The environmental impact of AI: a framework for energy consumption in machine learning services

International Journal of Quality and Service Sciences

Michael Möhring
School of Computer Science − HHZ, Reutlingen University, Reutlingen, Germany

Barbara Keller
Faculty of Business, DHBW, Stuttgart, Germany, and

Laura Di Pietro
Department of Business Studies, Roma Tre University, Rome, Italy

Received 4 December 2025
Revised 19 March 2026
11 June 2026
22 June 2026
Accepted 23 June 2026

Abstract

Purpose – This study aims to develop a framework to analyze the energy consumption of machine learning (ML) services. It enables a process-oriented understanding of the environmental impacts and identifies actionable levers for their mitigation. Indeed, ML services are increasingly integrated into organizational processes generating rising energy demand and environmental impact, highlighting the need for solutions to assess these issues.

Design/methodology/approach – Using a design sience research approach, first an initial framework drawing on Cross Industry Standard Process for Data Mining (CRISP-DM) was developed through a targeted literature review, and observations of different recommender systems. Then, a two-step assessment (pre-assessment and expert evaluation involving academic and practitioner stakeholders) was conducted. Insights from these evaluations enriched the final framework proposal.

Findings – This research reveals that along the CRISP-DM process different factors impact energy consumption and must be considered. Three different phases and related energy relevant aspects were named. The evaluation of the framework as an artifact underlines its statement and value for research and practice.

Research limitations/implications – The framework provides a holistic overview of relevant factors that must be considered in the context of energy consumption of ML-based artificial intelligence (AI) services. Limitations can be found in the neglected model accuracy, the selected sources and the applied research approach itself.

Practical implications – The developed framework can be used by practitioners to understand and evaluate important factors of the energy demand that should be considered before and when building a ML-based AI service in practice.

Originality/value – This paper develops and presents a framework for the environmental impact of AI, focusing on ML approaches. The framework provides a holistic view of energy consumption in ML services based on the previous literature and recommender systems. In addition, it is process-based and enables its usage at different milestones within a project.

Keywords AI, Machine learning, Environmental impact, Energy consumption, Framework, Sustainability, Information systems

Paper type Research paper

International Journal of Quality and Service Sciences
Emerald Publishing Limited
1756-669X
DOI 10.1108/IJQSS-12-2025-0340

1. Introduction

Artificial intelligence (AI) represents a transformative solution that is reshaping services throughout its potential to enhance decision-making and operational efficiency (Mustafa *et al.*, 2025; Di Vaio *et al.*, 2020; Oppioli *et al.*, 2023; Nishant *et al.*, 2020; Singh *et al.*, 2024) and simulate human intelligence and autonomously perform complex tasks (Hoehndorf and Queralt-Rosinach, 2017; Moloi and Obeid, 2024). In particular, the AI integration into service ecosystems is completely reconfiguring service innovation, acting as a key driver of economic growth over the coming years (Akter *et al.*, 2023; Huang and Rust, 2018).

This transformation unfolds within the broader context of the *twin transition*, described by the European Union (2022) as the simultaneous pursuit, digital and ecological transformations. Tabares *et al.* (2025, p. 4) attempted to define twin transition as “two parallel and mutually reinforcing digital and green transitions, which amplify each other leading to sustainable competitiveness for firms.” Accordingly, institutional, economic and social actors are striving to advance along this dual and integrated trajectory.

However, emerging inconsistencies warrant closer examination. *Digital transition* presents a dual nature: while it can act as a catalyst for sustainability by enabling the implementation of innovative solutions, it may also become a source of environmental pressure when generating uncontrolled impacts (Dörr, 2022).

AI and conjunct approaches like *machine learning* (ML) are driving unprecedented innovation and efficiency across sectors and organizations. AI-based services, such as automatic forecast services (IBM, 2024) or generative AI solutions (IBM, 2026; Sigala *et al.*, 2024), are gaining more importance for organizations every day (Yee *et al.*, 2024). Applications like ChatGPT, Gemini or Copilot enable users to generate text and retrieve information in natural language (IBM, 2026). Midjourney generates pictures based on keywords (prompts) transforming business and services (Sigala *et al.*, 2024). These systems provide intuitive and comfortable handling and subsequently, fostering their usage in practice through their intrinsic value. It is not surprising that most sectors are affected, and an almost endless list can be enumerated starting from marketing (Davenport *et al.*, 2020; Kumar *et al.*, 2026) and production (Fragapane *et al.*, 2022) up to medicine (Briganti and Le Moine, 2020), education (Holmes and Tuomi, 2022) and services more broadly. Huang and Rust (2021, p. 31) argued that AI can also support customer engagement across multiple service levels and along the various stages of the service process. Similarly, in the service quality field, traditional process efficiency approaches (i.e. Lean Six Sigma, Kaizen) based on continuous improvement methodologies are enhanced through the integration of AI-powered approaches including predictive modeling (Bukhari and Akhtar, 2024) and data analytics (Carneiro *et al.*, 2026). This evolution has contributed to the development of the *Quality 4.0 paradigm* (Broday, 2022) and facilitated the emergence of smart continuous improvement practices (Hossain and Dhanekula, 2023). These are only examples highlighting the importance of AI and adjunct ML-based approaches within the service domain and for AI as a service itself.

Despite the well-established economic benefits of AI in terms of efficiency and competitiveness, its environmental implications remain largely underexplored (Wang *et al.*, 2024; Keller *et al.*, 2024). This reveals a significant research gap that appears particularly critical in the context of the twin transition, where the ecological transition should also be taken into considerations. For AI to be used effectively and in line with this concept − generating economic benefits while also meeting sustainability goals − it is necessary to establish a clear, evidence-based understanding of how this can be achieved. Thus, the adverse environmental impacts associated with AI services and overlooked ecological costs are particularly interesting. Yet, their rapid expansion entails significant ecological costs and risks (i.e. energy consumption, electronic waste generation, water consumption, carbon footprint, etc.) that remain largely overlooked (Zhuk, 2023; Kaack *et al.*, 2022; Keller *et al.*, 2024). Most existing studies predominantly focus on the positive dimensions and neglect the environmental trade-offs associated with widespread adoption.

To address this gap, the present study adopts a design science research method (Hevner *et al.*, 2004; Peffers *et al.*, 2007) to develop a framework aimed to support a more comprehensive evaluation and mitigation of the environmental impacts associated with AI. The proposed framework is intended to support the evaluation of AI applications at various stages of a project, while ensuring alignment with the twin transition requirements. Accordingly, the paper investigates the following research question:

RQ1. What aspects should be considered in a framework for evaluating the environmental impact of ML-based AI services based on energy consumption?

The proposed framework provides two relevant implications for service research by contributing to the understanding of AI as a service and providing insights for conjunct and subsequential services along typical projects. The paper is structured as follows. First, the paper outlines the relevant background, including the role of AI in the twin transition and presenting key considerations on ML energy consumption. The research focus lies on ML as a key segment of AI, drawing on insights from prior literature and leveraging established tools such as HollerithEnergyML (Zanger *et al.*, 2024), Green Algorithm Calculator (Lannelongue *et al.*, 2021) and ML CO_2 Calculator (Lacoste *et al.*, 2019). After, the research method is described, and the findings are presented. Following a standard process in ML projects (CRISP-DM), the results are linked to the associated process steps. The framework is then demonstrated within a service management conference (QUIS19) and evaluated afterward with leading experts. Finally, the paper concludes with a discussion and implications.

2. Background

2.1 The ambivalent *role* of AI in the twin transition

As recognized by several authors (e.g. Hammerschmidt *et al.*, 2025; Tabares *et al.*, 2025; Christmann *et al.*, 2024), the twin transition (or twin transformation) consist in the integrated and synergetic implementation of digital and ecological (and/or sustainability) transitions. Some authors note that it can be viewed as the adoption of digital technologies, including AI, to achieve sustainable and circular economy outcomes (Bianchini *et al.*, 2023; Tabares *et al.*, 2025). Similarly, Christmann *et al.* (2024, p. 490) posited that "digital transformation can potentially enable insights about sustainability transformation effects, while sustainability transformation may guide the design of digital transformation solutions, realizing value in new ways." Nowadays, the twin transition is a focal point of the service domain, where services are grounded in digitally enabled resource integration and value co-creation processes (Barile *et al.*, 2020) but, at the same time, they are expected to contribute to sustainability-oriented outcomes (Guandalini, 2022). From any analytical standpoint, the twin transition evidences a synergistic interconnection between the principles and enabling tools that underpin the processes and strategic pathways of digitalization and sustainability. Following, the main positive and negative aspects are summarized.

Positive aspects: The use of analytical services such as ML services can simultaneously generate multiple benefits and challenges for organizations (Keller *et al.*, 2024; Wang *et al.*, 2024). For example, ML has become a key concept of Industry 4.0 through enabling real-time data analysis, predictive maintenance and automated decision-making (Ayvaz and Alpay, 2021; Möhring *et al.*, 2022). These capabilities have accelerated digital transformation. They improve efficiency, challenge traditional operational models and revolutionize quality management

(Malikah *et al.*, 2025; Zonnenshain and Kenett, 2020). In addition, ML services facilitate developing strategies and data-driven approaches to reduce greenhouse gas (GHG) emissions (Kaack *et al.*, 2022). In line with this argument, several authors highlight a direct relationship between the adoption of AI and ML as strategic levers to foster the implementation of circular economy models (i.e. Montes-Pineda and Garrido-Yserte, 2024) or specific circular practices (i.e. Mohammadian *et al.*, 2025; Tutore *et al.*, 2024; Ronaghi, 2023; Noman *et al.*, 2022; Jose *et al.*, 2020; Pagoropoulos *et al.*, 2017). This highlights the potential of these approaches to contribute to both economic and ecological values in organizations. Wang *et al.* (2024) argued that AI exerts a relevant influence in advancing the energy transition, followed by its effect in reducing the ecological footprint and then, in lowering the carbon emissions. Furthermore, considering that the adoption of environmental sustainability practices can generate positive spillover effects in terms of social benefits (Zhu and Dolnicar, 2025), it can be argued that through AI-based services the digital transition provides tools capable of advancing environmental and social outcomes at the same time.

Negative aspects: Besides these positive aspects, ML services can also have a negative impact on the environment (Kaack *et al.*, 2022). For instance, even the use of a trained model for computer vision or Natural Language Processing (NLP) is highly energy-intensive (Desislavov *et al.*, 2023). These negative and unfavorable environmental impacts and consequences are widely ignored so far (Keller *et al.*, 2024). For instance, the training and use of ML services have a high level of energy consumption (Kaack *et al.*, 2022). Therefore, the substantial energy demand of ML models poses significant challenges to environmental sustainability, directly affecting SDG 7 (Affordable and clean energy), SDG 12 (Responsible consumption and production) and SDG 13 (Climate action). The widespread adoption of AI raises concerns about the carbon footprint of ML models and their contribution to climate change (Tamburrini, 2022). This is aligned with the environmental digital responsibility as a core dimension of corporate digital responsibility, which addresses the ecological impact of digital technologies by promoting responsible recycling, to extend equipment life cycles, and to adopt ethical energy consumption practices (Oduro *et al.*, 2023). These are critical factors that must be taken into consideration when using and assessing the environmental impact of AI-based services.

2.2 Energy consumption in machine learning

Through ML, software systems are able to learn during a training phase without explicit and detailed programming (Samuel, 1959; Awad and Khanna, 2015). ML services are software-based services that enable end users to solve problems using trained ML algorithms (Philipp *et al.*, 2020; Yao *et al.*, 2017). Therefore, ML services provide tools that can be used immediately without considerations about e.g. detailed computing resources, data complexity and handling, model training or deployment (Philipp *et al.*, 2020; Yao *et al.*, 2017). Typical scenarios in service, for example, include the prediction of repeat purchases (Hwang *et al.*, 2020) or the implementation of personal dynamic pricing (Ban and Keskin, 2021). As a general approach, service automation can be mentioned, where customer value is generated through improved support (Heinonen *et al.*, 2020; Broekens *et al.*, 2009). As another example, within the more specific area of service research, topic modeling through ML can support the research process (Antons and Breidbach, 2018).

The use and the quality (in terms of improved accuracy) of ML models and subsequently ML services are increasing steadily, generating a high and rising demand for computational power and consequently energy (García-Martín *et al.*, 2018; Eilam *et al.*, 2023). OpenAI publicized that their computation effort was increasing exponentially and doubling every 3.4 months (OpenAI, 2018; García-Martín *et al.*, 2018). The importance of measuring and quantifying this demand is obvious. Thus, it becomes a necessary priority to develop tools that help quantify the carbon footprint of ML models and facilitate the transition to a more sustainable AI infrastructure (Dhar, 2020). Prior research suggests that the carbon impact of ML models should be reported through a Carbon Impact Statement (Henderson *et al.*, 2020). This stands in line with the twin transition theory described above. The authors propose a framework assessing the ML impact as an experimental tracker. Fischer and colleagues (2023) focused in their research on the importance of energy efficiency of ML models. They indicate that ML energy labels can be valuable for comparing ML efficiency (Fischer *et al.*, 2022). However, the introduction of such labels is challenging, because the measurement and estimation of energy consumption here, is not easy (Goel *et al.*, 2012; García-Martín *et al.*, 2018). Insights about how to implement this are provided by García-Martín *et al.* (2018, 2019), who defined in their work the basics of energy consumption and recommendations to measure it. Generally, these proposed measurements ground on power meters or direct measures at the hardware motherboard (García-Martín *et al.*, 2018). But, as commented by the authors themselves, these approaches lack on fine granularity. Energy consumption can be measured empirically or analytically through simulation (García-Martín *et al.*, 2018). The CO_2 emissions resulting from this energy consumption can then be calculated (Lacoste *et al.*, 2019; Lannelongue *et al.*, 2021). That means, for example, the carbon emission can be calculated by the energy consumption itself (e.g. in kWh) multiplied by the carbon intensity (e.g. CO_2e/kWh) of the energy grid (Henderson *et al.*, 2020). Also, the region has an impact (e.g. Canada and Finland), where low carbon intensity results in a lower carbon emission (Henderson *et al.*, 2020). In this way, the environmental compatibility of these services can be evaluated. Moreover, further precise measures are provided via performance monitor counters (García-Martín *et al.*, 2018). Also, using concrete options based on the code level such as Python libraries or software tools (e.g. Code Carbon, Nvidia Management Library, Intel RAPL or Power Gadget), enable measurements and estimations of energy consumption (Fischer *et al.*, 2022; Zanger *et al.*, 2024; Hasan *et al.*, 2025). Moreover, the energy measurement method depends on the data. It is important to distinguish whether the model is exclusively trained with a fixed and defined historical data set (offline) or online with continuous data (García-Martín *et al.*, 2018). The complexity can also increase by the usage of e.g. virtualization, docker or cloud environments that might cause issues for concrete energy measurement (Fischer *et al.*, 2022).

Prior research findings in the context of ML consumption are limited. Islam *et al.* (2023) investigated the energy consumption of frequently used algorithms. However, energy consumption is not solely driven by the algorithms themselves. Other authors already used different measurement approaches to assess the energy consumption of ML models (Strubell *et al.*, 2019; Mavromatis *et al.*, 2024). They focused on specific tasks (NLP, image classification) and investigated the energy consumption in these cases (Strubell *et al.*, 2019; Mavromatis *et al.*, 2024). They also provide considerations regarding hardware and infrastructure. Other authors like Schwartz *et al.* (2020) have discussed the topic of green AI more through the lens of general resource consumption and efficiency instead of focusing on energy in detail. However, to the best of our knowledge, no scientific findings provide a holistic and process-based view considering all necessary phases of an analytical services project exist. This research tries to fill this gap by developing a framework with more general applicability and contributes to prior research by enlarging the insights and recommendations.

3. Research method

To respond to the recognized gap and subsequently to develop a framework, a research project according to the design science research method was conducted (Hevner *et al.*, 2004; Peffers *et al.*, 2007).

3.1 Usage of design science research

Choice of method: In general, the underlying idea here is that a given problem can be understood, analyzed and solved through the development of an artifact and its specific application (Hevner *et al.*, 2004; Peffers *et al.*, 2007). In this context, artifacts are broadly defined and can be a construct, model or method, for example Hevner *et al.* (2004). According to Hevner *et al.* (2004), design science research is useful for creating and evaluating IT artifacts (e.g. frameworks) enabling organizations to understand aspects and to incorporate relevant information. In literature, the applicability of this research method in service is highlighted, too (e.g. Teixeira *et al.*, 2019). The objective was to develop a framework to facilitate the understanding of the environmental impact of AI services, particularly ML services, based on their energy consumption. This method relies on a problem-solving process that aims to develop an artifact and thus, to find a solution to a problem through evaluation steps (Hevner *et al.*, 2004). For this reason, this approach is particularly suitable in this case, as it combines behavioral and design sciences (Hevner *et al.*, 2004).

Implementation of the method: For the conducting of the research, the design science research process model proposed by Peffers *et al.* (2007) was applied. This research agenda consists of six sequential and interconnected steps (that means, between the steps, an iterative process is possible, too):

(1) problem identification and motivation;
(2) objectives of a solution;
(3) design and development;
(4) demonstration;
(5) evaluation; and
(6) communication.

In accordance with the methodological guidelines, the steps were implemented as follows.

Step (1): Problem identification and motivation is elaborated within the Introduction and the Background paragraph. The problem is highlighted and the motivation for the research is presented. In line with the defined research question, the main objective of the solution (Step 2) is that the developed framework can explain the environmental impact of ML services based on energy consumption. Furthermore, to ensure the usability of the artifact, the framework should be understandable and useful (King and He, 2006; Davis, 1985). Step (3) is described in paragraph 4. A literature review was conducted and observations regarding the energy consumption of different AI/ML algorithms based on recommender systems were integrated. These observations with sample data can show how energy consumption is affected by different input variables like sample size and provide additional information for the framework. Step (4) and Step (5) are discussed within paragraph 5, where the demonstration process at a leading service management conference presentation as well as the evaluation process are described, and the framework is adjusted. Step (6), which refers to the communication, is done through the paper itself. This communication approach is used by Peffers and co-authors (2007), too.

3.2 Literature review

Aiming to prepare the artifacts design and development, a literature review based on the central aspects outlined by Webster and Watson (2002) was conducted. Relevant scientifical contributions were collected, selected and analyzed. The literature search was conducted using leading scientific databases and search engines like Google Scholar, ScienceDirect and considered publications up to the year 2025. Here, citations, relevance to the topic, and/or mentions within the papers were taken into consideration. Supplementary paper describing theoretical energy optimizations of specific variants of ML models and papers general discussing algorithm complexity as well as efficiency were omitted. The literature review revealed that most existing studies use ML for energy consumption prediction, failing to cover the energy demand generated by the ML processes. Overall, several factors impacting the energy consumption of AI systems exist, particularly in terms of ML.

3.3 Expert evaluation design and sample

Expert sample: The evaluation of the developed framework was carried out with experts coming from practice and research as well. This procedure is an important part of the design science research method (Hevner *et al.*, 2004; Peffers *et al.*, 2007). It was conducted in the third quarter of the year 2025. In total, six different experts from different domains participated and were involved in the evaluation. Three experts are employed in leading European industry enterprises. All the organizations come from different sectors (1) fashion manufacturer, (2) agricultural machinery manufacturer and (3) a database vendor and operate internationally. The other three experts are AI and ML experts from different research institutions in Germany. All experts have related knowledge and experience in ML and AI (on average 10.7 years) and related experience in the resource consumption of these models. The evaluation was conducted in the German language.

Procedure and question: The participants were invited formally (e.g. email, telephone) to participate in the evaluation. After agreeing to contribute, a Microsoft Teams call was organized. During the session, the experts were asked to evaluate the framework by answering open as well as closed-ended questions. All questions are based on basic recommendations of the relevant literature (e.g. Davis, 1985; King and He, 2006). The statements adapted the TAM dimensions "Perceived usefulness" and "Perceived ease of use" (Davis, 1985) in a context-specific manner. According to Davis (1989) "Perceived usefulness" can be described as: "[…] *user believes in the existence of a positive use-performance relationship*" (Davis, 1985; p. 320). This dimension was assessed through the questions (1) and (3). "Perceived ease of use" or "[…] *the degree to which a person believes that using a particular system would be free of effort.* […]" (Davis, 1985; p. 320), was measured through question (2).

At first, closed-questions were asked to gather insights about whether (1) the framework can explain the impact of ML services based on energy consumption ("*The framework can explain the environmental impact of Machine Learning services based on energy consumption*"; five-point Likert Scale: 1 = completely agree to 5 = completely disagree), (2) is easy to understand ("*The framework is understandable*"; five-point Likert Scale: 1 = completely agree to 5 = completely disagree) and (3) is useful in terms of energy consumption ("*The framework is useful for gaining an overview of the environmental impact of Machine Learning services based on energy consumption*"; five-point Likert Scale: 1 = completely agree to 5 = completely disagree). Furthermore, an open question was asked to gain insights about missing details of the framework. The participants had to provide their opinion about missing aspects ("*What is missing in the framework?*").

4. Results

Given the research method chosen, the artifact itself constitutes the central finding of the research. The steps required for this are described in detail in the following subsections.

4.1 Design and development of the artifact

4.1.1 Design of the framework. Due to the variety of ML applications, some sort of standardization had to be carried out to ensure relevance and informative value of the findings. Therefore, the Cross Industry Standard Process for Data Mining also known as CRISP-DM (Wirth and Hipp, 2000) as a standard for analytical projects (Chapman *et al.*, 2000; Schröer *et al.*, 2021) was selected. Its mutually dependent and interrelated phases, business understanding, data understanding, data preparation, modeling, evaluation and deployment, were considered with regard to their respective energy consumption. In analytical projects, training and usage are an integral part of modeling, evaluation and deployment (Chapman *et al.*, 2000). In general, resources are required to train the model and, for its subsequent use (Kaack *et al.*, 2022). Additionally, analytical projects typically involve preliminary stages prior to model training (e.g. business understanding, data understanding and data preparation) as shown by Chapman *et al.* (2000) and this pre-training phase also influences the energy consumption level of AI services. The phases of the CRISP-DM are further assigned into pre-training phase, training phase as well as usage phase.

The analyzed literature in line with the literature review and aligned with identified phases is shown in the following table.

The most relevant research papers in line with the study scope are listed in Table 1.

4.1.2 Development of the framework. In the next step of this study, the framework was built. It merges the findings of the literature review with observations of the established recommender systems [ML CO_2 Calculator (Lacoste *et al.*, 2019), Green Algorithms Calculator (Lannelongue *et al.*, 2021) and HollerithEnergyML Recommender (Zanger *et al.*, 2024)] used and described within the different phases. The first version of the framework is depicted in Figure 1 and described in the following.

4.1.2.1 Pre-training phase. Before training a model, scope and context must be understood. In the CRISP-DM process, this step is known as *business understanding* (Chapman *et al.*, 2000). If this step has not been conducted seriously, the project must be stopped and restarted several times. Consequently, this may imply a higher energy consumption, because preliminary analyses must be conducted several times. Afterwards, the phase *data understanding* (Chapman *et al.*, 2000) must be passed. Relevant and usable data (e.g. data ownership, quality, actuality) for the ML model must be explored and collected to reach a defined goal. If it has been successfully checked that the data is suitable for the intended purpose, the next step *data preparation* (Chapman *et al.*, 2000) can be started. Here, the data must be transformed into the correct form before it can be processed to the training. In every step, energy is demanded and consumed. With every step running several times, the energy consumption level increases simultaneously. Additionally, the used hardware (e.g. personal computer, cloud environment, self-hosted server) might increase the energy consumption, too (Lannelongue *et al.*, 2021; Lacoste *et al.*, 2019; Patterson *et al.*, 2021).

4.1.2.2 Training phase. The training phase consists of the following identified influencing factors: algorithms, data and IT infrastructure. These are explained in the following. In general, the training phase includes model development and training (Kaack *et al.*, 2022).

Training data: Observations of a recommender system and previous literature findings indicate that the data amount as well as the variety of structures due to different variables (e.g. categorical vs numerical variables) can influence the energy consumption (e.g. Zanger *et al.*, 2024; García-Martín *et al.*, 2018). The greater the amount of data and the more complex the structure of the data set used for the training, the higher may be the energy demand (Patterson *et al.*, 2021).

Table 1. Reviewed literature regarding energy and resource usage and related influencing factors

		Phases of machine learning services			
Author	Year	*Pre-training Phase*	*Training phase*	*Usage phase*	Short description
Desislavov *et al.*	2023			X	Research about energy consumption and computations trends of deep learning models in the domain of computer vision and NLP
Fischer *et al.*	2023		X	X	Development of a framework for assessing energy efficiency in ML experiments / tasks
García-Martín *et al.*	2018		X		Paper describes how to measure energy consumption in different machine learning scenarios
García-Martín *et al.*	2019		X	X	A review of different possibilities to measure energy consumption in machine learning
Han *et al.*	2015		X	X	Research describes how energy can be saved due to approaches like pruning, quantization and Huffman coding
Henderson *et al.*	2020		X	X	A framework for energy and carbon footprints of machine learning was developed
Islam *et al.*	2023		X		Research on the energy consumption of different machine learning algorithms
Kaack *et al.*	2022		X	X	Development of a framework for understanding the effects of machine learning on GHG emissions
Lacoste *et al.*	2019	X*	X	X*	A machine learning emission calculator is developed with focus on training and hardware
Lannelongue *et al.*	2021	X*	X	X*	A calculator for carbon emission is developed based on e.g. hardware, location and algorithm running aspects
Mavromatis	2024		X	X	Examination of model architectures etc. in training and inference
Patterson *et al.*	2021	X*	X	X*	Research about the energy usage and carbon emission of large language models
Strubell *et al.*	2019		X		Research about the energy and environmental costs of NLP training
Wang *et al.*	2023		X	X	Energy aspects of language model finetuning, pre-training and inference, with the example of Google BERT
Yang *et al.*	2017			X	Research paper on energy-aware pruning to reduce energy consumption in convolutional neural networks
Zanger *et al.*	2024		X		Development of a recommender system to compare different machine learning classifiers during training

Note(s): *Research does not explicitly focus on the energy consumption in this phase. However, they investigate the energy consumption of the hardware which is used in all phases
Source(s): Created by the authors

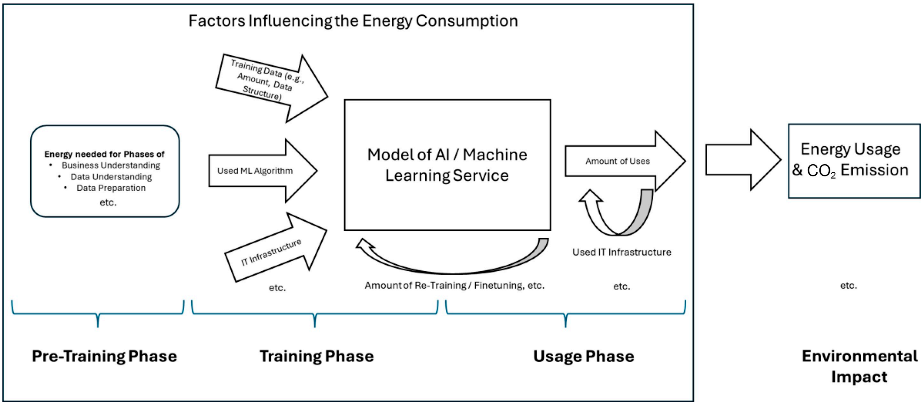

Figure 1. First version of the framework of environmental impact of machine learning services based on findings in literature and recommender system observations (demonstrated at QUIS19)
Source: Created by the authors

Algorithms: For every specific problem, various possibilities for a suitable ML algorithm exist. However, in advance it is not clear, which one represents the most favorable option ("No free lunch theorem"; Wolpert and Macready, 1997). In terms of energy consumption algorithms differ, too (Zanger *et al.*, 2024; Islam *et al.*, 2023; García-Martín *et al.*, 2018; Patterson *et al.*, 2021; Kaack *et al.*, 2022). For example, deep convolutional neuronal networks (CNN) have a high energy demand (Yang *et al.*, 2017; Nvidia, 2015). A glance at observations from the HollerithEnergyML Recommender (Zanger *et al.*, 2024) shows clearly, that in terms of classifying data, algorithms like the random forest algorithm are more energy inefficient than the K-Nearest-Neighbor algorithm or the Decision Tree algorithm (see Figure 2). Of course, larger models might enhance accuracy and consequently the results, but simultaneously the energy consumption rises (Patterson *et al.*, 2021). In addition, the model's training time (Lacoste *et al.*, 2019) also affects the energy demand. The longer the duration; the higher the energy consumption level. Algorithmic improvements supporting to manage the number of parameters and hyperparameter enhancements can save energy resources during the training phase (Fischer *et al.*, 2022; Desislavov *et al.*, 2023; Lacoste *et al.*, 2019). Energy leaderboards can be used to compare different ML experiments in terms of for example accuracy and energy efficiency (Henderson *et al.*, 2020). In conclusion, it can be stated that the design of the ML algorithms must be planned properly in advance to avoid unnecessary ML experiments and consequently energy waste (Lacoste *et al.*, 2019).

Furthermore, it must be considered that the energy consumption depends not only on the computation itself but also on the number of memory accesses required by the implemented ML algorithms (Yang *et al.*, 2017). Hence, sometimes small models (e.g. CNN with small size) can have higher energy consumption due to the amount of memory accesses (Yang *et al.*, 2017). Algorithm optimization possibilities like pruning, distillation, compression, quantization and compact models can reduce complexity and related energy consumption (García-Martín *et al.*, 2019; Yang *et al.*, 2017; Patterson *et al.*, 2021; Han *et al.*, 2015; Iandola *et al.*, 2016). Also, the complexity and different implementation (e.g. number of parameters, epochs) and execution possibilities can be relevant in terms of energy consumption (Fischer *et al.*, 2022). Here, different libraries and software environments can be used for improvements. Multiply accumulate operations can also be a good predictor of

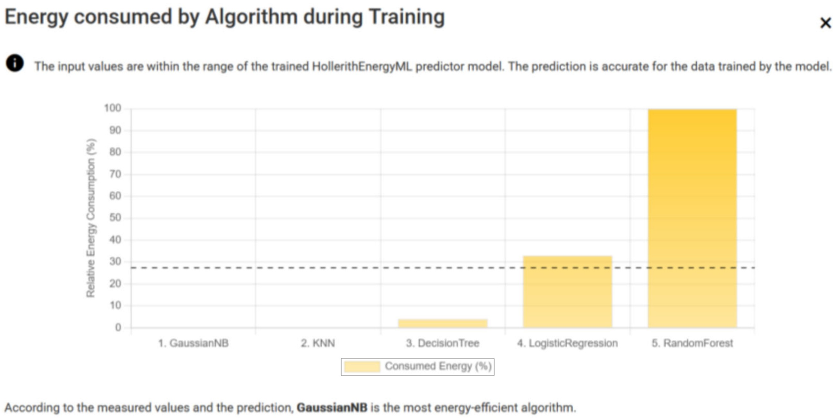

Figure 2. Example energy consumption for 20 numerical, 15 categorical and 1,500 data rows with HollerithEnergyML (ki-lab-region-stuttgart.de)

the complexity and the energy consumption of the implementation of the algorithm (Mavromatis *et al.*, 2024). Furthermore, floating point operations (FLOPs) could be a KPI for the computational requirements of ML models (Kaack *et al.*, 2022). Thus, an increasing number of FLOPs leads to a higher energy consumption in general, too (Kaack *et al.*, 2022).

IT-infrastructure: Another critical aspect regarding the energy and resource consumption is the existing IT infrastructure (Lacoste *et al.*, 2019; Patterson *et al.*, 2021; Fischer *et al.*, 2022; Strubell *et al.*, 2019; Lannelongue *et al.*, 2021). For instance, the use of CPU, GPU, Memory, Disk etc. is important and can have an impact (Henderson *et al.*, 2020; Mavromatis *et al.*, 2024; Lacoste *et al.*, 2019). Also, the usage of special processing units for neural networks (TPUs) can save energy (Patterson *et al.*, 2021). In line with the research of Lacoste *et al.* (2019) and observations from the related ML CO_2 Calculator for GPUs (mlco2.github.io), differences in the energy consumption level and its related CO_2 emission based on the hardware type and the hardware provider (e.g. AWS, Google, private infrastructure and region) become visible.

This effect is highlighted in Figure 3. It shows an example of information that can be obtained from the ML CO_2 Calculator Recommender System. The figure compares the energy consumption of different hardware ("RTX 3080" vs "RTX 4090") in the same environment ("AWS" − Amazon Web Services; EU Frankfurt region environment) needed to conduct a 100-h model training. On the "RTX 4090" 30 kWH were needed and 18.3 kg CO_2 have been emitted. In comparison, the energy consumption and emission of the "RTX 3090" hardware has been even higher (35 kWH and 21.35 kg CO_2). The hardware system can differ in terms of energy efficiency and the data center with regard to the natural energy mix and the server replacement strategy, too (Kaack *et al.*, 2022; Lacoste *et al.*, 2019; Patterson *et al.*, 2021; Strubell *et al.*, 2019). Furthermore, aspects such as data centers' power consumption efficiency are important to consider and to integrate into the used recommender system (Lacoste *et al.*, 2019). This is because energy is also needed for cooling, power conversion and other tasks (Lacoste *et al.*, 2019). In general, an energy efficient hardware system must be chosen (Lacoste *et al.*, 2019). Using more CPU or GPU cores to reduce required time may also have a negative impact on energy consumption and the carbon footprint, too (Lannelongue *et al.*, 2021). Besides using the ML CO_2 Calculator Recommender System

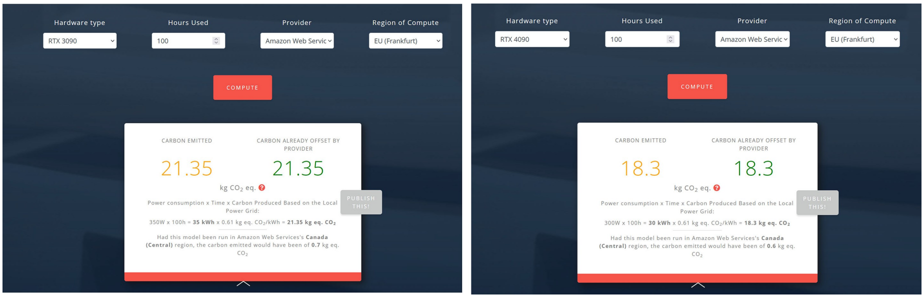

Figure 3. Example energy consumption for “RTX3090” and “RTX40900,” Cloud Provider with ML CO_2 Calculator (mlco2.github.io)

(mlco2.github.io) here the Green Algorithms Calculator can also be used (Lannelongue *et al.*, 2021) in the same way to get more insights.

Full-training or finetuning/pre-trained models: In particular, the training of large language models (LLMs) such as GPT or BERT consumes a significant amount of energy (Lannelongue *et al.*, 2021; Patterson *et al.*, 2021). Not in all business cases is it necessary to train an individual model from scratch (Lacoste *et al.*, 2019). Often, the fine-tuning of an existing model is a favorable option, leading to a lower energy consumption (Wang *et al.*, 2023; Lacoste *et al.*, 2019) while providing comparable accuracy (Patterson *et al.*, 2021). Besides the application of LLMs, there are also pre-trained models in other domains that can be used to save energy [e.g. adjusted Yolo Computer Vision Object detection with Coco data sets (Zhou, 2024)].

4.1.2.3 Usage phase. After the training and deployment of the algorithms (Chapman *et al.*, 2000), energy is necessary to run the ML model (Kaack *et al.*, 2022). Using the models in the usage phase (also referred to as inference) might need more resources than in the training phase (Desislavov *et al.*, 2023; Patterson *et al.*, 2021). However, spending more energy in the training phase to reach a more energy efficient model for usage might be a beneficial option (Patterson *et al.*, 2021). Here, the number of uses impacts the energy consumption. The more it is used; the more energy is needed. In special application areas like language models, the number of inputs or sequence length can have an impact (Wang *et al.*, 2023). Furthermore, the complexity of the model might influence the energy consumption (Mavromatis *et al.*, 2024). Approaches like pruning reduce the energy consumption of models during usage (Yang *et al.*, 2017). Like in the training phase, the existing IT infrastructure (Lacoste *et al.*, 2019; Patterson *et al.*, 2021; Lannelongue *et al.*, 2021) used to perform the model (e.g. on a server or a personal computer) impacts the energy consumption level, as well as factors such as the model size (Han *et al.*, 2015).

For this purpose, insights can be provided through observations from the Green Algorithms Calculator focusing on running algorithms in general (https://calculator.green-algorithms.org/; Lannelongue *et al.*, 2021). For instance, running a sales forecasting ML algorithm (e.g. linear regression) on a personal computer for 5 min (1 Core CPU, Core i7-4790, 16 GB RAM in Germany) lead to an energy consumption of 7.01 Wh and 2.37 g CO_2. In contrast, running the same analysis in the AWS Cloud Germany environment only leads to 5.39 Wh and 1.83 g CO_2 emissions. This example is shown in Figure 4. Differences in the hardware environment influence the energy consumption. Also, the ML CO_2 Calculator Recommender System (mlco2.github.io) can be used instead of the Green Algorithms

Figure 4. Example energy consumption in cloud environment for using a machine learning algorithm with Green Algorithm Calculator (https://calculator.green-algorithms.org)

Calculator (Lannelongue *et al.*, 2021). For running a ML service, it might be easy to estimate the energy consumption using the Green Algorithm Calculator.

In addition, it must be considered that, if changes in the data patterns occur (e.g. different consumer behavior, change in social trend), the model loses its value. In this case, it must be retrained again and subsequently, the analytics cycle (Chapman *et al.*, 2000) restarts. All prior phases must be repeated, inevitably leading to an increasing energy consumption.

5. Demonstration and evaluation

5.1 Demonstration

The first version of the developed framework was demonstrated within a session of an AI track during a leading service management conference (QUIS19) based on a conference abstract presentation in mid-2025 in Rome, Italy. After the presentation, the design, the sources and the concepts of the framework were discussed with the scientific community. Based on this feedback and insights, a further evaluation with experts was initiated and conducted.

5.2 Evaluation

Data analysis revealed that the experts underlined the framework's explainability regarding the energy consumption of ML services (Mean = 1.3). Additionally, they evaluated the framework as understandable (Mean = 1.3) and useful (Mean = 1.3) as well. Furthermore, the experts also highlighted extensions and further argumentations for the framework, which are described below.

Pre-training phase: All experts agree with the importance of this phase related to energy consumption. One expert (*Expert 1*) expressed the opinion that data understanding and related data acquisition are critical aspects for database customers in the context of energy consumption.

Training phase: All experts agree with the importance of this phase. Further comments were given by the experts. Expert 2 stated that generative AI has the potential to reduce the training effort and consequently the energy consumption through approaches such as Few-Shot, One-Shot Learning or Prompt-Engineering. Also, the complexity of the different algorithms and the need for (hyper)-parameter optimization was mentioned as a critical aspect regarding energy consumption worth considering. This opinion was shared by other experts, too (*Expert 5*). One highlighted the importance of hyperparameter optimization in this context. From his/her point of view, data complexity and algorithms belong together. Another expert (*Expert 3*) even suggested and proposed to train and optimize hyperparameter first for a subset or one specific field and afterwards to transfer and expand the findings to other areas. Additionally, as described in the framework, the use of pre-trained models to save effort and resources was highlighted by this expert (*Expert 3*). In this context, another expert mentioned the on-going discussion about models' performances and the effort for (re-)trainings and improvements (*Expert 5*). This argument was taken up by another person, who added that pattern identification might be helpful for the decision whether it is necessary to train a new model or not. Different experts agreed that everything is use case specific (*Expert 2*) and underlined the statement of the "No free lunch theorem"(Wolpert and Macready, 1997). One added the aspect, that from a data point of view everything is individual (*Expert 4*). Also, upcoming approaches where reusable software codes are generated to avoid that the same prompt is handled by a LLM repeatedly can save energy (*Expert 1*). Although the benefits, generative AI brings, also limitations and negative effects due to the energy consumption are mentioned. The same expert also stated that due to the phenomenon "AI slop" (Madsen and Puyt, 2025), energy is wasted because the resulting ML models are not usable for the intended task (*Expert 1*). S/he argued that currently the approaches lack on logical depth, hence the given output is unsuitable and inefficient (*Expert 1*).

Usage phase: Also, for this phase the experts agreed regarding its importance within the framework. It emerged that to achieve optimization of the energy consumption, it is important to know how often the model is used, respectively inferences will happen, as one expert mentioned (*Expert 3*). Additionally, s/he mentioned that it is important to have it via batch and not always on demand (*Expert 3*). Another expert, however, argued about the efficient and sustainable usage of hardware (on-premise vs public cloud) in this context (*Expert 4*). This statement was continued by the proposition to integrate ideas of idle time energy consumption, when the hardware system is not used; commonly in nonpublic cloud environments (*Expert 4*). In sum, the expert calls for a properly considered usage of hardware resources with less or no idle time. Expert 1 also named the use of reasoning models as a driver for energy consumption due to the longer inference phase. This calls for considerations whether and in which case these models are necessary. Furthermore, Expert 1 and Expert 2 underlined the importance of the loops in the framework, which should be highlighted. Loops are needed if, for instance, in the usage phase the ML services are not working as expected, one must go back to the training phase. This stands also in line with previous recommendations (Chapman *et al.*, 2000). In addition, the energy mix was named as a relevant aspect. This expert added that this aspect is also pushed by law, which underlines its importance (*Expert 6*). Therefore, where the data center is located must be considered and discussed in the context of its CO_2 footprint (*Expert 6*). As another potential parallelization was mentioned.

5.3 Re-design and re-definition: the final framework

The final framework, which integrates the different phases from pre-training to usage, is displayed in Figure 5. It includes the important extensions identified during the expert evaluation.

First, in the training phase the "Used ML Algorithm" as part of the model was revised. The aspects mentioned by the experts have been incorporated to explain and illustrate this step more clearly in the model. Now, the framework provides a more concrete description of the energy consumption in the training phase. This fosters the understanding of energy-intensive actions like hyper-parameter optimization. It highlights also possibilities to reduce energy consumption, as mentioned by using pre-trained models and generative AI. This result provides a better basis for decision-making.

Against the backdrop of the twin transition, it is now possible to weigh more effectively the ecological costs against the economic benefits and hence, to ensure a proper deployment. In the usage phase, expert ideas in terms of idle time and single versus batch processing to

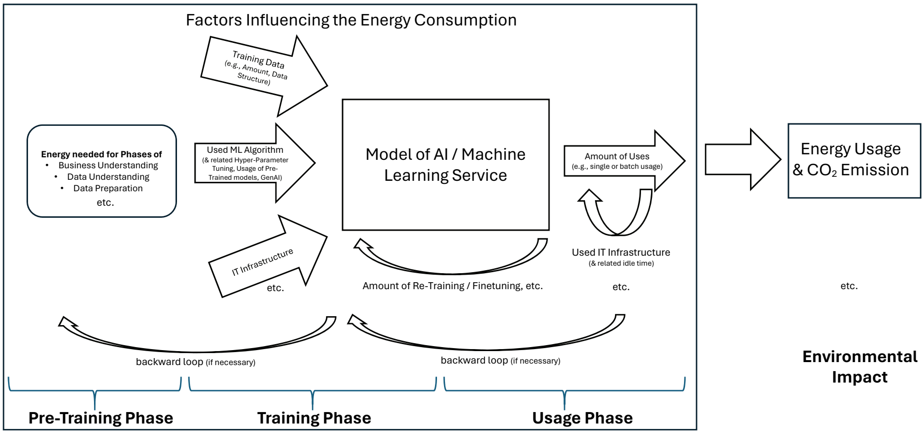

Figure 5. Final framework of environmental impact of machine learning services based on findings in literature and recommender system observations as well as expert evaluations

save energy were integrated into the model. These considerations are essential in the service domain. Considerations of these points help to understand and to evaluate under which conditions a service must always be available or on-premise. Since services cannot be stockpiled, there may be significant potential here to achieve the same quality and economic benefits while operating sustainably. Furthermore, a feedback loop was integrated for a better coverage of backward steps that might include internal or external factors as mentioned by the experts. This supports an awareness that a proposed solution − as common in the service sector − must be continually reevaluated and adjusted as necessary. For instance, in the field of AI services, as mentioned by the experts, new capabilities (reasoning) or legal regulations may play a role in this regard. These feedback loops further reinforce the framework's holistic nature, demonstrating its applicability to existing approaches and thereby helping to gradually integrate it into existing structures to foster twin transition.

6. Conclusion, discussion and implications

Besides the positive impacts, AI services also entail a shadow side, as they can generate adverse environmental consequences (Kaack *et al.*, 2022; Keller *et al.*, 2024). This paper focuses on the energy consumption of ML services by identifying key factors influencing energy demand and associated CO_2 emissions. Previous research findings and evidence from different recommender systems were integrated in one framework evaluated in terms of its content and validity. The proposed framework enables the energy impacts analysis across the service life cycle and within service delivery processes.

6.1 Theoretical implications

The present paper extends previous research (e.g. Keller *et al.*, 2024; Lannelongue *et al.*, 2021; Zanger *et al.*, 2024) by introducing and enabling a wider and more holistic view on this still underexplored topic. The research is also in line with the need formulated by Kaack *et al.* (2022) for a more operational and holistic understanding of ML and its influences on environment and climate change. The framework attempts to address the need to measure AI adoption in different service settings such as healthcare (Bratan *et al.*, 2024; Richie, 2022), education (Zhang *et al.*, 2026) and public and government sectors (Alhosani and Alhashmi, 2024; Van Noordt and Misuraca, 2022). Furthermore, it responds to Schwartz *et al.* (2020) by promoting more research and reporting of green AI by developing a framework and focusing more on energy consumption instead of overall resource efficiency as emphasized by the authors.

From a theoretical perspective, this study contributes to service research by introducing an energy-oriented perspective into the analysis of AI-based services (Hwang *et al.*, 2026). In particular, it extends service life cycle approaches by explicitly integrating energy consumption and environmental impact as critical dimensions of service design, delivery and usage. By conceptualizing energy consumption across the ML process, the framework advances current understanding of how digital service innovation processes can be aligned with sustainability objectives, thus contributing to the broader discussion on sustainable service and the twin transition.

In addition, the proposed framework enables the identification and systemic consideration of multiple factors. Current research insights on energy consumption levels in the context of ML were merged with well-known data analytics project steps from CRISP-DM (Chapman *et al.*, 2000) and enriched with recommender system observations. CRISP-DM is a major standard for analytical projects nowadays (Schröer *et al.*, 2021). Previous research on frameworks of ML observations assessing energy efficiency (Fischer *et al.*, 2022) and views of ML operations and energy consumption (Mavromatis *et al.*, 2024) was extended by needed steps from the pre-training phase as well as a more overall view of a data analytics project (Chapman *et al.*, 2000). The importance of this extension stands also in line with other data analytics process models for analytical projects like the TDSP, DASC-PM (Schulz *et al.*, 2020). Finally, the proposed framework was evaluated with experts from practice and research and shows its relevance for academic research and managerial practice.

6.2 Practical implications

From a managerial perspective, the framework supports the sustainable design of AI/ML-based services by enabling managers to integrate energy-related considerations from the early stages until usage. By integrating these stages into a unified perspective, the framework overcomes fragmented approaches to energy management and provides a holistic view of energy consumption. This perspective, as suggested by Hosseini *et al.* (2025), also encourages managers to critically reflect on whether AI-based solutions represent the most appropriate option in specific service contexts.

Practice benefits from the developed framework through its explainable character, which helps to identify and evaluate less visible drivers of energy demand, identifying actionable levers for their mitigation.

By making these factors explicit, the framework supports managerial decision-making and enables a more systematic planning of key steps influencing the energy consumption of ML-based services. In addition, it provides a structured basis for anticipating and preparing sustainability and energy reporting requirements. Indeed, energy consumption reporting and related carbon emission are important for both practice and research (Henderson *et al.*, 2020).

Furthermore, the experts involved in the evaluation underlined the practical usefulness of the developed framework. Its adoption is not only necessary to retain competitiveness in an environment increasingly oriented toward sustainability but also offers economic benefits such as improved operational efficiency and cost reduction through more energy-aware service design. This aligns with the needs to ensure both sustainability and efficiency of AI use in service (Alkire *et al.*, 2024; Hwang *et al.*, 2026).

The study results provide interesting implications also for end-users and, more generally, for society. More efficient ML services might reduce the environmental footprint of all AI-based interactions, supporting quality of life through greener services and strengthening confidence that AI can be aligned with sustainability objectives.

Finally, the framework can represent a supportive tool for policymakers, providing a practical basis for integrating energy-use criteria into AI governance, procurement policies, ESG reporting and national or regional sustainability goals. In this sense, the framework facilitates the integration of sustainability criteria into existing service systems and governance structures, supporting a more consistent alignment between digital innovation and environmental commitments.

In conclusion, the implications derived from this study support practitioners to aligning with the United Nations 2030 Agenda (SDGs 7, 12 and 13), by promoting energy efficiency, responsible production patterns and the mitigation of environmental impacts related to AI use in service.

6.3 Limitations and future research

Limitations can be found in that the model does not provide a detailed view of the identified impact factors and the so far missing empirical evaluation of the factors' quantification. Furthermore, the focus was set only on the energy consumption. Other important aspects such as the model's accuracy have been neglected. The integration of accuracy aspects should also be taken into account in the future as well as aspects of ML operations and ML life cycle (Fischer *et al.*, 2022; Mavromatis *et al.*, 2024). Sometimes a slightly more accurate model may need disproportionately more energy (Kaack *et al.*, 2022). Thus, it is important to investigate also accuracy aspects in the future. However, it has to be found a good balance between accuracy optimization and resource consumption in relation to red AI and green AI (Schwartz *et al.*, 2020). Future research is encouraged to address this gap by empirical investigations, further developments and through the conducting of case studies aiming to clarify its usefulness in organizational setups. The influence of carbon-free data centers and certificates should be considered in future research, too. Additionally, limitations arise from the used energy recommender systems themselves (Zanger *et al.*, 2024; Lannelongue *et al.*, 2021; Lacoste *et al.*, 2019) and reviewed literature. The usage and integration of other recommender systems or code-based frameworks could also be a promising starting point for future research. Besides energy demand, further important environmental aspects such as water consumption exist. In the future, these should be taken into account, too, to enlarge the objectives by investigating AI related areas besides ML.

Ethics statement

Before data collection within the expert evaluation, we provided to the participants a detailed information about the study's procedures and fully informed about the purpose of the research and the study's aims, their rights as respondents (e.g. the right to refuse participation or withdraw from the study), and the measures taken to ensure their confidentiality and anonymity. During the expert call with the participants, informed consent was obtained from all participants before taking part in the study.

Furthermore, the authors confirm that the study adheres to the relevant ethical guidelines for human participants, and that it followed ethical measures in compliance with institutional and international guidelines, including: maintaining and respecting the autonomy, confidentiality and anonymity of all respondents throughout the study via the virtual call. Using collected data solely for the purposes outlined in the study. Ensuring transparency in data handling and reporting. Avoiding any form of coercion or undue influence in securing participation. The statement is based on the template and the recommendations of the publisher.

Acknowledgements
The authors thank the editors and anonymous reviewers of the journal for supporting this work and the experts during the evaluation phase of the framework. Furthermore, the authors thank the QUIS19 conference community for discussions, where a preliminary version of this work was presented and demonstrated.

AI declaration statement
During the preparation of this article, B.K. used Microsoft Copilot and Deepl (free translator) for spelling correction, grammar and translations. L.P. used Microsoft Copilot for checking the English language. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article. The statement based on the template and the recommendations of the publisher.

References

Akter, S., Hossain, M.A., Sajib, S., Sultana, S., Rahman, M., Vrontis, D. and McCarthy, G. (2023), "A framework for AI-powered service innovation capability: review and agenda for future research", *Technovation*, Vol. 125, p. 102768.

Alhosani, K. and Alhashmi, S.M. (2024), "Opportunities, challenges, and benefits of AI innovation in government services: a review", *Discover Artificial Intelligence*, Vol. 4 No. 1.

Alkire, L., Bilgihan, A., Bui, M., Buoye, A.J., Dogan, S. and Kim, S. (2024), "RAISE: leveraging responsible AI for service excellence", *Journal of Service Management*, Vol. 35 No. 4, pp. 490-511.

Antons, D. and Breidbach, C.F. (2018), "Big data, big insights? Advancing service innovation and design with machine learning", *Journal of Service Research*, Vol. 21 No. 1, pp. 17-39.

Awad, M. and Khanna, R. (2015), "*Machine Learning*, " Efficient Learning Machines, Apress, Berkeley, CA.

Ayvaz, S. and Alpay, K. (2021), "Predictive maintenance system for production lines in manufacturing: a machine learning approach using IoT data in real-time", *Expert Systems with Applications*, Vol. 173, p. 114598.

Ban, G.Y. and Keskin, N.B. (2021), "Personalized dynamic pricing with machine learning: high-dimensional features and heterogeneous elasticity", *Management Science*, Vol. 67 No. 9, pp. 5549-5568.

Barile, S., Grimaldi, M., Loia, F. and Sirianni, C.A. (2020), "Technology, value co-creation and innovation in service ecosystems: toward sustainable co-innovation", *Sustainability*, Vol. 12 No. 7, p. 2759.

Bianchini, S., Damioli, G. and Ghisetti, C. (2023), "The environmental effects of the 'twin' green and digital transition in European regions", *Environmental and Resource Economics*, Vol. 84 No. 4, pp. 877-918.

Bratan, T., Heyen, N.B., Hüsing, B., Marscheider-Weidemann, F. and Thomann, J. (2024), "Hypotheses on environmental impacts of AI use in healthcare", *The Journal of Climate Change and Health*, Vol. 16, p. 100299.

Briganti, G. and Le Moine, O. (2020), "Artificial intelligence in medicine: today and tomorrow", *Frontiers in Medicine*, Vol. 7, p. 509744.

Broday, E.E. (2022), "The evolution of quality: from inspection to quality 4.0", *International Journal of Quality and Service Sciences*, Vol. 14 No. 3, pp. 368-382.

Broekens, J., Heerink, M. and Rosendal, H. (2009), "Assistive social robots in elderly care: a review", *Gerontechnology*, Vol. 23 No. s, pp. 94-103.

Bukhari, S.M.S. and Akhtar, R. (2024), "Leveraging artificial intelligence to revolutionize six sigma: enhancing process optimization and predictive quality control", *Contemporary Journal of Social Science Review*, Vol. 2 No. 04, pp. 1932-1948.

Carneiro, F., Lizarelli, F.L., Suárez-Barraza, M., Antony, J., Sony, M. and Ma, J. (2026), "Exploring the impact of kaizen practices on ESG performance in manufacturing: the role of digital technologies and organizational culture", *Production Planning and Control*, *Ahead-of-Print*.

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C. and Wirth, R. (2000), "CRISP-DM 1.0", *CRISP-DM Consortium*, Vol. 76 No. 3.

Christmann, A.S., Crome, C., Graf-Drasch, V., Oberländer, A.M. and Schmidt, L. (2024), "The twin transformation butterfly: capabilities for an integrated digital and sustainability transformation", *Business and Information Systems Engineering*, Vol. 66 No. 4, pp. 489-505.

Davenport, T., Guha, A. and Grewal, D. (2020), "How artificial intelligence will change the future of marketing", *Journal of the Academy of Marketing Science*, Vol. 48 No. 1, pp. 24-42.

Davis, F. (1985), "*A Technology Acceptance Model for Empirically Testing New End-User Information Systems - Theory and Results*, ", MA Institue of Technology.

Desislavov, R., Martínez-Plumed, F. and Hernández-Orallo, J. (2023), "Trends in AI inference energy consumption: beyond the performance-vs-parameter laws of deep learning", *Sustainable Computing: Informatics and Systems*, Vol. 38, p. 100857.

Dhar, P. (2020), "The carbon impact of artificial intelligence", *Nature Machine Intelligence*, Vol. 2 No. 8, pp. 423-425.

Di Vaio, A., Palladino, R., Hassan, R. and Escobar, O. (2020), "Artificial intelligence and business models in the Sustainable Development Goals perspective: a systematic literature review", *Journal of Business Research*, Vol. 121, pp. 283-314.

Dörr, S. (2022), "*Corporate Digital Responsibility: Managing Corporate Responsibility and Sustainability in the Digital Age*", 1st ed., Springer.

Eilam, T., Bello-Maldonado, P., Bhattacharjee, B., Costa, C., Lee, E.K. and Tantawi, A. (2023), "Towards a methodology and framework for AI sustainability metrics", In *Proceedings of the 2nd Workshop on Sustainable Computer Systems*, *ACM Digital Library*, pp. 1-7.

European Union (2022), "Towards a green and digital future – key requirements for successful twin transitions in the European Union, Publications Office of the European Union", available at: https://data.europa.eu/doi/10.2760/977331

Fischer, R., Jakobs, M., Mücke, S. and Morik, K. (2022), "A unified framework for assessing energy efficiency of machine learning", *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, Springer, pp. 39-54.

Fragapane, G., Ivanov, D., Peron, M., Sgarbossa, F. and Strandhagen, J.O. (2022), "Increasing flexibility and productivity in Industry 4.0 production networks with autonomous mobile robots and smart intralogistics", *Annals of Operations Research*, Vol. 308 Nos 1-2, pp. 125-143.

García-Martín, E., Lavesson, N., Grahn, H., Casalicchio, E. and Boeva, V. (2018), "How to measure energy consumption in machine learning algorithms", In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pp. 243-255.

García-Martín, E., Rodrigues, C.F., Riley, G. and Grahn, H. (2019), "Estimation of energy consumption in machine learning", *Journal of Parallel and Distributed Computing*, Vol. 134, pp. 75-88.

Goel, B., McKee, S.A. and Själander, M. (2012), "Techniques to measure, model, and manage power", *Advances in Computers*, Vol. 87, pp. 7-54.

Guandalini, I. (2022), "Sustainability through digital transformation: a systematic literature review for research guidance", *Journal of Business Research*, Vol. 148, pp. 456-471.

Hammerschmidt, J., Burtscher, J., Gast, J., Kraus, S. and Puumalainen, K. (2025), "Navigating the twin transformation: How digitalization and sustainability shape the future", *Strategic Change*, Vol. 35 No. 1.

Han, S., Mao, H. and Dally, W.J. (2015), "Deep compression: Compressing deep neural networks with pruning, trained quantization and huffman coding", arXiv preprint arXiv:1510.00149.

Hasan, S.M., Islam, T., Saifuzzaman, M., Ahmed, K.R., Huang, C.H. and Shahid, A.R. (2025), "Carbon emission quantification of machine learning: a review", *IEEE Transactions on Sustainable Computing*, Vol. 10 No. 6.

Heinonen, K., Kietzmann, J. and Pitt, L.F. (2020), "Guest Editorial: artificial intelligence and machine learning in service management", *Journal of Service Management*, Vol. 31 No. 2, pp. 137-143.

Henderson, P., Hu, J., Romoff, J., Brunskill, E., Jurafsky, D. and Pineau, J. (2020), "Towards the systematic reporting of the energy and carbon footprints of machine learning", *Journal of Machine Learning Research*, Vol. 21 No. 248, pp. 1-43.

Hevner, A., March, S., Park, J. and Ram, S. (2004), "Design science in information systems research", *MIS Quarterly*, Vol. 28 No. 1, pp. 75-105.

Hoehndorf, R. and Queralt-Rosinach, N. (2017), "Data science and symbolic AI: synergies, challenges and opportunities", *Data Science*, Vol. 1 Nos 1-2, pp. 27-38.

Holmes, W. and Tuomi, I. (2022), "State of the art and practice in AI in education", *European Journal of Education*, Vol. 57 No. 4, pp. 542-570.

Hossain, M.F. and Dhanekula, A. (2023), "Smart continuous improvement with artificial intelligence, big data, and lean tools for zero defect manufacturing systems", *American Journal of Scholarly Research and Innovation*, Vol. 2 No. 1, pp. 254-282.

Hosseini, M., Gao, P. and Vivas-Valencia, C. (2025), "A social-environmental impact perspective of generative artificial intelligence", *Environmental Science and Ecotechnology*, Vol. 23, p. 100520.

Huang, M.H. and Rust, R.T. (2018), "Artificial intelligence in service", *Journal of Service Research*, Vol. 21 No. 2, pp. 155-172.

Huang, M.H. and Rust, R.T. (2021), "Engaged to a robot? The role of AI in service", *Journal of Service Research*, Vol. 24 No. 1, pp. 30-41.

Hwang, B.N., Jitanugoon, S. and Puntha, P. (2026), "AI integration in service delivery: enhancing business and sustainability performance amid challenges", *Journal of Services Marketing*, Vol. 40 No. 2, pp. 263-281.

Hwang, S., Kim, J., Park, E. and Kwon, S.J. (2020), "Who will be your next customer: a machine learning approach to customer return visits in airline services", *Journal of Business Research*, Vol. 121, pp. 121-126.

Iandola, F.N., Han, S., Moskewicz, M.W., Ashraf, K., Dally, W.J. and Keutzer, K. (2016), "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size", arXiv Preprint, arXiv:1602.07360.

IBM (2024), "AI forecasting", IBM Planning Analytics Products, available at: www.ibm.com/products/planning-analytics/ai-forecasting

IBM (2026), "What are large language models (LLMs)?" IBM think", available at: www.ibm.com/think/topics/large-language-models

Islam, M.S., Zisad, S.N., Kor, A.L. and Hasan, M.H. (2023), "Sustainability of machine learning models: an energy consumption centric evaluation", *International Conference on Electrical, Computer and Communication Engineering*, *IEEE*, pp. 1-6.

Jose, R., Panigrahi, S.K., Patil, R.A., Fernando, Y. and Ramakrishna, S. (2020), "Artificial intelligence-driven circular economy as a key enabler for sustainable energy management", *Materials Circular Economy*, Vol. 2 No. 1, pp. 1-7.

Kaack, L.H., Donti, P.L., Strubell, E., Kamiya, G., Creutzig, F. and Rolnick, D. (2022), "Aligning artificial intelligence with climate change mitigation", *Nature Climate Change*, Vol. 12 No. 6, pp. 518-527.

Keller, B., Möhring, M. and Augenstein, F. (2024), "Data analytics as a service–challenges and opportunities", *Lecture Notes in Informatics*, Vol. 352, pp. 1481-1489.

King, W.R. and He, J. (2006), "A meta-analysis of the technology acceptance model", *Information and Management*, Vol. 43 No. 6, pp. 740-755.

Kumar, A., Singh, S., Srivastava, N., Srivastava, A., Jain, A., Ayaz, M., Solomon, P. and Kumar, A. (2026), "Artificial intelligence (AI) powered marketing strategy: a systematic literature review and future research direction", *International Journal of Quality and Service Sciences*.

Lacoste, A., Luccioni, A., Schmidt, V. and Dandres, T. (2019), "Quantifying the carbon emissions of machine learning".

Lannelongue, L., Grealey, J. and Inouye, M. (2021), "Green algorithms: quantifying the carbon footprint of computation", *Advanced Science*, Vol. 8 No. 12, p. 2100707.

Madsen, D.Ø. and Puyt, R.W. (2025), "The 7Vs of AI slop: a typology of generative waste", *available at SSRN 5558018*.

Malikah, W, Ayat, M. and Shin, W.S. (2025), "Quality 4.0: adoption approaches and future directions", *International Journal of Quality and Service Sciences*, Vol. 17 No. 4.

Mavromatis, I., Katsaros, K. and Khan, A. (2024), "Computing within limits: an empirical study of energy consumption in ML training and inference", arXiv preprint arXiv:2406.14328.

Mohammadian, N., Valilai, O.F. and Schlüter, A. (2025), "Sustainable design and repair: leveraging circular economy and machine learning for product development", *Journal of Open Innovation: Technology, Market, and Complexity*, Vol. 11 No. 1, p. 100469.

Möhring, M., Keller, B., Schmidt, R., Schönitz, F., Mohr, F. and Scheuerle, M. (2022), "Analytics in Industry 4.0: Investigating the challenges of unstructured data", *Lecture Notes in Business Information Processing*, Vol. 462, pp. 113-125.

Moloi, T and Obeid, H. (2024), "Perceptions of South African accountants on factors with a role in the adoption of artificial intelligence in financial reporting", *Journal of Risk and Financial Management*, Vol. 17 No. 9, p. 389.

Montes-Pineda, Ó. and Garrido-Yserte, R. (2024), "Artificial intelligence and circular economy: what is new for business model innovation?", *Artificial Intelligence and Business Transformation: Impact in HR Management, Innovation and Technology Challenges*, pp. 41-59.

Mustafa, F., Smolarski, J. and Elamer, A.A. (2025), "The convergence of artificial intelligence and sustainability reporting: a systematic review of applications, challenges and future directions", *Business Strategy and the Environment*, Vol. 34 No. 8.

Nishant, R., Kennedy, M. and Corbett, J. (2020), "Artificial intelligence for sustainability: challenges, opportunities, and a research agenda", *International Journal of Information Management*, Vol. 53, p. 102104.

Noman, A.A., Akter, U.H., Pranto, T.H. and Haque, A.K.M. (2022), "Machine learning and artificial intelligence in circular economy: a bibliometric analysis and systematic literature review", *Annals of Emerging Technologies in Computing*, Vol. 6 No. 2, pp. 13-40.

Nvidia (2015), "*GPU-Based Deep Learning Inference: A Performance and Power Analysis*, ", Nvidia Whitepaper.

Oduro, S., De Nisco, A. and Mainolfi, G. (2023), "Do digital technologies pay off? A meta-analytic review of the digital technologies/firm performance nexus", *Technovation*, Vol. 128, p. 102836.

OpenAI (2018), "OpenAI: AI and Compute", available at: https://openai.com/index/ai-and-compute/

Oppioli, M., Sousa, M.J., Sousa, M. and de Nuccio, E. (2023), "The role of artificial intelligence for management decision: a structured literature review", *Management Decision*, Vol. 63 No. 10, pp. 3262-3280.

Pagoropoulos, A., Pigosso, D.C. and McAloone, T.C. (2017), “The emergent role of digital technologies in the circular economy: a review”, *Procedia CIRP*, Vol. 64, pp. 19-24.

Patterson, D., Gonzalez, J., Le, Q., Liang, C., Munguia, L.M., Rothchild, D. and Dean, J. (2021), “Carbon emissions and large neural network training”.

Peffers, K., Tuunanen, T., Rothenberger, M.A. and Chatterjee, S. (2007), “A design science research methodology for information systems research”, *Journal of Management Information Systems*, Vol. 24 No. 3, pp. 45-77.

Philipp, R., Mladenow, A., Strauss, C. and Völz, A. (2020), “Machine learning as a service: Challenges in research and applications”, In *Proceedings of the 22nd International Conference on Information Integration and Web-based Applications and Services*, *Chiang Mai*, *Thailand*, pp. 396-406.

Richie, C. (2022), “Environmentally sustainable development and use of artificial intelligence in health care”, *Bioethics*, Vol. 36 No. 5, pp. 547-555.

Ronaghi, M.H. (2023), “The influence of artificial intelligence adoption on circular economy practices in manufacturing industries”, *Environment, Development and Sustainability*, Vol. 25 No. 12, pp. 14355-14380.

Samuel, A.L. (1959), “Some studies in machine learning using the game of checkers”, *IBM Journal of Research and Development*, Vol. 3 No. 3, pp. 210-229.

Schröer, C., Kruse, F. and Gómez, J.M. (2021), “A systematic literature review on applying CRISP-DM process model”, *Procedia Computer Science*, Vol. 181, pp. 526-534.

Schulz, M., Neuhaus, U., Kaufmann, J., Badura, D., Kuehnel, S., Badwitz, W. and Lanquillon, C. (2020), “*Introducing DASC-PM: A Data Science Process Model*”, ACIS, AIS, Wellington.

Schwartz, R., Dodge, J., Smith, N.A. and Etzioni, O. (2020), “Green AI”, *Communications of the ACM*, Vol. 63 No. 12, pp. 54-63.

Sigala, M., Ooi, K.B., Tan, G.W.H., Aw, E.C.X., Cham, T.H., Dwivedi, Y.K. and Wirtz, J. (2024), “ChatGPT and service: opportunities, challenges, and research directions”, *Journal of Service Theory and Practice*, Vol. 34 No. 5, pp. 726-737.

Singh, A., Kanaujia, A., Singh, V.K. and Vinuesa, R. (2024), “Artificial intelligence for Sustainable Development Goals: bibliometric patterns and concept evolution trajectories”, *Sustainable Development*, Vol. 32 No. 1, pp. 724-754.

Strubell, E., Ganesh, A. and McCallum, A. (2019), “Energy and policy considerations for deep learning in NLP”, In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pp. 3645-3650.

Tabares, S., Parida, V. and Chirumalla, K. (2025), “Twin transition in industrial organizations: conceptualization, implementation framework, and research agenda”, *Technological Forecasting and Social Change*, Vol. 213, p. 123995.

Tamburrini, G. (2022), “The AI carbon footprint and responsibilities of AI scientists”, *Philosophies*, Vol. 7 No. 1, p. 4.

Teixeira, J.G., Patrício, L. and Tuunanen, T. (2019), “Advancing service design research with design science research”, *Journal of Service Management*, Vol. 30 No. 5, pp. 577-592.

Tutore, I., Parmentola, A., Di Fiore, M.C. and Calza, F. (2024), “A conceptual model of artificial intelligence effects on circular economy actions”, *Corporate Social Responsibility and Environmental Management*, Vol. 31 No. 5, pp. 4772-4782.

Van Noordt, C. and Misuraca, G. (2022), “Artificial intelligence for the public sector: results of landscaping the use of AI in government across the European Union”, *Government Information Quarterly*, Vol. 39 No. 3, p. 101714.

Wang, Q., Li, Y. and Li, R. (2024), “Ecological footprints, carbon emissions, and energy transitions: the impact of artificial intelligence (AI)”, *Humanities and Social Sciences Communications*, Vol. 11 No. 1, pp. 1-18.

Wang, X., Na, C., Strubell, E., Friedler, S. and Luccioni, S. (2023), "Energy and carbon considerations of fine-tuning BERT".

Webster, J. and Watson, R.T. (2002), "Analyzing the past to prepare for the future: writing a literature review", *MIS Quarterly*, Vol. 26 No. 2, pp. xiii-xxiii.

Wirth, R. and Hipp, J. (2000), "CRISP-DM: towards a standard process model for data mining", *Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining*, *Semantic scholar*, pp. 29-39.

Wolpert, D.H. and Macready, W.G. (1997), "No free lunch theorems for optimization", *IEEE Transactions on Evolutionary Computation*, Vol. 1 No. 1, pp. 67-82.

Yang, T.J., Chen, Y.H. and Sze, V. (2017), "Designing energy-efficient convolutional neural networks using energy-aware pruning", In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5687-5695.

Yao, Y., Xiao, Z., Wang, B., Viswanath, B., Zheng, H. and Zhao, B.Y. (2017), "Complexity vs. performance: empirical analysis of machine learning as a service", In *Proceedings of the 2017 Internet Measurement Conference*, pp. 384-397.

Yee, L., Chui, M., Robert, R. and Isler, M. (2024), "Mc Kinsey Technology Trends Outlook 2024", available at: www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-top-trends-in-tech.(29.11.2024)

Zanger, M., Schulz, A., Grodmeier, L., Agaj, D., Schindler, R., Weiss, L. and Möhring, M. (2024), "HollerithEnergyML: a prototype of a machine learning energy consumption recommender system", *Lecture Notes in Informatics*, Vol. 352, pp. 1519-1523.

Zhang, W., Song, X., Lin, M. and Vasile, E. (2026), "Artificial intelligence enters higher education: does research drive carbon reduction or accelerate emissions?", *International Journal of Sustainability in Higher Education*, pp. 1-18.

Zhou, Y. (2024), "A YOLO-NL object detector for real-time detection", *Expert Systems with Applications*, Vol. 238, p. 122256.

Zhu, O.Y. and Dolnicar, S. (2025), "Achieving social impact through business research aimed at contributing to climate change mitigation", *Journal of Social Impact in Business Research*, Vol. 1 No. 2, pp. 53-69.

Zhuk, A. (2023), "Artificial intelligence impact on the environment: hidden ecological costs and ethical-legal issues", *Journal of Digital Technologies and Law*, Vol. 1 No. 4, pp. 932-954.

Zonnenshain, A. and Kenett, R.S. (2020), "Quality 4.0—the challenging future of quality engineering", *Quality Engineering*, Vol. 32 No. 4, pp. 614-626.

Corresponding author
Michael Möhring can be contacted at: michael.moehring@reutlingen-university.de