Zhang, K. (2023). What-is and How-to for Fairness in Machine Learning: A Survey, Reflection, and Perspective. *ACM Computing Surveys*, 55, 1–37. <https://doi.org/10.1145/3597199>
- Thomas, M., Costa, D., & Oliveira, T. (2016). Assessing the role of IT-enabled process virtualization on green IT adoption. *Information Systems Frontiers*, 18(4), 693–710. <https://doi.org/10.1007/s10796-015-9556-3>
- Tomašev, N., Cornebise, J., Hutter, F., Mohamed, S., Picciariello, A., Connelly, B., Belgrave, D. C. M., Ezer, D., Haert, F. C. V. D., Mugisha, F., Abila, G., Arai, H., Almiraat, H., Proskurnia, J., Snyder, K., Otake-Matsuura, M., Othman, M., Glasmachers, T., Wever, W. D., ... Clopath, C. (2020). AI for social good: Unlocking the opportunity for positive impact. *Nature Communications*, 11(1), 2468. <https://doi.org/10.1038/s41467-020-15871-z>
- Tornede, T., Tornede, A., Hanselle, J., Mohr, F., Wever, M., & Hüllermeier, E. (2022). Towards Green Automated Machine Learning: Status Quo and Future Directions. *Journal of Artificial Intelligence Research*. <https://doi.org/10.1613/jair.1.14340>
- Truant, E., Borlatto, E., Crocco, E., & Bhatia, M. (2023). ESG performance and technological change: Current state-of-the-art, development and future directions. *Journal of Cleaner Production*, 429, 139493. <https://doi.org/10.1016/j.jclepro.2023.139493>
- Tsang, A., Frost, T., & Cao, H. (2023). Environmental, Social, and Governance (ESG) disclosure: A literature review. *The British Accounting Review*, 55(1), 101149. <https://doi.org/10.1016/j.bar.2022.101149>
- van Wynsberghe, A. (2021). Sustainable AI: AI for sustainability and the sustainability of AI. *AI and Ethics*, 1(3), 213–218. <https://doi.org/10.1007/s43681-021-00043-6>
- van Giffen, B., & Ludwig, H. (2023). How siemens democratized artificial intelligence. *MIS Quarterly Executive*, 22(1), 3.

- van Noorden, R., & Perkel, J. M. (2023). AI and science: What 1,600 researchers think. *Nature*, 621(7980), 672–675. <https://doi.org/10.1038/d41586-023-02980-0>
- van Giffen, B., Herhausen, D., & Fahse, T. (2022). Overcoming the pitfalls and perils of algorithms: A classification of machine learning biases and mitigation methods. *Journal of Business Research*, 144, 93–106. <https://doi.org/10.1016/j.jbusres.2022.01.076>
- Vassilakopoulou, P., & Hustad, E. (2023). Bridging Digital Divides: A Literature Review and Research Agenda for Information Systems Research. *Information Systems Frontiers*, 25(3), 955–969. <https://doi.org/10.1007/s10796-020-10096-3>
- Vassiliadis, P. (2009). A Survey of Extract Transform Load Technology: *International Journal of Data Warehousing and Mining*, 5(3), 1–27. <https://doi.org/10.4018/jdwm.2009070101>
- Veit, D. J., & Thatcher, J. B. (2023). Digitalization as a problem or solution? Charting the path for research on sustainable information systems. *Journal of Business Economics*, 93(6–7), 1231–1253. <https://doi.org/10.1007/s11573-023-01143-x>
- Verdecchia, R., Sallou, J., & Cruz, L. (2023). A systematic review of Green AI. *Wires Data Mining and Knowledge Discovery*, 13(4), e1507. <https://doi.org/10.1002/widm.1507>
- Vial, G., Cameron, A.-F., Giannelia, T., & Jiang, J. (2023). Managing artificial intelligence projects: Insights from an consulting firm. *Information Systems Journal*, 33(3), 669–691. <https://doi.org/10.1111/isj.12420>
- Vinuesa, R., Azizpour, H., Leite, I., Balaam, M., Dignum, V., Domisch, S., Felländer, A., Langhans, S. D., Tegmark, M., & Fuso Nerini, F. (2020). The role of artificial intelligence in achieving the Sustainable Development Goals. *Nature Communications*, 11(1), 233. <https://doi.org/10.1038/s41467-019-14108-y>
- Visengeriyeva, L., Kammer, A., Bär, I., Knish, A., & Plöd, M. (2023). *MLOps and Model Governance*. <https://ml-ops.org/content/model-governance>. Accessed 4 Jun 2023.
- vom Brocke, J., Hevner, A. R., & Maedche, A. (2020a). *Design science research*. Springer.
- Vössing, M., Kühl, N., Lind, M., & Satzger, G. (2022). Designing Transparency for Effective Human-AI Collaboration. *Information Systems Frontiers*, 24(3), 877–895. <https://doi.org/10.1007/s10796-022-10284-3>
- Wang, C., Han, B., Patel, B., & Rudin, C. (2023). In Pursuit of Interpretable, Fair and Accurate Machine Learning for Criminal Recidivism Prediction. *Journal of Quantitative Criminology*, 39(2), 519–581. <https://doi.org/10.1007/s10940-022-09545-w>
- Wang, C., Wu, Q., Weimer, M., & Zhu, E. (2021). *FLAML: A Fast and Lightweight AutoML Library*. Fourth Conference on Machine Learning and Systems (MLSys 2021). <https://www.microsoft.com/en-us/research/publication/flaml-a-fast-and-lightweight-automl-library/>
- Wanner, J., Heinrich, K., Janiesch, C., & Zschech, P. (2020). How much AI do you require? Decision factors for adopting AI technology. *International Conference on Information Systems*. Forty-First International Conference on Information Systems, India.
- Weerts, H., Dudík, M., Edgar, R., Jalali, A., Lutz, R., & Madaio, M. (2023). Fairlearn: Assessing and Improving Fairness of AI Systems. *Journal of Machine Learning Research*, 24, 1–8.
- Westenberger, J., Schuler, K., & Schlegel, D. (2022). Failure of AI projects: Understanding the critical factors. *Procedia Computer Science*, 196, 69–76. <https://doi.org/10.1016/j.procs.2021.11.074>
- Wirth, R., & Hipp, J. (2000). CRISP-DM: Towards a standard process model for data mining. In *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining* (Vol. 1, pp. 29–39).
- Wormeck, L., Crome, C., Meyer-Hollatz, T., Hinsin, S., & Wassermann, M. E. (2024). Evaluating digital sustainability-oriented innovations: Criteria for the frontend of innovation. In *ECIS 2024 Proceedings* (Vol. 13). European Conference on Information Systems 2024, Cyprus.
- Wu, C.-J., Raghavendra, R., Gupta, U., Acun, B., Ardalani, N., Maeng, K., Chang, G., Behram, F. A., Huang, J., Bai, C., Gschwind, M., Gupta, A., Ott, M., Melnikov, A., Candido, S., Brooks, D., Chauhan, G., Lee, B., Lee, H.-H. S., ... Hazelwood, K. (2022). *Sustainable AI: Environmental Implications, Challenges and Opportunities*. <https://doi.org/10.48550/ARXIV.2111.00364>
- Xiang, L., Luo, J., & Rosenberg, C. (2013). Compressed Data Aggregation: Energy-Efficient and High-Fidelity Data Collection. *IEEE/ACM Transactions on Networking*, 21(6), 1722–1735. <https://doi.org/10.1109/TNET.2012.2229716>
- Xu, T. (2022). *These simple changes can make AI research much more energy efficient*. MIT Technology Review. <https://www.technologyreview.com/2022/07/06/1055458/ai-research-emissions-energy-efficient/>
- Yarally, T., Cruz, L., Feitosa, D., Sallou, J., & van Deursen, A. (2023). *Uncovering Energy-Efficient Practices in Deep Learning Training: Preliminary Steps Towards Green AI*. <https://doi.org/10.48550/ARXIV.2303.13972>
- Young, D. S., & Casey, E. A. (2018). An Examination of the Sufficiency of Small Qualitative Samples. *Social Work Research*. <https://doi.org/10.1093/swr/svy026>
- Yu, J. (2014). *Big Data vs. Relevant Data: Intelligence That Matters*. HuffPost. [https://www.huffpost.com/entry/big-data-vs-relevant-data\\_b\\_5022792](https://www.huffpost.com/entry/big-data-vs-relevant-data_b_5022792). Accessed 17 Sept 2023.
- Yurrita, M., Murray-Rust, D., Balayn, A., & Bozzon, A. (2022). Towards a multi-stakeholder value-based assessment framework for algorithmic systems. *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, 535–563. <https://doi.org/10.1145/3531146.3533118>
- Zacharias, J., von Zahn, M., Chen, J., & Hinz, O. (2022). Designing a feature selection method based on explainable artificial intelligence. *Electronic Markets*, 32(4), 2159–2184. <https://doi.org/10.1007/s12525-022-00608-1>
- Zhang, B. H., Lemoine, B., & Mitchell, M. (2018). Mitigating Unwanted Biases with Adversarial Learning. *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*, 335–340. <https://doi.org/10.1145/3278721.3278779>
- Zhu, S., Ota, K., & Dong, M. (2022). Green AI for IIoT: Energy Efficient Intelligent Edge Computing for Industrial Internet of Things. *IEEE Transactions on Green Communications and Networking*, 6(1), 79–88. <https://doi.org/10.1109/TGCN.2021.3100622>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Daniel Leuthe** is a researcher at the FIM Research Center for Information Management of the University of Applied Sciences Augsburg and at the Branch Business & Information Systems Engineering of the Fraunhofer FIT. His research focuses on the implications of data analytics and machine learning for existing products and services. He furthermore works as a Machine Learning Engineer and has published articles in various outlets such as Computers & Security, Energy and Buildings, and Schmalenbach Journal of Business Research.

**Tim Meyer-Hollatz** is a researcher at the FIM Research Center for Information Management and at the Branch Business & Information Systems Engineering of the Fraunhofer FIT. His research is positioned at the intersection of sustainability and digitalization, with a particular focus on the convergence of sustainability and AI/ML. This includes exploring both the AI for Sustainability and the Sustainability of AI.

**Tobias Plank** is currently pursuing his Master's degree in Finance & Information Management at the Technical University of Munich and the University of Bayreuth. He is a research assistant at the FIM Research Center for Information Management and the Branch Business & Information Systems Engineering of the Fraunhofer FIT. In his studies and research, Mr. Plank works on the intersection of applied advanced technologies (i.e., artificial intelligence and blockchain) and sustainability. Throughout his research, his goal is to equally drive the sustainable use and use for sustainable development of these technologies by various stakeholders.

**Anja Senkmüller** is a Master's student in Finance & Information Management at the Technical University of Munich and the University of Bayreuth. Her academic pursuits and research focus on the convergence of applied advanced technologies, such as artificial intelligence and machine learning, with sustainability. Ms. Senkmüller's research aims to promote both the sustainable application of these technologies and their role in advancing sustainable development by diverse stakeholders.