Rising like a Phoenix: Emerging from the Pandemic and Reshaping Human Endeavors with Digital Technologies ICIS 2023

IT Implementation and Adoption

Dec 11th, 12:00 AM

Artificial Intelligence as a Service: Trade-Offs Impacting Service Design and Selection

Kathrin Brecker
Karlsruhe Institute of Technology, brecker@kit.edu

Sebastian Lins
Karlsruhe Institute of Technology, sebastian.lins@kit.edu

Manuel Trenz
University of Goettingen, trenz@uni-goettingen.de

Ali Sunyaev
Karlsruhe Institute of Technology, sunyaev@kit.edu

Artificial Intelligence as a Service: Trade-Offs Impacting Service Design and Selection

Completed Research Paper

Kathrin Brecker
Karlsruhe Institute of Technology
Karlsruhe, Germany
brecker@kit.edu

Sebastian Lins
Karlsruhe Institute of Technology
Karlsruhe, Germany
lins@kit.edu

Manuel Trenz
University of Goettingen
Goettingen, Germany
trenz@uni-goettingen.de

Ali Sunyaev
Karlsruhe Institute of Technology
Karlsruhe, Germany
sunyaev@kit.edu

Abstract

AI as a Service (AIaaS) is a promising path to leverage AI capabilities from the cloud. However, there is no one-size-fits-all service, but heterogenous service options since interdependent AIaaS characteristics require trade-offs. We lack knowledge on these trade-offs and how they result from conflicting characteristics. Therefore, we interviewed 39 AIaaS providers, customers, and consultants to provide rich descriptions of interdependent characteristics and uncover resulting trade-offs and their consequences. This study contributes to a better understanding of the inner functioning and interplay of AIaaS characteristics and discusses how this complex nature of service offerings impacts providers' design and customers' selection decisions.

Keywords: AIaaS, AI services, trade-offs, design, selection, adoption

Introduction

Nowadays, organizations have the disruptive opportunity to acquire AI capabilities as external services. Eminent cloud providers, such as Amazon, Google, IBM, Microsoft, and small and medium-sized enterprises have started offering AI as a Service (AIaaS) (Sundberg & Holmström, 2022; Zapadka et al., 2020). AIaaS refers to cloud-based systems providing services to organizations and individuals to deploy, develop, train, manage, and use AI models (Lins et al., 2021). With the help of AIaaS, organizations of any size can realize AI-related benefits (Bughin et al., 2018; Pandl et al., 2021) and access scalable, on-demand AI capabilities. Despite the glaring opportunities of AIaaS for organizations, in reality, industry faces a "trough of disillusionment" with cloud-based AI services (Gartner, 2022), while B2C AI services such as ChatGPT are already celebrated as major milestones in the consumer market. Many heterogeneous AI services have evolved so that organizations are confronted with a complex and diverse system landscape comprising AI service platforms (e.g., to train and deploy AI models), generic AI technology services (e.g., general purpose text recognition), and also industry-specific services (e.g., chatbots for hospitals) (Geske et al., 2021; Lins et al., 2021; Sundberg & Holmström, 2022). Each AI service has unique characteristics requiring customers' attention when comparing and selecting offerings that fit their business problems.

Assessing fit and identifying a matching service is of great importance as service characteristics are highly interdependent, often implying trade-offs for customers' selection decisions and AIaaS providers' service design and development. For example, a provider can decide to design a very generic AI service to ensure its broad applicability for customers and to achieve economies of scale. At the same time, such a service might reduce suitability for customers that face industry-specific use cases and require a high prediction accuracy of AI services. Hence, there is no one-size-fits-all approach to providing AIaaS, but rather heterogenous services defined by interdependent characteristic manifestations. Deciding on certain characteristics of a service may result in an (unintentional) decision against other characteristics (e.g., high generalizability but reduced use case or industry specificity). To resolve such conflicting characteristics, providers and customers must make trade-off decisions based on suitability to particular use cases, industries, or groups of customers. Guidance on key characteristics, their interdependencies, and resulting trade-offs for cloud-based AI services is needed to ensure that, on the one hand, service developers are aware of potential pitfalls in the design phase. On the other hand, customers can ensure the applicability of AIaaS for their use cases. Nevertheless, given AIaaS' novelty and service complexity, it is challenging for providers and customers to oversee potential interdependencies, mitigate trade-offs, and make informed decisions.

AIaaS research is at an early stage, grounding on solid cloud research and ever-increasing AI research. Three AIaaS research streams are relevant when examining interdependent service characteristics and their resulting trade-offs: (1) Research focusing on the conceptualization of AIaaS, comparing prevalent service (types) and highlighting key service characteristics (e.g., Geske et al., 2021; Lins et al., 2021). (2) Research designing and implementing (cloud-based) AI services, thereby dealing with characteristic manifestations (e.g., Boag et al., 2018; Romero et al., 2021). (3) Research studying why and how customers use and adopt AIaaS (e.g., Pandl et al., 2021; Zapadka et al., 2020). While providing valuable contributions, prior research across all streams mainly focuses on one or a few specific characteristics of AI services with a limited view of interdependencies among them. When considering characteristics in isolation, we lack an understanding of how critical interdependencies impact proposed service designs and customers' decisions to use the service. To be aware of the possibilities and implications of designing and using AIaaS, practitioners and researchers require a broader and more detailed understanding of characteristics, interdependencies, and their trade-off manifestation and consequences. Without such an understanding, the risk of designing and using services that deviate from the targeted behavior due to overseen trade-offs increases. We, therefore, aim to bridge prior research streams and answer the following research questions (RQs):

RQ 1: What are the key trade-offs arising from AIaaS' interdependent characteristics?

RQ 2: What are the consequences for AIaaS design and selection?

To answer the RQs, we applied an inductive research approach and interviewed 39 AIaaS customers, providers, and consultants to gain insights into their perspectives on AIaaS. Overall, this research provides a comprehensive understanding and rich descriptions of key interdependent characteristics of AIaaS, resulting trade-offs, and their consequences for service design and selection.

Our insights contribute to the prevalent IS research streams on AIaaS. Examining interdependencies helps to resolve the conceptual ambiguity associated with the term "what is AIaaS" because we reveal the inner functioning of characteristics and their interplay. We expand prior IS research by examining characteristics of AIaaS across diverse use cases, organizations, and service offerings, enabling us to uncover novel trade-offs, centered on service generalizability, that are relevant to AIaaS design and key to understanding diverse service options. We provide rich descriptions of these trade-offs, including the conditions causing the trade-off, the resulting complex nature of service offerings, and consequences for providers and customers. Thereby, we extend our knowledge base on customers' sourcing decisions in the context of AIaaS and potential service design options for providers. For practice, we provide a starting point that can help providers to derive suitable AIaaS characteristic combinations and allow customers a deeper assessment regarding a service's fit for purpose and making suitable trade-off decisions.

The remainder of this paper proceeds as follows. We provide an introduction to AIaaS and related research. Then, we detail our research approach to identify AIaaS key characteristics and interdependencies for trade-off decisions. We also summarize the aggregated key characteristics. Afterwards, we describe the derived interdependencies between characteristics that lead to trade-offs. Lastly, we summarize our main findings and contributions, this study's limitations, and starting points for future research before concluding our research.

Background

AI as a Service

AIaaS combines AI capabilities and the typical cloud service models (Lins et al., 2021), which are known for "enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources […] that can be rapidly provisioned" (Mell & Grance, 2011, p. 2). In general, AI services can "perform cognitive functions that we associate with human minds such as perceiving, reasoning, learning, interacting with the environment, problem solving, decision-making, and even demonstrating creativity" (Rai et al., 2019, p. iii). Offerings range from analytics to conversational services, machine, deep learning, or neural networks as services and cocreation options for developing services (Sundberg & Holmström, 2022). The spectrum ranges from highly integrated ready-to-use services to (no-code) development platforms for model training and optimization (Geske et al., 2021).

In general, AIaaS emerge from the convergence of cloud service and AI technologies and therefore inherit typical cloud characteristics, such as on-demand self-service, scalability, and pay-as-you-go pricing models (Lins et al., 2021). Nevertheless, literature already discusses unique and innovative characteristics of AIaaS that enable organizations to use AI in their contexts effectively, including AIaaS' complexity abstraction, automation, and customizability features (e.g., Geske et al., 2021; Lins et al., 2021; Sundberg & Holmström, 2022). These unique characteristic of AIaaS help to address organizations' struggles to leverage AI to its full potential (Lorica & Nathan, 2019; Ransbotham et al., 2019; Zapadka et al., 2020), while avoiding in-house resource requirements and development efforts (Krogh, 2018; Newlands, 2021). For example, AI developer services guide its users through developing, deploying, and using data analytics models without having to learn complex algorithms or tools (Elshawi et al., 2018). Cloud-based AI services are, in particular, a promising alternative for organizations that struggle with the complex process of AI adoption (Dutta, 2018), high failure rates of AI implementations (Ransbotham et al., 2019), finding the right AI talent (Chui & Malhotra, 2018), lack of expertise for AI in-house development (Geske et al., 2021), high run costs or knowledge-intense deployment and configuration decisions (Yao et al., 2017). Overall, AIaaS can be a means to overcome these AI difficulties while allowing organizations to benefit from the vast opportunities that AI holds.

In this study, we particularly focus on AI software services that are ready-to-use AI applications and pre-trained models, which are AI models already trained by the AIaaS provider (or other parties) and then made available to users (Lins et al., 2021). Such pre-trained models are also referred to as inference as a service and are the most prominent and frequently used type of AIaaS. Using AI services, customers can pursue their core competencies without having to concern themselves with installation, maintenance, and technical deployment (Boag et al., 2018).

However, although organizations desire to overcome resource requirements by acquiring AI capabilities externally, there is no one-size-fits-all AIaaS. Organizations still require different resources to leverage and interact with the various AIaaS types (Geske et al., 2021; Halpern et al., 2019). For example, depending on the characteristics of the AIaaS, the service can offer customization options and room for customer adjustments. Still, these functionalities require customers to have development resources and invest more effort (Geske et al., 2021). In contrast, the desire to keep development efforts low comes with reduced customer freedoms, such as design and development options (Geske et al., 2021). This example shows that AI providers need to make trade-offs between offering highly integrated services or high customer freedom. AIaaS providers need to be aware of conflicting service characteristics that matter to their customers to make and communicate their trade-off design decisions and adequately manage customers' expectations. Hence, for AIaaS providers, clarity on possible trade-offs is crucial for AIaaS design, market positioning, and target customer identification. An analysis of interdependencies among key characteristics could also improve the quality of AIaaS by informing about potential pitfalls and synergies to look out for in the design phase. At the same time, customers need to know which AIaaS characteristics interfere with each other to avoid disappointments and sunk costs, and engage in trade-off discussions for informed purchasing decisions. Yet, practitioners' knowledge of trade-offs is still scarce, given AIaaS' novelty.

Related Research

Research on AIaaS is steadily increasing, covering diverse disciplines including information system (IS), computer science, and management. This study builds on and bridges three key research streams: AIaaS' conceptual foundations, design, and adoption (Table 2). We also harness extant research on cloud computing (e.g., defining service quality characteristics based on the well-known SERVQUAL measures; Zheng et al., 2014) and AI (e.g., discussing tensions between principles of AI; Jiang et al., 2022; Thiebes et al., 2021). Table 1 summarizes example cloud and AI research findings and their implications for this study.

Research Area	Example Research Findings and their Implications for this Study
Cloud Computing	- the model CLOUDQUAL comprising quality dimensions and metrics for general cloud services (Zheng et al., 2014), supporting us in rooting AIaaS characteristics in extant cloud research - adoption trade-offs related to weighting the benefits and risks of cloud services (e.g., Gashami et al., 2016; Schneider & Sunyaev, 2016), enabling us to dissociate AIaaS-specific trade-offs
Artificial Intelligence	- AI principles (e.g., trustworthy, accuracy, explainability; Smit et al., 2020; Thiebes et al., 2021), helping us compare AIaaS characteristics - the concept of situation Awareness for understanding human-AI interaction and resolving AI tensions (e.g., automation vs human agency; Jiang et al., 2022), guiding our study to identify means for resolving AIaaS trade-offs

Table 1. Example Related Cloud and AI Research

First, IS research focusing on the AIaaS' conceptual foundations studies the novel concept of AIaaS providing detailed conceptualizations and service classification to reduce conceptual ambiguity. For example, prior IS research already discusses key characteristics of AIaaS, such as inheriting common cloud (e.g., scalability, resource pooling, on-demand access) or AI characteristics (e.g., accuracy), but also highlights AIaaS' novel characteristics (e.g., complexity abstraction, integration, and automation) (Geske et al., 2021; Lins et al., 2021; Sundberg & Holmström, 2022). This research stream provides valuable contributions to determine unique characteristics for AIaaS but neglects to consider whether these characteristics are interdependent.

Related research focusing on AIaaS design deals with designing, implementing, and evaluating (cloud-based) AI services to provide solutions for specific use cases and harness the unique characteristics of AIaaS. For example, services are proposed to automatically generate AI model variants, helping AI developers to navigate the large model space and switch between differently optimized variants to meet their requirements best (Romero et al., 2021). Likewise, related research measures service implementations' qualities (e.g., availability, reliability, and fault tolerance) (Boag et al., 2018). It discusses first trade-offs, for instance, between latency and accuracy resulting from different service designs (Halpern et al., 2019). While informing this study by explicitly or implicitly arguing for interdependencies between characteristics, extant research in this stream studies only a small set of interdependent characteristics in isolation, mainly in context-specific use case scenarios, but neglects to elaborate on trade-offs' implications for service design and customers' usage decision.

Finally, practitioners and researchers are concerned with AIaaS adoption, its advancement, success factors, and hurdles. The third research stream has shown that an organization's characteristics (e.g., own training data), service qualities (e.g., complexity abstraction), and external market pressure (e.g., competitive pressures) impact customers' usage of AIaaS, while some customers tend to overestimate service qualities (Pandl et al., 2021; Philipp et al., 2021; Zapadka et al., 2020). This research stream provides valuable insights into customers' attitudes and opinions toward AIaaS but neglects to take the variety of AIaaS offerings into account and thereby falls short in explaining how prevalent interdependencies between characteristics impact customers' usage decisions.

While all literature streams provide valuable insights into the field of AIaaS, they do not consider trade-off decisions that guide AI providers' design and customers' selection decisions for AIaaS. We therefore lack a deeper understanding of interrelated characteristics, which are not only conflicting but may also be congruent, and of their impact on AIaaS design and usage. These trade-offs finally determine AIaaS design options that can be offered to organizations, those AIaaS customers hoping to acquire AI capabilities externally. Therefore, researchers and practitioners may oversee potential pitfalls based on the lack of knowledge of trade-off decision options. Organizations might also not be aware of the trade-off design decisions as root causes for disappointments and success with AIaaS. With our study, we aim to shed more light on how AIaaS characteristics interdepend and form design options based on trade-offs.

Research Stream	Description	Exemplary Studies
AIaaS' Conceptual Foundations	Conceptualizing AIaaS, embedding AIaaS in the research discourse.	- Geske et al., 2021 - Lins et al., 2021
AIaaS Design	Studying AIaaS design and implementation.	- Romero et al., 2021 - Boag et al., 2018
AIaaS Adoption	Examining AIaaS adoption factors and experience.	- Pandl et al., 2021 - Zapadka et al., 2020

Table 2. Related (IS) Research Streams on AIaaS

Research Approach

We chose an inductive approach to examine interrelations between AIaaS characteristics that result in trade-offs and imply consequences for service design and selection that literature has not yet explored. 39 interviews with AI providers, consultants, and customers were conducted to gain rich insights into industry stakeholders' views on AIaaS. Our aim to understand the interrelations by which AI characteristics impact AIaaS design and usage guided our decision to follow inductive methods (Corbin & Strauss, 2015). Inductive methods are a useful means to gain insights and an understanding of phenomena and their underlying processes to be able to create an abstract analytical schema (Creswell, 2013; Gasson & Waters, 2013). We particularly relied on coding techniques and supportive practices (i.e., memoing, constant comparison) suggested by the Straussian grounded theory methodology (Corbin & Strauss, 2015; Wiesche et al., 2017).

Data Collection

In order to gain insights into AIaaS characteristics' interrelations, we followed a purposeful sampling approach (Palinkas et al., 2015) and selected industry professionals from AIaaS provider, customer, and consultancy job roles. We included AIaaS providers (e.g., sales, development) in particular for their view on designing and positioning their AIaaS offerings. AIaaS customers were chosen to complement the view with customers' perception of AIaaS characteristics and their understanding for AIaaS design. AIaaS consultants were acquired given their expertise on both the service's and the customer's side. Data collection started in August 2020 and ended in September 2022. 39 experts were acquired, 15 of those from a provider's, 10 from a consultant's, and 14 from a customer's perspective. All interviewees live and work in Europe, mostly in Germany. Therefore, most interviews were conducted in the German language.

The AIaaS providers interviewed consisted of start-ups (40%) that provide innovative AI services (e.g., in health, finance, telco), enterprises of various sizes that offer AI services (33.3%, SMEs to software companies with more than 100,000 employees) and large cloud service providers (26.7%). Considering job roles, (senior) product management (33.3%), and technical roles such as solution architects (40%) and for start-ups founders and "c-level" roles (26.7%) were included.

AIaaS customer organizations interviewed cover a wide range of sizes ranging from smaller organizations of 300-5,000 employees (28.6%) to larger organizations with more than 250,000 employees (35.7%), including sizes in-between 12,000 employees up to more than 95,000 employees (35.7%). We cover a wide range of organizational sizes and representation across industries (e.g., insurance, retail, construction, pharma, travel and transportation, telco, and more). Customer job roles include department heads responsible for data and AI (35.7%), project leads and coordinators for AI-related projects (50%), as well as technical specialists such as senior architects (14.3%).

AIaaS consultants cover a wide range of consultancies, from small consultancies with 250 or 1000 employees, across medium-sized boutiques up to leading professional services firms with more than 200,000 employees. Interview partners included managers in the field of data analytics and AI (40%) as well as technical roles such as ML specialists and engineers (60%).

We conducted qualitative one-to-one interviews following the guidelines by Myers (2013). Interviews took 38 minutes on average, and 1,474 minutes altogether (excl. 1 interview based on written answers). All interviews were recorded, anonymized, and transcribed. A semi-structured interview method provided structure for interviewing the industry experts, while leaving room for aspects that were not or could not be considered while preparing the interview guide (Myers, 2013). We developed the interview guide by building on extant AIaaS research. Experts were asked for their view on AIaaS characteristics, congruent and conflicting ones, and their interdependencies, as well as for trade-off decisions that they see. The interview guide was constantly adjusted to consider novel findings and improvements. Non-judgmental form of listening (Walsham, 1995) was considered as well as maintaining distance (Patton, 2015).

Data Analysis

Our data analysis was started in parallel to data collection, directing our choice of questions based on emerging concepts, that are, AIaaS characteristics and interdependencies and resulting trade-offs (Abraham et al., 2013; Corbin & Strauss, 2015). This procedure enabled us to detect relationships between concepts (of characteristics) in an iterative process of constant comparison between initial data collected and analysis (Corbin & Strauss, 2015).

We applied a structured and iterative coding approach, applying open and axial coding so that data was analyzed with increasing abstraction level (Abraham et al., 2013; Corbin & Strauss, 2015; Wiesche et al., 2017). These coding techniques are widely adopted across IS research (Wiesche et al., 2017). During the coding process labels of text are assigned to interview transcripts based on the words mentioned by the interviewees. Atlas.ti as a software tool and manual annotations were used for coding the transcripts.

Open coding as the first coding stage encompasses fracturing the data according to concepts in the data that might describe relevant AIaaS characteristics, their interdependencies, and resulting trade-offs (Abraham et al., 2013; Corbin & Strauss, 2015). During open coding within the 434 pages of transcripts from the 39 interviews conducted, codes were assigned to 687 textual segments that were aggregated to 25 characteristics. For example, the passage *"it becomes commercially unviable when we use such amount of GPUs"* [interview partner (IP) 02, AIaaS provider, start-up founder] was coded as the AIaaS characteristic 'cost efficiency'. Further examples of the 25 characteristics are: 'improvability', 'particularity', 'trialability', or 'economies of scale'. All characteristics are listed in Table 3.

We then performed axial coding to uncover trade-offs between characteristics and aggregate our findings to achieve higher abstraction levels. We reviewed coded text segments for each characteristic to extract information about potential interdependencies. We followed Corbin and Strauss' (2015) suggestions to look for causes (i.e., conditions why AIaaS characteristics are interrelated), interactions (e.g., opposing forces spanning up the nature of the trade-off), and consequences (i.e., ramifications for service design and selection). For example, we revealed that a generic AI service can come with performance limitations (i.e., referring to interactions) and identified several reasons why AI services for the broad mass often achieve less accuracy (i.e., referring to causes). Hence, axial coding applied for our study allowed us to understand the causes why AIaaS characteristics are interrelated, the nature of the interdependency, and the actions that providers and customers take for dealing with it (e.g., customization freedom vs. ease of use).

In addition, axial coding enabled us to compare codes and classify them under more abstract themes that comprise related AIaaS characteristics (Abraham et al., 2013; Corbin & Strauss, 2015). By deriving trade-offs and elaborating on the key characteristics being interdependent, we were able to identify seven overarching characteristic themes, namely (1) ease of use, (2) generalizability, (3) cost advantage, (4) performance, (5) security, (6) reliability, and (7) ethical compliance, summarized in Table 3.

Our analysis was guided by constant comparison (Corbin & Strauss, 2015; Wiesche et al., 2017). First, we compared congruent and conflicting characteristics in the design of AIaaS, especially between the different stakeholder groups of AI providers and customers to triangulate our data. Second, while we did not specify theoretical perspectives prior to collecting and analyzing data (Abraham et al., 2013), we also compared coded interview data with related research. Building on our previous literature reviews on AIaaS (e.g., Lins et al., 2021; Pandl et al., 2021), we compared characteristics discussed in extant research with our coded characteristics to deepen our understanding of identified characteristics, align characteristic descriptions with extant research, and examine whether our interview findings confirm, extend, or oppose prior research. Table 3 summarizes this comparison.

To guide and structure our data analyses to uncover AIaaS trade-offs, we also looked for extant research examining similar trade-offs and thereby providing a theoretical foundation that our study can build on.

Related IS and management research particularly relied on the work by Smith and Lewis (2011) who argue that trade-offs can arise from contradictory yet interdependent elements leading to paradoxical tensions (e.g., characteristics seem inconsistent when juxtaposed) or dilemmas (e.g., competing alternatives pose clear advantages and disadvantages). We used their work as a theoretical lens for three key reasons: (1) Smith and Lewis (2011) describe different types of tensions (i.e., paradox, dualities, dilemma, dialectic) and differentiate their nature, helping us to better characterize and understand our interview findings; (2) they explain how such tensions emerge, thereby providing a theoretical foundation to explain our interview findings; and (3) describe means to manage tensions (e.g., splitting, integration, acceptance), helping us to identify ramifications for providers and customers. For example, we compared the different types of tensions with our emerging trade-offs, helping us in classifying the trade-off *"Generalizability vs. Specialization"* as a dilemma, that is, AIaaS providers have competing alternatives, each posing advantages and disadvantages when offering AI services (Smith & Lewis, 2011). In total, we were able to identify five key trade-offs for pre-trained AI models from the cloud, which we present in the following section in detail.

Characteristic Theme (#codes)	**Description**	**Assigned Characteristics**	**Example Related Research**
Ease of Use (#173)	The extent to which the AIaaS is intuitively understandable, easy to use, and leads to complexity reduction for the customer.	Ease of use (110), complexity abstraction (22), automation (19), trialability (13), modularity (9)	Geske et al., 2021; Lins et al., 2021; Pandl et al., 2021; Sundberg & Holmström, 2022
Generalizability (#151)	The degree to which an AIaaS is universally usable for various business problems. A low degree of generalizability means that an AIaaS readily incorporates particularities (e.g., industry specifics, details of a certain use case) or allows for customization and improvements to adjust for specifics.	Generalizability (59), customizability (46), particularity (33), improvability (13)	Geske et al., 2021; Lins et al., 2021
Cost Advantage (#121)	The extent to which the AIaaS meets the sweat spot for customer cost efficiency and provider profitability.	Cost efficiency (53), pricing (31), economies of scale (24), profitability (13)	Geske et al., 2021; Newlands, 2021; Pandl et al., 2021; Romero et al., 2021
Performance (#93)	The percentage of accurate and efficient output provided by the AIaaS to solve the business problem.	Accuracy (54), latency (27), scalability (9), performance (3)	Boag et al., 2018; Halpern et al., 2019; Romero et al., 2021; Zapadka et al., 2020
Security (#85)	The extent to which leveraging the AIaaS complies with organizations' security objectives (e.g., wrt. data or infrastructure).	Data protection (66), (IT) security (14), availability (5)	Pandl et al., 2021; Zheng et al., 2014
Reliability (#38)	The extent to which the AIaaS' results are reliable at the moment of deployment and during operation over time.	Reliability (22), AI Impermanence (16)	Myllyaho et al., 2021; Zheng et al., 2014
Ethical Compliance (#26)	The extent to which leveraging the AIaaS helps to achieve ethical, fair, and trustworthy AI (e.g., by providing explainability and transparency).	Explainability & transparency (20), ethical AI (4), fairness (2)	Jiang et al., 2022; Philipp et al., 2021; Thiebes et al., 2021

Table 3. AI as a Service Characteristics

AI as a Service Trade-Offs Impacting Service Design and Selection

Trade-off: Generalizability vs. Specialization (G vs. S). For service design, AIaaS providers have the option to either invest in generic AI models, or to invest in solving very specific business problems and tailored solutions.

<u>Conditions for trade-off occurrence and relevance:</u> We observed that many providers offer generic AI services that are universally usable for various use cases, such as, providing AI-based object detection, image

recognition, or translation services. AIaaS providers then share IT resources among customers to achieve the economies of scale needed for profitability. However, offering generic services comes with costs for AIaaS providers because it requires investments in research and development to ensure that the generic service performs well in different customer environments and industries, and on previously unknown data. Similarly, it is costly for providers to offer general explainability for AI behavior, especially if customers add further customizations.

Inversely, interviewees reported that AIaaS providers invest in solving very specific business problems and tailored solutions for certain customers. Offering a more tailored service requires industry or domain expertise and access to corresponding data, which is costlier for providers, but allows services to achieve higher performance (refer to trade-off G vs. P).

<u>Resulting options:</u> Decision between the costs of a generic service and the costs of a specific service.

<u>Trade-off consequences for providers:</u> AIaaS providers have the option to either keep AI models generic and achieve economies of scale, or to invest in solving very specific business problems and tailored solutions for certain customers. "*AI-as-a-Service should be a solution that addresses a very broad mass or solves a very specific problem very well"* [IP10, AIaaS provider, solution architect]. As the cost structure of AIaaS makes it difficult to perform well in a general and a very specific problem space, service providers' options to develop viable business models are either focused on providing generic or more specific services. In our interviews, we discovered six types of such business models.

When deciding to offer generic AIaaS, providers offer (1) a generic AI technology as a service (e.g., computer vision, natural language processing) or (2) develop a service that solves standard business problems (e.g., translation or document extraction services). For example, various providers offer chatbots that customers can easily deploy at their websites to manage end users' inquiries, while the chatbots' configuration relies on a no- or low-code approach. AIaaS providers rely on generic data models to stay industry agnostic and cost-efficient, such as an AI service that recognizes the handwriting of letters that remain the same across industries and use cases. If possible, providers can also (3) abstract services for a specific industry, thus, customizing AI models to meet the industry-specifics, but still offering a mass product.

If providers decide to offer generic services, they cannot account for various specific customer set-ups (e.g., data formats, systems, region, or market differences) because it would incur additional costs, diminishing profitability achievements from scaling their service across a wide set of customers and industries (refer to trade-off G vs. P). Generic service providers' profit margin is limited at the point where, for example, hiring a human to perform the task manually is cheaper than adopting the AIaaS. Providers' options to balance service prices range therefore between the lowest operating cost possible and the maximum that they can charge compared to human labor.

When deciding to offer tailored AIaaS, providers can focus on (4) specific niche use cases offering rich customization options, or (5) differentiating themselves by providing integrated solutions (e.g., covering all tasks of daily work of a radiologist). To meet customers' demand for customization and performance improvements, several providers have started to (6) customize generic service offerings for a group of customers, thus becoming a partner of generic service providers or a market intermediary; ultimately leading to entangled cloud supply chains. "*I can well imagine that there is also a large space and a large market for special solutions or specialized chatbots*" [IP20, AIaaS customer, lead for digital].

If providers decide to focus on tailoring their service (e.g., to a specific industry segment), providers require domain expertise and additional training data, among other efforts, thereby increasing service development costs. In addition, tailored AI models are often more complex and therefore computationally intensive when providing inferences, which can reduce operational performance and increase costs. Depending on the use case, tailored AI services may further require solutions to reduce the latency of computationally intensive predictions, adding additional costs. Providers' options are guided by the assumption of whether the customer segment is promising enough to compensate costs for tailored services.

Trade-off: Generalizability vs. Cost Efficiency (G vs. CE). Customers have the option to decide whether to use cost-efficient generic services that bear the risk of hidden costs, or to use tailored services that are costlier but induce less configuration and adjustment costs.

<u>Conditions for trade-off occurrence and relevance:</u> AIaaS are characterized by the commonly known pay-per-use pricing model for cloud services. Customers may rely on cost-efficient generic services to experiment with AI and quickly deploy first proof-of-concepts. However, customers may face hidden costs for integration, configuration, and data preparation not included in the pay-per-use pricing model when consuming generic AIaaS. "*How well is [the AI service] already adapted to me, and how much cost and time do I have to put into it? Because the worse the basic product is adapted to me, the more work I have to do.*" [IP16, AIaaS customer, project lead].

In contrast, tailored services rely on costlier pay-per-use pricing models but induce less customization and integration costs for customers because tailored services are designed to fit customers' industry and set-up requirements.

<u>Resulting options:</u> Decision between generic services with hidden costs and cost-intensive tailored services with high specificity.

<u>Trade-off consequences for customers:</u> Considering the diverse set of AIaaS business models, customers first require awareness of the diverging service options and their implications (as outlined in the trade-offs), and second, select a viable service offering that meets their business requirements: either using a generic or a more specific service.

In general, AIaaS come with a cost-beneficial pricing model enabling infrastructure resources payment as needed and covering peaks through cloud scalability. AI services require less capacity planning for organizations and offer a degree of capacity that would usually not be possible on-premise. Possessing a comparable number of processing power in-house for model training and prediction would not be economically viable for many customers, so that AIaaS makes AI affordable for many organizations. Without the need to host and manage their own AI-optimized infrastructure, customers can also ensure focus on their core business activities. *"It's incredibly easy to build a chatbot using [a provider's] services."* [IP21, AIaaS customer, department lead].

Nevertheless, for generic AIaaS, interviewees reported costs for integration and configuration as well as obtaining and funding external resources that perform these tasks on customers' behalf if needed (e.g., consultancy). Such costs depend on the gap between customers' internal set-up and the generic services. For example, customers may need to prepare internal data to meet the service's generic format or organize data so that it can be accessed on the cloud infrastructure of the AIaaS provider. Covering and calculating such costs can be challenging for customers as they are not included in the service pricing model and customers lack experience in costs for training and using AI services. In addition, costs are less predictable because model complexity may vary over time, leading to higher costs for making inferences or storing data in the cloud. Consequently, generic services may be cost-efficient at first sight but can come with hidden costs that need consideration when evaluating options. In contrast, tailored services are generally more expensive because they cannot harness economies of scale, but they usually require fewer adaptions and configurations that are costly for customers.

Trade-off: Generalizability vs. Performance (G vs. P). Using generic AI services can come with performance limitations.

<u>Conditions for trade-off occurrence and relevance:</u> While generic AI services are less costly due to economies of scale, they also come with limited performance for specific business problems – a drawback that has been frequently not known in advance or at least underestimated by interviewed customers. Generic models suffer from limited prediction accuracy for deviating customer data in particular. *"It worked to create quick results in many cases, but it often wasn't enough for us."* [IP23, AIaaS customer, chief digital officer].

On the one hand, generic services achieve moderate prediction accuracy for very general tasks and business problems (e.g., translations or image recognition) because cloud providers have access to a large amount of training data and high technical (AI) expertise. *"We very often experience that the quality of our services is also better because we have gained a lot of experience from a lot of customers from a lot of countries."* [IP13, AIaaS provider, solution architect]. AIaaS providers also benefit from customers' feedback, lessons learned over time, and technical improvements, enabling them to achieve steady service performance improvements, as witnessed, for example, in the case of translation services.

On the other hand, interviews revealed that AI services for the broad mass often achieve less accuracy for specific customer use cases when compared to custom AI solutions (i.e., own models based on customer data). First, generic services might include aspects that are not relevant to customers' business problems and miss parts that would be relevant. For example, a generic AI service may fail to recognize doctors' handwriting accurately or interpret industry-specific forms. Second, generic services may deviate from customers' data and insights required. For example, a generic service may be inadequate for processing specific customer data and attributes sensitive to bias or discrimination. Third, a generic service may not meet the performance requirements of sensitive business cases, such as when customers are operating in a safety-critical industry like automotive engineering. Fourth, some interviewees also reported that generic services may lack configuration and customization options, hampering customers from achieving performance improvements.

Conversely, tailored AI services provide higher accuracy needed for the business problem because they focus on industries or specific customer segments and thus meet the specific customer set-up typical for the target industry or profession.

Resulting options: Decision between generic services with moderate to high performance (depending on the use case) and tailored services with high performance.

Trade-off consequences for customers: Customers require awareness that generic services often cannot (fully) resolve their business problem but require adjustments to achieve high levels of performance. Given the trade-off, customers have the option to decide to either (1) use a generic service with moderate performance, (2) customize a generic service, or (3) purchase a tailored AI service.

First, customers need clarity on which performance level is required for their specific business problem. For example, an interviewee reported that they integrated a generic chatbot providing customer support. Even though the chatbot's prediction accuracy was relatively low, the company still gained several benefits from embedding a generic chatbot, including 24/7 customer support, resource cost savings, and increased end user satisfaction. Given AIaaS's trialability, customers can also easily and quickly run a proof-of-concept using a generic service and then assess whether harnessing AI seems in general promising or not. Another option to solve the business problem with a generic AI service despite relatively low performance is to leverage the service to save time on simple business tasks and have a human in the loop for validation.

If customers decide that the generic services' performance is not sufficient to solve their business problem, they may have the option to adapt and customize the service. For example, a generic AI service may be unable to process customers' tickets because ticketing systems sometimes are highly individualized and contain product-related information. Customers can then calculate whether they invest to adjust and customize the service (refer to trade-off G vs. E). In the AI context, success and accuracy are not always foreseeable with certainty, so that customization comes with risks for customers. If accuracy targets are not met, service integration investments are lost. Also, customization requires that customers have access to data that can be used to retrain generic models. Internal or external AI knowledge may similarly be needed for configuration especially for sensitive use cases that need an awareness for, among others, distributions or protected attributes leading to bias or discrimination.

In case customization and adaption options are not sufficient for solving the business problem because they cannot reach the targeted performance either technically or economically, customers can opt for tailored AI services right from the start.

Trade-off consequences for providers: Given the performance differences and responsibilities in generic vs. tailored AIaaS, providers have the option to issue clear service descriptions to manage customers' expectations, including information on usage requirements, service limits, and performance. Particularly, relevant in providers' communication is which aspects of service performance are dependent on customers (i.e., their data quality or use case specificity). This also includes open and honest communication of potential uncertainties and pitfalls to make customers aware of the nature of AI that can make it challenging to foresee pre-trained models' success on customer data and use cases. Otherwise, it may result in disappointing customers' expectations up to a canceled project in extreme cases. Metrics such as accuracy, precision, and error rates provide an orientation, but need context and explanation to be interpreted correctly by customers, hence requiring customer education.

(Limited) trialability periods that enable customers to (quickly) conduct proof-of-concepts and evaluate whether the service meets their expectations can help as an orientation to evaluate performance fit. However, such proof-of-concept trials require information with regard to expectations and customers' efforts for transferal from proof-of-concept environments and use cases (often showcases) to the organization in production environments (e.g., skill requirements, data preparation efforts etc.). Otherwise, unexpectedly deviating performance when compared to proof-of-concept cocreations with providers raises questions for customers. Based on our interviews it is important that providers highlight risks and uncertainties that could lead to missed performance targets in the interest of long-term customer satisfaction and to protect the overall reputation of the AIaaS and its provider. Otherwise, customers may conclude that AIaaS are lacking technology or service capabilities when performance targets are missed.

In case providers offer generic solutions, as far as profitability allows, our interviews show a demand to invest in ease of use, allowing for simple configuration and customization options, and reduced upfront efforts for data preparation (refer to trade-off G vs. E).

Trade-off: Generalizability vs. Ease of Use (G vs. E). Generic AI services often come with a need for adaptions and customizations to solve specific business problems, thereby diminishing ease of use.

<u>Conditions for trade-off occurrence and relevance:</u> The extent to how easy customers can use offered AI services varies greatly for each individual service. Interviewees reported that services may provide APIs or user interfaces, which may even offer drag-and-drop functionalities (e.g., for Excel files), to provide inferences on pre-trained models. In addition, customers' perception of what ease of use means varies significantly: some consider an AI service easy to use if they can easily make adjustments, tailor it, and understand exactly how it works. Others do not want to have this level of detail but prefer rather simple access to use the service successfully. These conflicting views also illustrate the diverse user groups of AIaaS, ranging from data scientists to employees of business units.

More tailored AI services typically exhibit a high degree of ease of use for organizations because they are already customized to meet business specifics (e.g., business vocabulary). Also, many generic models were perceived as easy to use, particularly if they fit customers' use cases. Interviewees reported that high ease of use can compensate for relatively low service performance as long as there is business value.

Nevertheless, interviewees are concerned that generic services require adaptions, customizations, and creativity to address the customers' business problems. "*Make or buy, I think it is a combination of both. Not just the buy because we have to customize*" [IP15, AIaaS customer, senior architect digitalization]. For example, many generic AI services come with individualization levels or developer options to offer a degree of specialization and allow for necessary adjustments or performance improvements. Customers have the option to make investments to adjust and tailor the system for success and accuracy. Holding this balance is particularly difficult in the context of AIaaS because performance success is not always foreseeable with certainty, given AI's experimental nature of retraining, trial and error, and statistics. Hence, generic services can come with customization requirements that conflict with customers' wishes for ease of use.

<u>Resulting options:</u> Decision between the probably limited ease of use of a generic service offering compared to the ease of use associated with highly tailored services.

<u>Trade-off consequences for providers:</u> First of all, providers require clarity on the target user group in order to communicate services' out-of-the-box functionalities and adjustment options in a clear manner to manage customers' expectations. AIaaS providers have the option to enable customers to customize and adjust their services. However, concerns were raised that customer enablement comes with providing insights into how providers' services are working, bearing the risk for providers of losing intellectual property. "*We have a service that [...] can be customized to your own case. [...] We explain quite precisely and we also write whitepapers on how we train this model. But I don't think we explain [...] how weights are applied or even provide the model. Of course, that would also be a competitive advantage to a certain extent, which we would be giving out of our hands.*" [IP10, AIaaS provider, solution architect]. Interviewees reported that services may enable service and model configurations, further training, or re-training with customers' data, or even applying transfer learning, referring to a method in which a model and associated data developed for a particular task are used as a building block to solve a different problem (Samreen et al., 2022). Another profit-driven option can be to offer data preparation tools to support customers in making inferences easier because most generic services still require customers to prepare and adjust their data for making inferences.

Accounting for easy customization options and additions to the service not only increases AIaaS' production costs and efforts, potentially lowering profitability or increasing the service price. But providers' level of customer support and consultancy efforts can also threaten their business profitability. This is particularly relevant for generic AI services requiring a lot of data science expertise to leverage, customize, and integrate services tailored to a very specific business problem. AIaaS providers may tend to avoid building up consultancy resources themselves but rather engage in (strategic) partnerships so that partners take over implementing customization requirements at the customer site.

<u>Trade-off consequences for customers:</u> Pre-trained services are of great value for customers for a heads-start to AI use. But it was reported that it is challenging to fulfill the customers' wishes to easily use AIaaS without the necessary technical expertise, given the burden of customization. Enabling customers to perform the necessary customizations can be challenging because customers may lack the expertise or data. Customers that do not have the skill to perform configuration or customization activities do not profit from customizable AIaaS unless they hire internal or external resources. Hiring (internal) resources can be challenging due to the scarcity of AI talent, which is among the points that customers try to overcome when purchasing AIaaS rather than in-house development. Indeed, customization's resource requirements can undermine customers' cost-efficiency arguments for opting for AIaaS. Interviewees claimed that as long as experts are missing, they would rather consider AI services without customization as easy to use, acknowledging that generic services require adjustments for high performance. Similarly, customers need to have access to high-quality data for adjustment training, achieving better business problem fit. Access to such data in the format needed may not be feasible for many customers. Some interviewees reported frustrating projects when services did not meet accuracy targets and cost savings were not achieved.

Still, vast customization options allow customers to benefit from generic services and tailor them to a specific use case that would otherwise not be profitable for the provider and thus not offered (i.e., due to a small customer segment). Another option for customers is to purchase a readily tailored service, coming with a higher ease of use, but also higher costs due to lack of economies of scale. Hence, customers need to thoroughly analyze their in-house capabilities (e.g., knowledge and skills, resource capacities), and access to and knowledge about data enabling customizations, before making trade-off decisions.

Trade-off: Generalizability vs. Security (G vs. Sec). Generic AI services are designed to be applied across organizations and industries raising concerns regarding the protection of the data processed.

<u>Conditions for trade-off occurrence and relevance:</u> Apart from general data protection concerns about sharing data with cloud providers for storage, AI services presuppose data processing. Ensuring security and data protection is more critical in AI contexts given the amount of data used to make inferences or customizations (e.g., retraining), and risks of human failures. For example, one interviewee mentioned that they integrated a generic chatbot but end users provided sensitive information to this bot (e.g., assurance identifiers or e-mail addresses), leading to unforeseen security risks.

Interviewees reported that generic services are typically offered by large providers (e.g., hyperscalers) that provide high data protection and AIaaS development standards compared to in-house alternatives. However, generic pre-trained AI models are often used, trained, and improved across customer organizations. On the one hand, models can then be more elaborate than an in-house development of a single organization, demonstrating an advantage as AIaaS providers can improve and train models on this set of data. On the other hand, it often remains unclear for customers to what extent and how safely data flowing into the model is anonymized and what exactly the model takes in. "*Does my data now end up in there [the AI service] making the platform product smarter for everyone? Yes, I have to think about that relatively carefully*" [IP23, AIaaS customer, chief data officer]. Some customers particularly fear to support competing organizations using the same service via model training based on their data. Therefore, interviewees stated that they are frequently confronted with the decision between the advantages of generic AI services (e.g., model quality, economies of scale) and potential data protection risks, or at least, protection uncertainty.

In regard to more tailored services, customers acknowledge that services offered by startups and small and medium-sized enterprises, either differentiate themselves by focusing on data protection aspects or are perceived as less elaborate on protection and security features compared to larger providers, but more suitable to their use cases from a performance perspective.

Resulting options: Decision between advantages of generic services and potential data protection risks. If using tailored services, need to assess data protection levels of startups and small and medium-sized enterprises.

Trade-off consequences for providers: The high data protection and development standard that AI providers offer compared to customers' own infrastructure and in-house development is not always conveyed successfully to customers. This includes communication of providers' options regarding the security offered by the elaborate design of a professional AI service and the potential vulnerabilities of in-house developments. Providers can undergo third-party assessments to prove security and data protection, including traditional security and data protection certifications (e.g., ISO/IEC 27001) and novel AI quality seals (e.g., ensuring transparency and fairness). However, our interviews have shown that this cannot always mitigate a perceived lack of trust. AIaaS providers further may react to concerns by implementing protective measures, such as anonymization. Nonetheless, these measures can significantly limit processing options, accuracy, and possible service offerings. Also, if the provider cannot access the data, supporting customers with their AIaaS and potential issues is more complicated. Finally, customers are prone to making (security-related) mistakes, such as accidentally sharing data publicly (e.g., if settings are set or adjusted by customers that lack skill), so that providers should decide on options to adjust settings from the user interface that could infringe data protection, following the principles of privacy by default and design.

Trade-off consequences for customers: Some interviewees are afraid of losing control over their data when using AIaaS, which is still common depending on customers' regions. Interviewees face issues that typical protection measures such as encryption are not necessarily fully applicable to AI as it limits data processing options and thereby contradicts the value of AIaaS. Interviewees emphasized that using AIaaS may require a cultural change in organizations, becoming more open to use cloud services, thereby reducing the number of customers that may not opt for using AI services. "*Use common sense within the company and think about which data is so important to me that I don't want to take any risks*" [IP23, AIaaS customer, chief data officer]. They also emphasize that if they decide for own developments, they are confronted with maintenance issues and ease of use and security considerations that are otherwise covered by generic AI services. Relying on third-party assessments and carefully studying data processing agreements are typical remedies to mitigate concerns, as proposed by the interviewees. Customers may opt for additional data protection measures at an additional cost. Another option may be deciding to use AI developer service platforms for model training and keep data stored on own infrastructure for operation or train the model in the cloud and then deploying it in-house (e.g., on the edge).

Discussion

Principal Findings

We conducted a qualitative interview study to deepen our understanding of AIaaS characteristics, their interdependencies, and resulting trade-off manifestation. We identified 25 characteristics relevant when designing and adopting AIaaS, covering seven overarching characteristic themes, including AI services' ease of use, generalizability, cost advantage, performance, security, reliability, and ethical compliance. By examining interdependencies between these characteristics, we were able to uncover five trade-offs, their conditions, resulting options, and consequences for provider design decisions and customer service selection.

Notably, our research findings highlight the centrality of AI services' generalizability because it impacts many manifestations of trade-offs and determines the viability of service offerings. The importance of service generalizability originates from the driving force behind cloud-based AI services. Namely, offering (pre-trained) AI models that are easy to use by organizations to leverage AI capabilities. Interestingly, customers reported that trade-off manifestations related to generalizability have been frequently not known in advance or at least underestimated by interviewed customers. Due to this lack of understanding of trade-offs, customers purchased generic services and invested in integration and configuration, only to discover that a tailored solution would have been needed instead. Interviewees strongly emphasized that there are AIaaS providers dominating in niche applications for very specific use cases where larger cloud providers currently cannot compete due to the limitations stemming from pursuing a generic strategic approach to AIaaS. Currently, many AIaaS offerings are evaluated in "light-house" projects and proof of concepts that are not yet fully in production, come with pro bono presales consulting services, or do not have to be profitable yet. Interviewees confirmed that increased clarity on the generalizability trade-offs – as uncovered in this research – would help to understand the value proposition of generic services vs. specific services to calculate the business case more adequately and select AIaaS guided by realistic expectations of economic viability.

Nevertheless, our findings reveal that elaborating on trade-off manifestations require often a basic understanding from customers about AI and about corresponding performance metrics in particular to be able to estimate a (generic) service's benefits and performance. However, this circumstance contradicts the core promise of AIaaS, offering services for the broad mass. A robust requirement analysis by customers, complemented by clear and honest expectation management from providers or consultancies before service usage, and a critical internal assessment by customers when using the service are required to prevent customers' failures. "*In principle, the barrier is more about finding out afterwards how far this service and these platform services will take me, and where they may no longer help me. And that's where I need the expertise […]. That's not a problem of the service, in fact at that point, it's simply that the deeper I go into it, the more expertise I have to bring from in-house*" [IP23, AIaaS customer, chief digital officer].

Finally, we note that identified trade-offs are highly interdependent, leading to cascading effects once a service design or selection decision is made. For example, a service provider may embed customization and configuration functionalities to mitigate the trade-off consequences of a generic service that lacks high performance. Adding such functionalities may improve performance but puts burden on customers that require more skills to make use of them effectively; ultimately hampering the service's ease of use. Our findings thereby confirm the theoretical view that characteristics being oppositional to one another yet are also synergistic and interrelated within a larger system (Smith & Lewis, 2011). These entangled service characteristics further emphasize that there is no one-size-fits-all AI service requiring customers to make well-informed selection decisions.

Implications for Research

Our study has implications for the three key research streams of AIaaS, rooted in the IS and computer science disciplines: AIaaS' conceptual foundations (1), design (2) and adoption (3). First, we complement existing IS research on AIaaS' foundations that highlights key characteristics of AIaaS (e.g., Geske et al., 2021; Lins et al., 2021; Sundberg & Holmström, 2022) by examining whether these characteristics are interdependent. Taking a deep dive into these interdependencies between characteristics enabled us to uncover trade-offs and explain how they emerge. Our considerations on interdependencies also help to resolve the ambiguity associated with the term "what is AIaaS" because we reveal the inner functioning of characteristics and their interplay. In contrast to IS research that emphasizes key characteristics of AIaaS, we discuss viable service variants of AIaaS in detail based on the complex network of interdependent characteristics, thereby guiding future research when examining cloud-based AI services.

Second, existing AIaaS design literature and related computer science research provides valuable information on specific AIaaS implementations (e.g., Boag et al., 2018; Romero et al., 2021). Nevertheless, prior research has examined characteristics and potential interdependencies in isolation to implement a specific AI service. We expand this research by examining characteristics of AIaaS across diverse use cases, organizations, and service offerings, enabling us to reveal different service design options, while discussing the trade-offs. Our study also provides further evidence for selected design trade-offs mentioned in prior research, for instance, between latency and accuracy resulting from different service designs (Halpern et al., 2019), by constantly comparing our findings to the literature. More importantly, we extend the AIaaS design literature stream by uncovering novel trade-offs, centered on service generalizability, that are relevant to AIaaS development and key to understanding AIaaS' design success. We provide rich descriptions of these trade-offs, including the conditions causing the trade-off, resulting service options for providers, and consequences for service design. Taking a theoretical lens on tensions (Smith & Lewis, 2011) also helped us to reveal the nature of design trade-offs, mostly pertaining to dilemmas (e.g., competing alternatives pose clear advantages and disadvantages for providers). Researchers can refer to and build on our rich descriptions when designing, developing and accessing cloud-based AI services, making informed decisions.

Finally, we extend IS research examining AIaaS adoption (e.g., Pandl et al., 2021; Philipp et al., 2021; Zapadka et al., 2020) by providing rich descriptions of the trade-offs, and more importantly, their consequences on customers' selection. Our study reveals that trade-offs challenge existing knowledge on AIaaS adoption because neglecting interdependencies between characteristics may mislead customers when making selection decisions. Particularly, our interviewees reported examples for selection pitfalls stemming from trade-offs, such as not knowing in advance that generic services require customization to achieve adequate performance. This study provides deep insights into the complex nature of AI provisioned as a service and how this nature impacts customers' balancing of trade-off manifestations, extending our knowledge base on customers' sourcing decisions in the context of AIaaS. Our findings emphasize that IS research should not only examine AIaaS characteristics and their impact on customers' service adoption intentions and behavior, but also control for potential interdependencies between these characteristics and confounding effects.

Implications for Practice

Our results also inform practice. We support customers to make more informed purchasing decisions on AIaaS by highlighting key characteristics of AIaaS and explaining how their interdependencies lead to the emergence of trade-offs (e.g., informing about tailored and generic services, associated characteristics, costs for configuration and benefits). Notably, even though customers wish that AIaaS work plug and play, our findings reveal that they often need support for preparation and configuration (e.g., how to bring data in the right format, adjust to setup). Illustrating different design options based on trade-offs enables customers to understand the AIaaS landscape and differences in offerings, contributing to the awareness that based on the nature of cloud-based AI services there is no one-size fits all solution. We show that customers need to make trade-off decisions, hence, choosing one pole of a trade-off that comes with advantages but also disadvantages for the customer. By supporting customers in making well-informed sourcing decisions, we strive to accelerate the usage of AIaaS which can overcome prevalent AI challenges (e.g., lack of AI resources) and realize cloud-related benefits (e.g., cost savings, scalability). Due to the importance of AI to customers' businesses and the wide-ranging consequences of initial adoption decisions, AIaaS providers become a strategically relevant partner for customers' future planning. Looking ahead, customers should prepare strategies to deal with vendor lock-in and price dependencies (e.g., multi-cloud); considerations well-known from general cloud computing research.

Our study also provides rich insights for AIaaS providers by illustrating ways to improve services, their marketing and positioning, as well as design decision-making and communication (e.g., highlighting design decision drawbacks and pitfalls due to interdependencies). Trade-offs related to generalizability require providers to determine their customer segment carefully and then identify potential users of their services (e.g., employees from business units, data scientists, or machine learning engineers) to meet customers' expectations about performance, ease of use, and customizability. For future perspectives, the AIaaS market is highly dynamic so that further options for providers are lying ahead, as developments among ChatGPT and GPT-4 are showing. AIaaS providers need to identify viable business models that consider our proposed trade-offs and will sustain in the long term.

Limitations and Future Research

Our study is subject to limitations, paving the way for future research. First, given the number and depth of interviews we conducted, our study has limitations concerning the generalizability of our findings. Our interviewee sample is relatively diverse regarding the interviewees' backgrounds, ranging from AIaaS providers to consultancies and AIaaS customers. However, at the same time, it is homogenous given that all interviewees live and work in Europe, mostly in Germany. Conducting further interviews in different countries to increase the generalizability of findings is recommended. Besides, our study may be subject to interpretation and selection biases due to the ambiguity of language (Myers & Newman, 2007), or the involvement of the interviewers' perspective when constructing knowledge (Myers & Newman, 2007). We noticed that key characteristics and resulting interdependencies and trade-off manifestations were repeatedly found across interviewees. However, it could be possible that examining further AI services may provide additional information on trade-offs that have not yet been revealed. The interviewees may have found it difficult to verbalize some challenges related to AIaaS trade-offs due to AIaaS' conceptual ambiguity and novelty. Future work can complement our research approach using quantitative methods to test our propositions and triangulate our findings.

Conclusion

AIaaS trade-offs can only be considered during AIaaS design and selection if interdependencies between characteristics are understood. This qualitative interview study provides an overview along with explanations on how key characteristics of AIaaS are interdependent and lead to trade-offs that help to understand how design options are determined. Thereby, we contribute to an understanding of AIaaS along with implications for providers and customers associated with characteristic prioritization and trade-off decisions. Our results indicate that examining AIaaS in a differentiated manner rather than presuming a one-size-fits-all quick path to overcoming issues with AI adoption offers the chance for customer organizations to fulfill their goals and business cases while providers can avoid disappointments and frustration with their offerings.

References

Abraham, C., Boudreau, M.-C., Junglas, I., & Watson, R. (2013). Enriching our theoretical repertoire: the role of evolutionary psychology in technology acceptance. *European Journal of Information Systems*, *22*(1), 56–75.

Boag, S., Dube, P., El Maghraoui, K., Herta, B., Hummer, W., Jayaram, K. R., Khalaf, R., Muthusamy, V., Kalantar, M., & Verma, A. (2018). Dependability in a Multi-tenant Multi-framework Deep Learning as-a-Service Platform. In *Proceedings of the 48th Annual IEEE/IFIP International Conference on Dependable Systems and Networks Workshops (DSN-W),* Luxembourg.

Bughin, J., Seong, J., Manyika, J., Chui, M., & Joshi, R. (2018). *Notes from the AI frontier: Modeling the impact of AI on the world economy*. https://www.mckinsey.com/featured-insights/artificial-intelligence/notes-from-the-ai-frontier-modeling-the-impact-of-ai-on-the-world-economy#/

Chui, M., & Malhotra, S. (2018). *AI adoption advances, but foundational barriers remain*. https://www.mckinsey.com/featured-insights/artificial-intelligence/ai-adoption-advances-but-foundational-barriers-remain#/

Corbin, J. M., & Strauss, A. L. (2015). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (4th ed.). SAGE Publications, Inc.

Creswell, J. W. (2013). *Qualitative inquiry and research design: Choosing among five approaches* (3rd ed.). SAGE Publications, Inc.

Dutta, S. (2018). An overview on the evolution and adoption of deep learning applications used in the industry. *WIREs Data Mining and Knowledge Discovery*, *8*(4), 1–12.

Elshawi, R., Sakr, S., Talia, D., & Trunfio, P. (2018). Big Data Systems Meet Machine Learning Challenges: Towards Big Data Science as a Service. *Big Data Research*, *14*, 1–11.

Gartner (Ed.). (2022). *What's New in Artificial Intelligence from the 2022 Gartner Hype Cycle*. https://www.gartner.com/en/articles/what-s-new-in-artificial-intelligence-from-the-2022-gartner-hype-cycle

Gashami, J. P. G., Chang, Y., Rho, J. J., & Park, M.-C. (2016). Privacy concerns and benefits in SaaS adoption by individual users. *Information Development*, *32*(4), 837–852.

Gasson, S., & Waters, J. (2013). Using a grounded theory approach to study online collaboration behaviors. *European Journal of Information Systems*, *22*(1), 95–118.

Geske, F., Hofmann, P., Lämmermann, L., & Schlatt, V.M Urbach, N. (2021). Gateways to Artificial Intelligence: Developing a Taxonomy for AI Service Platforms. In *Proceedings of the Twenty-Ninth European Conference on Information Systems (ECIS),* Online Conference.

Halpern, M., Boroujerdian, B., Mummert, T., Duesterwald, E., & Janapa Reddi, V. (2019). One Size Does Not Fit All: Quantifying and Exposing the Accuracy-Latency Trade-Off in Machine Learning Cloud Service APIs via Tolerance Tiers. In *Proceedings of the 2019 IEEE International Symposium on Performance Analysis of Systems and Software (ISPASS),* Madison, WI, USA.

Jiang, J., Karran, A. J., Coursaris, C. K., Léger, P.-M., & Beringer, J. (2022). A Situation Awareness Perspective on Human-AI Interaction: Tensions and Opportunities. *International Journal of Human–Computer Interaction*, 1–18.

Krogh, G. von (2018). Artificial Intelligence in Organizations: New Opportunities for Phenomenon-Based Theorizing. *Academy of Management Discoveries*, *4*(4), 404–409.

Lins, S., Pandl, K. D., Teigeler, H., Thiebes, S., Bayer, C., & Sunyaev, A. (2021). Artificial Intelligence as a Service. *Business & Information Systems Engineering*, *63*(4), 441–456.

Lorica, B., & Nathan, P. (2019). *AI Adoption in the Enterprise* (1st ed.). O'Reilly Media, Inc.
Mell, P., & Grance, T. (2011). *The NIST definition of cloud computing*. https://doi.org/10.6028/NIST.SP.800-145
Myers, M. D. (2013). *Qualitative research in business & management* (2nd ed.). SAGE Publications.
Myers, M. D., & Newman, M. (2007). The qualitative interview in IS research: Examining the craft. *Information and Organization*, *17*(1), 2–26.
Myllyaho, L., Raatikainen, M., Männistö, T., Mikkonen, T., & Nurminen, J. K. (2021). Systematic literature review of validation methods for AI systems. *Journal of Systems and Software*, *181*, 1–22.
Newlands, G. (2021). Lifting the curtain: Strategic visibility of human labour in AI-as-a-Service. *Big Data & Society*, *8*(1), 1-14.
Palinkas, L. A., Horwitz, S. M., Green, C. A., Wisdom, J. P., Duan, N., & Hoagwood, K. (2015). Purposeful Sampling for Qualitative Data Collection and Analysis in Mixed Method Implementation Research. *Administration and Policy in Mental Health*, *42*(5), 533–544.
Pandl, K. D., Teigeler, H., Lins, S., Thiebes, S., & Sunyaev, A. (2021). Drivers and Inhibitors for Organizations' Intention to Adopt Artificial Intelligence as a Service. In *Proceedings of the 54th Hawaii International Conference on System Sciences (HICSS),* Online Conference.
Patton, M. Q. (2015). *Qualitative research & evaluation methods: Integrating theory and practice* (4th ed.). SAGE.
Philipp, R., Mladenow, A., Strauss, C., & Völz, A. (2021). Revealing Challenges within the Application of Machine Learning Services-A Delphi Study. *Journal of Data Intelligence*, *2*(1), 1–29.
Rai, A., Constantinides, P., & Sarker, S. (2019). Next-Generation Digital Platforms. *Management Information Systems Quarterly*, *43*(1), iii–ix.
Ransbotham, S., Khodabandeh, S., Fehling, R., LaFountain, B., & Kiron, D. (2019). Winning With AI: Pioneers Combine Strategy, Organizational Behavior, and Technology. *MIT Sloan Management Review*, 1–22.
Romero, F., Li, Q., Yadwadkar, N. J., & Kozyrakis, C. (2021). INFaaS: Automated Model-less Inference Serving. In *Proceedings of the 2021 USENIX Annual Technical Conference,* Virtual Conference.
Samreen, F., Blair, G. S., & Elkhatib, Y. (2022). Transferable Knowledge for Low-Cost Decision Making in Cloud Environments. *IEEE Transactions on Cloud Computing*, *10*(3), 2190–2203.
Schneider, S., & Sunyaev, A. (2016). Determinant Factors of Cloud-Sourcing Decisions: Reflecting on the IT Outsourcing Literature in the Era of Cloud Computing. *Journal of Information Technology*, *31*(1), 1–31.
Smit, K., Zoet, M., & van Meerten, J. (2020). A Review of AI Principles in Practice. In *Proceedings of the 24th Pacific Asia Conference on Information Systems (PACIS),* Dubai, UAE.
Smith, W. K., & Lewis, M. W. (2011). Toward a Theory of Paradox: A Dynamic equilibrium Model of Organizing. *Academy of Management Review*, *36*(2), 381–403.
Sundberg, L. T., & Holmström, J. (2022). Towards 'Lightweight' Artificial Intelligence: A Typology of AI Service Platforms. In *Proceedings of the 2022 Americas' Conference on Information Systems (AMCIS),* Minneapolis, Minnesota, US.
Thiebes, S., Lins, S., & Sunyaev, A. (2021). Trustworthy artificial intelligence. *Electronic Markets*, *31*(2), 447–464.
Walsham, G. (1995). Interpretive case studies in IS research: Nature and method. *European Journal of Information Systems*, *4*(2), 74–81.
Wiesche, M., Jurisch, M. C., Yetton, P. W., & Krcmar, H. (2017). Grounded Theory Methodology in Information Systems Research. *MIS Quarterly*, *41*(3), 685–701.
Yao, Y., Xiao, Z., Wang, B., Viswanath, B., Zheng, H., & Zhao, B. Y. (2017). Complexity vs. performance: empirical analysis of machine learning as a service. In *Proceedings of the 2017 Internet Measurement Conference (IMC),* London, United Kingdom.
Zapadka, P., Hanelt, A., Firk, S., & Oehmichen, J. (2020). Leveraging "AI-as-a-Service": Antecedents and Consequences of Using Artificial Intelligence Boundary Resources. In *Proceedings of the 41st International Conference on Information Systems,* Virtual Conference.
Zheng, X., Martin, P., Brohman, K., & Da Xu, L. (2014). CLOUDQUAL: A Quality Model for Cloud Services. *IEEE Transactions on Industrial Informatics*, *10*(2), 1527–1536.