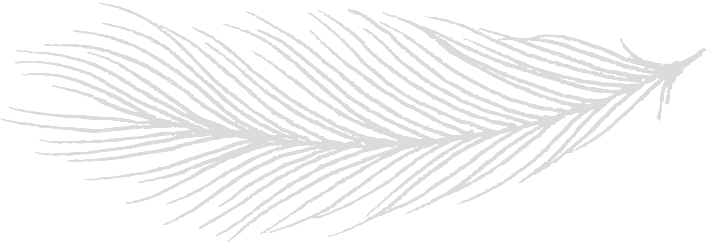
Master Thesis

# Conceptualizing the Business Model of Machine Learning as a Service Platform

Jo Lynn Tan
MSc Management of Technology
Delft University of Technology
2025



Cover photo © Unsplash

# Conceptualizing the Business Model of Machine Learning as a Service Platform

By

**J. L. Tan**

Student number: 6055443

In partial fulfilment of the requirements for the degree of:

**Master of Science**

**in Management of Technology**

Faculty of Technology, Policy and Management

at Delft University of Technology,

to be defended publicly on Friday August 29, 2025, at 10:00 AM.

**Thesis Committee**

| | |
|---|---|
| Chairperson | Prof.dr.ir. G.A. de Reuver, Information and Communication Technology |
| First supervisor | Prof.dr.ir. G.A. de Reuver, Information and Communication Technology |
| Second supervisor | Dr. H. Khodaei, Delft Center for Entrepreneurship |

An electronic version of this thesis is available at [http://repository.tudelft.nl/](http://repository.tudelft.nl/).

# Preface

The journey that culminated in this thesis began nearly 24 weeks ago—though in truth, the seed of curiosity was planted long before. For months, I had observed a phenomenon in the technology industry that intrigued me, yet remained just beyond my grasp—an elusive puzzle I couldn’t quite articulate, much less solve. Pursuing knowledge is a privilege, but it is also a humbling and often brutal endeavour. It forces us to confront our assumptions, challenge our biases, and refine our understanding of the world. Yet for those fortunate enough to have the right conditions—support, guidance, and a spark of justified curiosity—there comes a moment when the fog lifts, and a glimpse of clarity emerges. This thesis represents one such moment for me.

None of this would have been possible without the invaluable contributions of those who guided me along the way. My deepest gratitude goes to my first supervisor, Mark de Reuver, who witnessed my journey from initial confusion to eventual conviction. His patience, wisdom, and ability to gently steer me in the right direction were instrumental in shaping this work. I am equally grateful to my second supervisor, Hanieh Khodaei, whose expertise, encouragement, and thoughtful feedback helped refine my ideas and strengthen my arguments. But beyond academia, this journey has been driven by something far more personal: my family. To my wife—thank you for your unwavering belief in me, even when my attention was divided and my presence at home was fragmented. Your support was the foundation that allowed me to complete this work. And to my son—I hope that one day, you too will experience the joy of pursuing a question that captivates you, of chasing knowledge not just for its own sake, but for the way it expands your understanding of the world.

This thesis is not just the product of research; it is the culmination of encouragement, perseverance, and the quiet moments of insight that come only after pursuit of my curiosity. I am grateful to everyone who played a part in making it possible.

Best regards,
Jo Lynn Tan
July 12th, 2025

# Executive Summary

This thesis systematically conceptualizes the business model of Machine Learning-as-a-Service (MLaaS) platform by developing a taxonomy and deriving archetypes grounded in both theoretical frameworks and empirical analysis. The study addresses a critical gap in research: *while MLaaS platforms have become central to AI deployment, their business models remain understudied*. The central research question asks: “*How can MLaaS platform business models be classified into a taxonomy, and what archetypes emerge from their key dimensions and characteristics?*”

The emergence of MLaaS platforms the reflects the broader "*platformization*" of artificial intelligence (AI), where cloud-based infrastructures, dominated by Big Tech firms like Amazon, Google, and Microsoft, standardize access to machine learning tools while consolidating control over computational resources (van der Vlist et al., 2024). These platforms, offered primarily by these Big Tech firms, have become essential infrastructure for AI adoption. They abstract the complexity of machine learning workflows while simultaneously creating new forms of dependency on their proprietary infrastructure. Despite their growing industrial importance, academic research has largely focused on technical architectures rather than business models, leaving significant gaps in understanding how MLaaS platforms create, deliver, and capture value in practice. Three key limitations persist in current literature:

1. Conceptual boundaries remain ambiguous, with terms like “*AI Service Platform*” (Geske et al., 2021), "*AIaaS*" (Syed et al., 2025), “*AI Software Service*” (Lins et al., 2021), “*AI Platform*” (Mucha and Seppälä, 2020) and "*MLaaS platform*" (Philipp et al., 2020) often used interchangeably without clear differentiation. This terminological inconsistency and overlapping scope lead to misclassification and makes comparative analysis difficult.
2. Second, existing studies (Geske et al., 2021; Philipp et al., 2020; Lins et al., 2021), fail to systematically analyse business model dimensions specific to MLaaS platforms, particularly how they create, deliver and capture value.
3. Third, prior classifications (Geske et al., 2021; Sundberg et al., 2022) lack further insights about the common business model archetypes of MLaaS platforms, limiting our understanding of how specific platform design choices, such as prioritizing scalability of platform architecture or specialization in customer segments, influence the platform's ability to create value, deliver value, and capture value.

To address the three aforementioned research gaps, a qualitative, iterative methodology structured around four phases: *taxonomy design*, *development*, *evaluation* and *analysis,* was employed. This study focuses specifically on MLaaS platforms that support the complete machine learning lifecycle, from data preprocessing to model deployment and monitoring. The scope deliberately excludes adjacent categories including single-purpose AI APIs, Software-as-a-Service (SaaS) applications with AI features, and generative AI platforms. The defining characteristics of platforms included in our study are their modular services, support for multisided interactions between different user groups, and provision of shared infrastructure for machine learning workflows. In the first taxonomy design phase, we followed Möller et al.’s (2022) business model taxonomy design guidelines by establishing the data input, development approach, representation format and final application of the taxonomy upon completion. We first conducted a scoping review of 19 academic papers that helped formulate the definition of “*MLaaS platform business model*” concept while clarifying the interdependencies between digital platform, business model and MLaaS platform research domains. Building upon Osterwalder et al.'s (2005) hierarchical framework identified in the scoping review, we then developed a meta-model that integrated Teece's (2010) value creation, delivery, and capture mechanisms with Osterwalder and Pigneur's (2010) Business Model Canvas components. This meta-model established the foundation for the study by defining the key meta-characteristic as "*characteristics representing MLaaS platforms' value creation, delivery, and capture mechanisms.*" The meta-model resolved definitional ambiguities present in prior literature by first clearly distinguishing MLaaS platforms from concepts such as “*AI as a service*” (AIaaS) and “*AI service platform*”. It then focused on highlighting the value creation, delivery and capture mechanisms to describe the business model of MLaaS platforms. We synthesized Teece's (2010) framework of value creation, delivery, and capture, combined with Osterwalder and Pigneur's (2010) Business Model Canvas components to provide a starting point and guidance on the platform activities that are related to the business model characteristics.

In the taxonomy development phase, we constructed our case database by sampling 24 MLaaS platforms from the G2, business-to-business (B2B) software review website, following the predetermined research scope. We employed an iterative process based on Nickerson et al.’s (2013) taxonomy development method, iterating between conceptual grounding and empirical refinement. Inductive content analysis, including open and axial coding, was applied continuously to platform documentation and third-party publications from the public domain, enabling progressive validation and enrichment of the taxonomy. We began with the conceptual-to-empirical phase by integrating insights from an earlier scoping review and a rapid review of 12 additional studies, which identified relevant business model dimensions from adjacent domains such as AI service platforms, AI startup business models, and platform business model taxonomies. At the end of three conceptual iterations, four MLaaS platform cases were purposefully sampled to validate the dimensions and characteristics, resulting in the preliminary business model taxonomy developed from only conceptual sources. Next, we proceeded to the empirical phase, where we further refined and enriched the business model taxonomy. For each of the three empirical iterations, a different set of four cases was selected from the case database for inductive analysis, employing open and axial coding techniques. This iterative approach culminated in a final business model taxonomy consisting of 16 dimensions and 46 characteristics that systematically captured how MLaaS platforms create, deliver, and capture value. While many taxonomy components were refined during development, the conceptualization of the cost dimension remained limited due to scarce empirical data on MLaaS platform expenditures and capital allocation.

During the taxonomy evaluation phase, the taxonomy was tested against four more previously unsampled MLaaS platforms (BigML, Microsoft Azure AI, DataRobot, and VESSL AI) to assess its structural validity and practical usability. The evaluation followed Szopinski et al.'s (2019) criteria for assessing structural validity of a taxonomy, where the taxonomy should demonstrate *comprehensiveness* (covering all relevant business model aspects), *mutual exclusivity* (no overlapping dimensions), and *clarity* (precise terminology). Miles and Huberman's (1994) usability criteria were then applied to verify whether the taxonomy could be meaningfully applied to real-world cases. Each platform was analysed by mapping its business model characteristics to the taxonomy's dimensions. For example, VESSL AI's support for hybrid edge/cloud deployment validated the *deployment flexibility* dimension, while Microsoft Azure AI's integration of large language models confirmed the relevance of *model scope* in differentiating platforms. The evaluation did not seek to prove superiority but rather to determine whether the taxonomy could consistently classify the different MLaaS platform business models of the four sampled cases without gaps or redundancies. The evaluation confirmed that the taxonomy could effectively categorize the 46 business model characteristics of MLaaS platforms. While, some dimensions required refinement, such as clarifying distinctions between closely related characteristics, no fundamental restructuring was needed.

The taxonomy analysis concluded with a cross-case analysis of all 24 platforms in the case database, identifying consistent patterns across the sample that led to four different business model archetypes. *Cloud orchestrators* leverage their ownership of cloud infrastructure to deliver horizontally scalable solutions, with value creation centred on tight integration of cloud infrastructure with complementary services and value capture relying predominantly on consumption-based pricing models. *Data orchestrators* differentiate themselves through unified data ecosystems that bridge existing data analytics value propositions and support for machine learning workflows, employing hybrid pricing models that combine subscriptions with usage-based fees. These platforms emphasize data governance capabilities as complementary services. *Aggregator* platforms curate third-party tools to complement for their lack of infrastructure ownership, utilizing freemium models to attract users while depending heavily on partner networks for complementary value delivery. *Niche specialist* platforms target specific vertical markets with customized solutions, employing outcome-based pricing, proprietary integration services and consultations.

These findings indicate how value propositions of the MLaaS platforms are largely influenced by platform envelopment where cloud orchestrators like Amazon SageMaker and Google Vertex AI continue to strengthen their dominant positions by bundling MLaaS functionalities into their platform infrastructures. This aligns with "*platform envelopment*" phenomena whereby dominant platforms leverage their existing user base and technical infrastructure to bundle services to expand into adjacent market domain while generating lock-in (Eisenmann et al., 2011; Teece, 2010). *Data orchestrators* such as Snowflake and Databricks similarly achieve ecosystem lock-in through deep integration of data governance and machine learning pipelines, emphasizing process efficiency for enterprise clients (Jacobides et al., 2018; Hagiu & Wright, 2015a). In contrast, *aggregators* rely more on partnerships and modularity, while *niche specialists* secure defensible positions through verticalized workflow integrations and domain expertise, often addressing the customer needs overlooked by other business model archetypes. The findings further highlight the critical role of marketplaces and referral mechanisms in delivering value, where they not only expand service offerings but also intensify user dependence. Marketplaces serve as curated hubs where third-party tools, datasets, and models are aggregated, expanding the platform's functionality and reinforcing user dependency by embedding external resources directly into operational workflows (Hagiu & Wright, 2015a; Teece, 1986). For example, Amazon Web Services (AWS) Marketplace facilitates access to a wide range of machine learning services tightly integrated with AWS's compute and storage infrastructure, generating demand-side economies of scope and strengthening the platform's ecosystem (Teece, 2010; van der Vlist, 2024). Referral mechanisms, operationalized via partner directories, formalize relationships with technology providers and system integrators, providing users, especially in regulated or complex domains, with trusted access to expert implementation and customization support (Biyalogorsky et al., 2001; Parker et al., 2016). While marketplace and referral mechanism amplify platform credibility and adoption, they also require rigorous ongoing partner curation to mitigate risks associated with partner dependency, inconsistent service quality, which may cause potential reputational damage (Eisenmann et al., 2011; Jacobides et al., 2018).

In this research we made several contributions to the academic understanding of MLaaS platforms. First, it provides clarity in the definitions of MLaaS platform by systematically distinguishing MLaaS platform from adjacent concepts. The process of establishing the theoretical foundation of this study analysed and compared MLaaS platforms with AIaaS and AI service platform. This addresses inconsistencies present in prior literature where terms like MLaaS and AIaaS were often used interchangeably without clear differentiation. Second, the business model taxonomy with 16 dimensions advances beyond existing classifications by capturing previously overlooked aspects such as integrated managed cloud services, marketplace and referral mechanisms that prove critical to understanding real-world platform operations. Third, the taxonomy analysis yields new insights into how *niche specialist* platforms can maintain competitive positions against dominant providers through vertical specialization and customized value propositions, challenging the deterministic narrative of how Big Tech dominates in AI infrastructure markets (van der Vlist et al., 2024).

Lastly, this research also make practical contribution by offering actionable insights through the development of a meta-model, business model taxonomy, and archetypes that support MLaaS platform operators, particularly startups and new market entrants, in designing and refining their business models. The taxonomy provides a diagnostic tool enabling startups to map their existing capabilities against the four archetypes (*cloud orchestrator*, *data orchestrator*, *aggregator*, and *niche specialist*) and identify strategic positioning by highlighting critical interdependencies across dimensions such as *deployment flexibility*, *integrated resources*, and *sales channel*. The archetypes also deliver prescriptive insights for value creation, demonstrating how resource constraints can be addressed by focusing on select dimensions, such as developer experience or partner integration, as seen in *aggregator* platforms. Furthermore, the taxonomy helps established platforms benchmark their progress and anticipate the broader impacts of strategic shifts, including pricing and partner strategies, emphasizing that partner network development typically follows a sequential path from referral mechanisms to marketplaces. By conceptualizing taxonomy dimensions as design parameters rather than rigid categories, MLaaS platform operators are empowered to undertake systematic experimentation and strategic adjustments grounded in industry patterns identified through this research (Eisenmann et al., 2011; Parker et al., 2016).

# Table of Contents

# List of Abbreviations

| | |
|---|---|
| AI | Artificial Intelligence |
| AIaaS | AI as a Service |
| API | Application Programming Interface |
| AWS | Amazon Web Services |
| B2B | Business-to-business |
| C2E | Conceptual-to-empirical |
| E2C | Empirical-to-conceptual |
| GPU | Graphics Processing Unit |
| IaaS | Infrastructure as a Service |
| IoT | Internet of Things |
| LLM | Large Language Model |
| MLaaS | Machine Learning as a Service |
| MLOps | Machine Learning Operations |
| SaaS | Software as a Service |
| SDK | Software Development Kit |
| SLA | Service Level Agreement |
| SME | Small and Medium-sized Enterprise |
| VP | Value Proposition |

# List of Tables

# List of Figures

# Glossary

| **Terms** | **Definitions** |
|---|---|
| Agentic AI | AI systems capable of autonomous, goal-directed actions |
| API | Interface allowing software components to communicate and interact |
| Archetypes | Typical examples representing a category or pattern |
| Artificial intelligence | Computer systems simulating human intelligence and decision-making |
| As-a-service | Delivering software or resources over the internet, on demand |
| AutoML | Automated machine learning model selection and optimization |
| Compute infrastructure | Hardware and software resources that process and execute computational tasks for ML workloads. |
| Data drift | Change in data distribution over time that affects models. |
| Data lake | Central repository for raw, unstructured data. |
| Data lakehouse | Combines data lake flexibility with data warehouse performance. |
| Data pipeline | Series of processes moving and transforming data between systems |
| Data warehouse | Structured storage for processed, organized data. |
| Deep learning | Machine learning using neural networks with multiple layers |
| Graphics Processing Unit | Processor optimized for gaming and machine learning. |
| High performance computing | Ability to rapidly process data and execute complex computations. |
| Hyperscale | Scalable IT infrastructure efficiently managing massive computing resources on demand |
| Large language models | AI models trained on large dataset for language-related tasks |
| Machine learning | Algorithms learning patterns from data to make predictions |
| Machine learning pipeline | Automated process for data preparation, training, and deployment |
| Machine learning workflow | Sequence of steps to develop, train, and deploy models |
| MLOps | Practices for managing and deploying machine learning models |
| Modularity | Independent, interchangeable modules in systems |
| Multi-agent system | Multiple actors interacting within a shared environment |
| Multisidedness | Platform serving multiple distinct user groups simultaneously |
| Network effects | Value increases as more users join a platform |
| Platform | Foundational architecture enabling interactions among users or services |
| Software Development Kit | Collection of tools for building applications on a platform |
| Service-oriented architecture | Software design organizing functions as reusable, independent services |
| Storage infrastructure | Systems and technologies used to store, manage, and retrieve data for machine learning operations. |
| Taxonomy | Structured classification system organizing concepts or entities |
| Tensor Processing Unit | Specialized Google-designed chip accelerating machine learning and AI computations |

# 1. Introduction

## 1.1 Background

Machine learning, a study of algorithms that enable systems to learn and improve from data without explicit programming (Jordan & Mitchell, 2015), has evolved from a niche academic discipline into a core component of commercial deployment of artificial intelligence (AI) technologies. This transformation is not simply the result of technical progress; rather, it reflects the growing "*platformization*" of AI, the shift toward embedding machine learning capabilities within centralized, cloud-based platforms operated by dominant technology firms (Luitse, 2024; Helmond, 2015). These platforms play a critical role in standardizing access to machine learning tools, reducing development complexity, and enabling adoption across sectors such as healthcare, finance, and logistics. At the same time, they consolidate infrastructural control and create new forms of dependency for businesses and developers (Van der Vlist et al., 2024).

The widespread adoption of machine learning has been driven by three mutually reinforcing trends: exponential increases in computing power, unprecedented volumes of data generated by connected devices, and the rapid evolution of cloud infrastructure that supports scalable, distributed computing (Philipp et al., 2020). These developments have enabled machine learning models to be trained and deployed at scale. However, for most organizations, especially small and medium-sized enterprises (SMEs), the technical and financial barriers to developing software systems with machine learning models in-house remain prohibitive. High-performance computing resources involve substantial upfront investment, and the scarcity of skilled machine learning engineers further complicates internal development efforts. Moreover, the end-to-end lifecycle of machine learning models, from data preprocessing and model training to deployment and monitoring, requires a level of expertise and infrastructure management that many firms cannot sustain (Philipp et al., 2020). In response to these challenges, Machine learning as a service (MLaaS) platforms, with different degree of scalability and specialization, have emerged as a valuable solution, offering modular, cloud-based environments that lower the entry threshold to develop machine learning-based systems in the organizations. By abstracting complex infrastructure and providing standardized, user-friendly interfaces, these platforms allow organizations, such as the SMEs, to leverage powerful machine learning capabilities without needing to build or maintain systems from the ground up (Lins et al., 2021). In doing so, MLaaS platforms not only democratize access to machine learning but also further strengthen the role of a few dominant providers in shaping how AI is built, accessed, and governed (Van der Vlist et al., 2024).

The MLaaS platform market is dominated by Big Tech firms ( e.g., Amazon, Microsoft, and Google), with hyperscale data centres, who leverage their existing cloud infrastructures allowing to offer vertically integrated solutions with high scalability and low specialization (Van der Vlist et al., 2024; Luitse, 2024). These hyperscale cloud service providers operate massive, globally distributed data centre infrastructures to deliver scalable computing, storage, and networking services (Ma et al., 2024). The dominance of Big Tech in this space raises critical questions about market concentration and dependency. MLaaS platforms are not neutral infrastructures; many of them are tightly controlled by a handful of Big Tech firms that dictate access to computational resources, pricing models, and technical standards. Big Tech lures startups with cloud credits, training, and technical support. Microsoft's "*AI Grant*" program and Google's "*Startups Cloud Program*" incentivize reliance on their platforms (Yang & Gokturk, 2023). These initiatives serve a dual purpose: fostering innovation while ensuring dependence on proprietary ecosystems. Moreover, the proprietary nature of many platforms obscures transparency in areas like data governance, algorithmic bias, and model interpretability which are increasingly scrutinized by regulators and civil society (Whittaker, 2021). Despite these concerns, the economic and operational advantages of MLaaS platforms ensure their continued growth, making it imperative to conceptualize their underlying business models.

From an academic perspective, MLaaS platforms remain understudied despite their industrial impact. Existing research has primarily focused on technical architectures (e.g., Zou et al., 2024; Mei et al., 2025), with limited attention to their business models. Traditional digital platforms like Uber monetize transactions between users, whereas MLaaS platforms monetize infrastructure, tools, and data flows (Lins et al., 2021). Platform theory, which explains value creation in multisided markets and economies of scale (Hagiu & Wright, 2015a), provides a foundational lens to conceptualize digital platforms but lacks characteristics specific to MLaaS platforms. Moreover, the machine learning and broader AI landscape is also changing quickly. New developments, such as traditional batch training's transition to the more advanced federated learning, are reshaping how models are trained and deployed. In parallel, the rise of "*agentic AI*," an enterprise buzzword referring to AI system design that is based on deep learning, is also shifting the field. These systems can now operate with a high level of autonomy, making decisions and completing tasks with little or no human input (Ng, 2025). Not only technical, these changes affect how technology companies make money from AI. For example, new consumption-based pricing models such as pay-per-token are emerging, where users are charged based on each interaction with an AI system. However, existing business model frameworks do not fully explain or capture the logic behind these new forms of value creation and monetization (Mucha & Seppälä, 2020). This further challenges existing classifications, highlighting the need for a dedicated framework to analyse how MLaaS platforms create, deliver and capture value.

The strategic importance of MLaaS platforms extends beyond their technical utility. They are becoming critical infrastructure for innovation, similar to operating systems in software development. By providing reusable tools and services, these platforms enable a division of labour where third-party developers, data scientists, and enterprises collaborate within shared ecosystems to develop machine learning models for AI solutions. For example, Snowflake’s model marketplace allows developers to share pre-trained models, reducing duplication of effort and accelerating AI adoption. While platform openness promotes collaboration and ecosystem growth, their potential is often constrained by underlying commercial interests. This tension is particularly visible in business models with freemium pricing model, where free access draws in users, but key functionalities, such as graphics processing unit (GPU) acceleration, are restricted to paid tiers (Yao et al., 2017). Therefore, MLaaS platforms must balance fostering platform openness to attract and retain users, while also maintaining value capture mechanisms for monetization.

The societal implications of MLaaS platforms are equally significant. Their design choices influence which organizations can adopt AI. Current pricing models favour large corporations that can absorb costs for high-volume training, while SMEs rely on limited free tiers or open-source alternatives (Luitse, 2024). Geopolitically, the concentration of MLaaS platforms in the United States and China raises questions about data sovereignty, as platforms enforce jurisdictional compliance (e.g., AWS’s regional data centres) while retaining control over the core algorithms (Miceli et al., 2022). These issues highlight the urgency of conceptualizing the value creation, delivery, and capture mechanisms of MLaaS platforms to support future research that further examines how their technical, economic, and governance structures collectively shape the AI landscape.

## 1.2 Scope

This study focuses specifically on MLaaS platforms, a distinct class of cloud-based platforms with varying degrees of scalability and specialization. These platforms provide modular infrastructure and tools for building, training, and deploying custom machine learning models (Philipp et al., 2020). The broader AI as a service (AIaaS) offerings typically deliver specific AI functions, such as sentiment analysis or object detection, through a variety of "*boundary resources*" (Ghazawneh & Henfridsson, 2013), including Application Programming Interfaces (APIs), Software Development Kits (SDKs), developer guidelines, data and software. Unlike AIaaS offerings, MLaaS platforms enable users to specifically customize machine learning that spans the full model development lifecycle (Lins et al., 2021). These platforms abstract away the complexity of infrastructure provisioning and integrate capabilities such as data preprocessing, model selection, training, testing, hyperparameter tuning, and deployment. The scope is deliberately constrained to platforms that support this end-to-end machine learning lifecycle, excluding tools and services that either provide only a narrow slice of the process or deliver pre-built, non-customizable AI outputs.

This restriction is essential because MLaaS platforms operate within a multisided ecosystem that fundamentally differs from AI services with traditional Software as a Service (SaaS) models or those based on APIs. MLaaS platforms are designed to be modular and extensible, allowing different user groups, such as data scientists, machine learning engineers, business analysts, and third-party developers, to use the platform despite their different needs. This modularity also supports different pricing models and fosters innovation through reusable components like data versioning and labelling tools, and model registries (Geske et al., 2021). So, this study will focus on MLaaS platforms with varying degrees of scalability and specialization, to attain a broad understanding of the MLaaS platform domain applicable for practitioners in the SMEs. External publications tend to favour coverage of general-purpose MLaaS platforms with high scalability and less specialization. This bias is reflected in the larger presence of such platforms in the sampling frame used in this study.

To establish conceptual boundaries, the study applies Gawer’s (2014) framework for technological platforms, which identifies three core features: (1) facilitation of interactions among multiple user groups, (2) provision of shared digital infrastructure, and (3) delivery of modular services. Platforms such as Amazon SageMaker, Google Cloud Vertex AI, Snowflake, and Dataiku clearly exhibit these features. They support multisided interactions by enabling collaboration between model developers, business users, third party service providers through the platform. They offer shared infrastructure such as scalable compute resources, model hosting environments, and managed data storage. Finally, they often allow users to configure machine learning model development by selecting from a range of interoperable tools, including built-in model templates and third-party tool integrations.

The study excludes several adjacent categories that do not meet these criteria. Single-purpose tools, libraries and machine learning frameworks such as TensorFlow, PyTorch, or Scikit-learn are not included, as they do not function as platforms in the sense of coordinating interactions among multiple parties or providing integrated infrastructure services. Similarly, analytics platforms like Tableau or Power BI are out of scope because they focus primarily on data analytics and visualization rather than iterative model development. While they may integrate with predictive machine learning tools, they are not designed to support the entire machine learning lifecycle in a modular, platform-based architecture. We also excluded SaaS applications that offer AI functionality through a static predefined interface within pre-defined use cases but without supporting model customization or third-party integration.

Generative AI platforms that develop foundational large language models (LLMs) such as OpenAI's ChatGPT or Anthropic's Claude are also outside the scope of this study. These platforms provide user-facing access to large pre-trained models, but they do not support user-led model training or infrastructure management, nor do they enable modular machine learning model development workflows. Their business models resemble SaaS subscriptions or API monetization more closely than the infrastructure-centric and multi-user business model seen in MLaaS platforms. Although some MLaaS platforms are integrating foundational LLMs or LLM-based workflow automations into their offerings, this study focuses on platforms where user-led machine learning model development remains the primary value proposition. Thus, platforms that offer data analytics, AutoML, prebuilt LLMs, inference as a service, and deployment as a service as primary value proposition will be excluded.

Finally, the study prioritizes value creation, delivery and capture mechanisms, reflected in Teece's (2010) and Osterwalder & Pigneur's (2010) frameworks, such as pricing model, customer relationships, and value proposition, over technical architectures. While technical configurations are relevant, the primary emphasis is on how MLaaS platforms create, deliver and capture value.

## 1.3 Research Gap

Despite the growing adoption of MLaaS platforms, current research lacks a systematic understanding of their business models, where few studies examine how these platforms create, deliver and capture value (Mucha & Seppälä, 2020). The literature suffers from three key gaps: fragmented definitions, inconsistent scope, and a lack of business model taxonomies tailored to the MLaaS platforms.

First, the conceptual boundaries of MLaaS platforms remain ambiguous. Researchers use overlapping but distinct terms, such as "*AI service platform*" (Geske et al., 2021), "*MLaaS Platform*" (Philipp et al., 2020), and "*AIaaS*" (Syed et al., 2025) without clear differentiation. Geske et al. (2021) define AI service platforms broadly as infrastructures enabling AI application and development, while Lins et al. (2021) categorize MLaaS platforms under "*AI Software Services*," emphasizing their modular service components. Only Philipp et al. (2020) explicitly define MLaaS platforms as systems providing "*the ability to build, train, and deploy machine learning models in a single toolset,*" a definition that aligns with industry offerings like Amazon SageMaker and Google Vertex AI. This terminological inconsistency leads to conceptual blurring. For example, Mucha and Seppälä (2020) classify open-source tools like TensorFlow as "*AI platforms,*" despite their lack of multisided ecosystems or service modularity which are features central to MLaaS platforms (Philipp et al., 2020). Without precise definitions, research risks misclassifying platforms or misattributing characteristics, making it harder to compare them accurately.

Second, existing studies fail to systematically analyse MLaaS platform business models. Most focus on the technical capabilities rather than their value creation, delivery or capture mechanisms. From a service-oriented view, Lins et al. (2021) classify MLaaS at the top of three-layer AIaaS stack ("*AI Software Services*", "*AI Developer Services*", "*AI Infrastructure Services*") but overlook business model dimensions specific to MLaaS platform like pricing model (e.g., consumption-based vs. tiered model) or revenue streams. Geske et al. (2021) identify platform-oriented dimensions in his taxonomy, such as complementary services, but do not explore how these translate into value capture. Philipp et al. (2020) detail MLaaS platform functionalities, such as model training interfaces, but do not extend their analysis to value creation, delivery and capture mechanisms. This oversight leaves companies in search of outsourcing their AI capabilities without frameworks to evaluate trade-offs in MLaaS platforms, such as balancing open-source collaboration with platform monetization.

Third, prior classifications do not reveal patterns in MLaaS platform business models. Sundberg and Holmström (2022) proposed four types of AI service platforms (*analytic*s, *learning*, *conversational*, and *distributed*) based on Makarius et al.'s (2020) notions of novelty (*low*, *moderate* or *high*) and scope (*content-changing*, and *context-changing*). Geske et al. (2021) found that AI service platforms have three main purposes, which vary from requiring a lot of development work and offering much flexibility in design, to needing little development work and providing limited options for design and development. Both studies lack insights on archetypes derived from business model dimensions of MLaaS platforms. For instance, they do not explain why platforms like Amazon SageMaker adopt freemium models while others, like Microsoft Azure ML, rely on enterprise subscriptions. This gap limits the ability to predict how platform choices, such as modular services or the variety of pre-trained machine learning model offerings, affect the platform's abilities to create, delivery and capture value.

These gaps highlight the need for a taxonomy of MLaaS platform business models. By focusing on value creation, delivery and capture dimensions of these platforms, the taxonomy will address inconsistencies in prior literature, providing a foundational understanding of the MLaaS platform business models for future researcher and business practitioners from the.

## 1.4 Research Objectives & Questions

To address the identified research gap surrounding the understudied business models of MLaaS platforms, this study has two interrelated research objectives. First, it seeks to systematically conceptualize the business model characteristics of operational MLaaS platforms by developing a taxonomy grounded in both existing literature and empirical cases. Second, it aims to explain the recurring patterns, or archetypes, observed in how different business model dimensions and characteristics are combined in practice across different MLaaS platforms. These objectives are guided by the conceptual foundations of business models as articulated by Teece (2010), who emphasizes the mechanisms through which firms create, deliver, and capture value, and Osterwalder & Pigneur (2010), whose Business Model Canvas framework identifies the core business model components such as customer segment, value proposition, key resources, and revenue streams.

The central research question(**RQ**) driving this work is:

*How can the business models of MLaaS platforms be systematically classified into a taxonomy, and what common archetypes emerge from the combination of their key dimensions and characteristics?*

This question (**RQ**) targets the lack of structured understanding around the business model of the MLaaS platforms. Given the increasing importance of MLaaS platforms in AI deployments and the platformization of cloud infrastructure required in AI development, a classification of their business models is needed to clarify how these platforms create, deliver and capture value. We therefore create a business model taxonomy, which follows by deriving the archetypes to reveal the recurring patterns amongst the business model dimensions and characteristics of these MLaaS platforms. The following four sub-questions guide the research process and collectively support the taxonomy's design, development, evaluation and analysis.

***SQ1:** To what extent can the business models of MLaaS platforms be conceptualized using current literature?*

This sub-question (**SQ1**) lays the conceptual foundation for the taxonomy by drawing insights from multiple domains: cloud computing, machine learning as a service platform, digital platform, and business model research. The main goal is to define the key components and boundaries of MLaaS platform business model. To do this, a working definition of MLaaS business models is required, along with an assessment of whether current literature offers adequate support for this task. Given the interdisciplinary and fragmented nature of existing research, this synthesis is essential to ensure the taxonomy is theoretically grounded and relevant across all relevant domains (Nickerson et al., 2013).

***SQ2:** What are the key characteristics of MLaaS platform business models that are systematically captured within the business model taxonomy?*

This sub-question (**SQ2**) supports the identification and organization of business model characteristics, using the iterative methodology proposed by Nickerson et al. (2013) to develop the taxonomy. It focuses on identifying relevant dimensions and features from both empirical cases and conceptual literature, particularly those related to value creation, value delivery, and value capture (e.g., *pricing models*) (Teece, 2010). The aim is to explore how the actual characteristics of MLaaS platforms can be accurately represented in a taxonomy.

***SQ3**: To what extent is the business model taxonomy of MLaaS platform structurally valid and practically usable?*

This sub-question (**SQ3**) evaluates the structural validity and practical usability of the proposed MLaaS platform business model taxonomy by testing its relevance to real-world cases. Using established criteria for structural validity (Szopinski et al., 2019) and usability (Miles & Huberman, 1994), the evaluation examines whether the taxonomy can effectively categorize different MLaaS platforms based on their business model dimensions and characteristics. The assessment also identifies any limitations, such as overlaps or gaps in the taxonomy's dimensions. By demonstrating the ability to systematically assess MLaaS platforms according to predefined criteria, this step highlights the taxonomy's value as a classification tool for both researchers and practitioners.

---

***SQ4:** What archetypes can be derived from the recurring patterns in the MLaaS platform business model taxonomy?*

---

This sub-question (**SQ4**) aims to identify recurring patterns, from which archetypes can be derived, within the business model taxonomy developed for MLaaS platforms. It builds on the dimensions and characteristics of the taxonomy by analysing how certain combinations of dimensions and characteristics appear repeatedly. While the process of deriving archetypes goes beyond the formal taxonomy development approach outlined by Nickerson et al. (2013), it offers an additional layer of insight. Specifically, it supports the empirical understanding of business models by grouping similar cases into generalized types, making the taxonomy easier to interpret and apply (Möller et al., 2022).

## 1.5 Research Approach & Design

This study adopts a qualitative approach to develop a business model taxonomy for MLaaS platforms and subsequently deriving business model archetypes from the taxonomy dimensions and characteristics. Following Osterwalder et al.'s (2005) Business Model Concept Hierarchy, the research is structured through a three-layer conceptual hierarchy by beginning with a definition and meta-model of MLaaS platform business model, followed by a taxonomy of business model dimensions and characteristics, and archetypes derived from the observed recurring patterns. To ensure methodological rigor and transparency, we follow Möller et al.'s (2022) framework to inform the overall taxonomy design, while using Nickerson et al.'s (2013) taxonomy development method for developing the taxonomy through iterative refinement. To address methodological gaps, the study integrates other qualitative techniques in data sampling, data analysis and taxonomy evaluation.

The scope of this research is intentionally defined to focus on MLaaS platforms that support the end-to-end machine learning lifecycle, excluding pre-built AI applications that lack customization capabilities. This delineation ensures the taxonomy captures the value creation, delivery, and capture mechanisms in a busienss model, rather than just platform-specific functionalities. To establish the sampling frame, platforms are drawn from G2's "*Data Science and Machine Learning Platforms*" and "*Machine Learning Operations* (*MLOps) Platforms*" categories (G2 Crowd Inc., 2025). After screening product descriptions on G2 and platform's website, 24 cases are identified and included in a case database. Purposive sampling ensures diversity of cases with service-bundling (e.g., service-bundling by cloud or data platform vs. standalone platforms ) and different model scope (e.g., machine learning vs. large-language model support).

During the taxonomy design, we define the four meta-dimensions proposed by Möller et al. (2022): *Data* (sampling and collection), *Development* (methodological and technological scope), *Representation* (visualization and exclusivity), and *Analysis* (clustering techniques and applications). The taxonomy development phase follows Nickerson et al. (2013) by first conducting a scoping review to define the "*MLaaS platform business model*" concept and rapid review to gather input to support the conceptual-to-empirical phase of taxonomy development. The definition is then used to formulate the taxonomy's meta-characteristic: "*the characteristics representing MLaaS platforms' value creation, delivery, and capture mechanisms.*" This addresses conceptual ambiguities noted in prior literature (Philipp et al., 2020; Lins et al., 2020) and aligns with Möller et al.'s (2022) emphasis on balancing empirical data with theoretical constructs. The iterative process is governed by eight objective and five subjective ending conditions, proposed by Nickerson et al. (2013), which determine when the taxonomy can be finalized. The method iteratively proceeds through the conceptual-to-empirical and empirical-to-conceptual phases where the former builds on existing literature, while the latter derives from the real-world MLaaS platform cases. Elo & Kyngäs'(2008) inductive content analysis is applied on secondary data to identify dimensions and characteristics during the taxonomy development

In the following taxonomy evaluation phase, Yin's (2018) within-case analysis is used to evaluate the taxonomy's relevance to the real-world MLaaS platforms' business model characteristics. We analyse the application of the finalized taxonomy to four previously-unsampled platform cases, including BigML, Microsoft Azure AI, DataRobot, and VESSL AI, validating the structural validity and practical usability of the taxonomy while refining the taxonomy iteratively until all conditions are met (Szopinski et al., 2019; Miles & Huberman, 1994). In the final phase, we map the business model characteristics of all the 24 cases in the case database to our business model taxonomy and analyse their similarities and differences using cross-case analysis (Eisenhardt, 1989), where recurring patterns of business model characteristics are identified and derived into business model archetypes. During each phases in our research process, findings are synthesized, with conclusions drawn following the iterative consideration of sub-questions. The process (see Figure 1) is carefully documented to ensure transparency, and limitations are explicitly discussed.

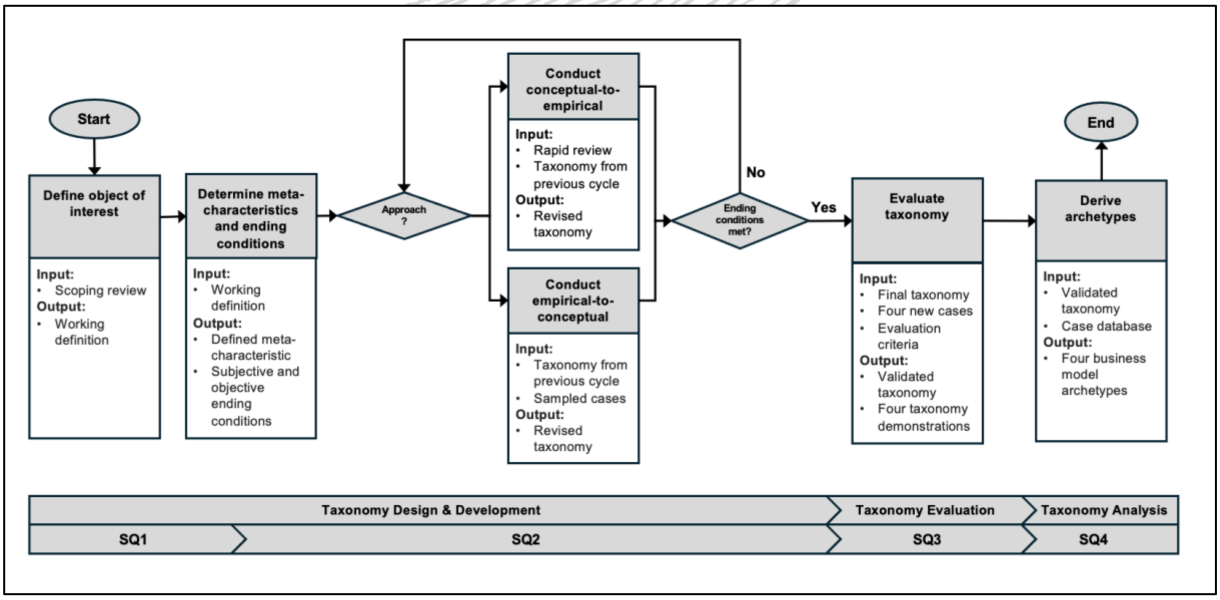


*Figure 1: Research process*

## 1.6 Relevance to MSc Management of Technology (MoT)

This research aligns closely with the objectives and syllabus of the MoT programme by examining the MLaaS platform functions and the strategic use of resources that enable firms to leverage AI technologies for competitive advantage. The study adopts a structured, hierarchical approach by defining a meta-model, developing a business model taxonomy, and deriving archetypes, to systematically analyse how these platforms create, deliver, and capture value. The research methodology reflects the MoT programme's emphasis on analytical frameworks for technology management, particularly in assessing technological markets and commercialization strategies. By focusing on the business model of MLaaS platforms, a critical enabler of AI adoption, the study bridges technical and managerial perspectives, mirroring the programme's interdisciplinary approach. Furthermore, the analysis of value creation, delivery and capture mechanisms resonates with the programme's goal of equipping students to evaluate technology's commercial impact in global, competitive contexts. This research provides a structured framework for assessing how MLaaS platforms affect corporate productivity, profitability, and strategic decision-making. It contributes foundational work that supports the MoT's mission in technology management education.

## 1.7 Thesis Structure

The remaining of this thesis is structured as the following. In Chapter 2, we present the theoretical concepts that are used throughout this study. In Chapter 3, we explain our research design where we complement the Nickerson et al.'s (2013) taxonomy development methodology with different research methods. In Chapter 4, we explain the use of the relevant literature to develop a working definition of MLaaS platform, and present the conceptual framework that guides the rest of our research. In Chapter 5, we then present the findings from the literature review and the resulting preliminary business model taxonomy. In Chapter 6, the final business model taxonomy will be evaluated and explained. Chapter 7 presents the evaluated final business model taxonomy and explain its dimensions and characteristics. In Chapter 8, we analyse and explain the similarities and differences in all 24 cases that lead to the derived business model archetypes. Lastly, we discuss the findings from this study in Chapter 9 with the final conclusions presented in Chapter 10.

# 2. Theoretical Foundations

In this chapter, we explain the core concepts and definitions, including MLaaS platforms, business models, and platform business models, that ground our study. Lastly we review the extent of prior taxonomies in capturing MLaaS platform's business model characteristics and the applicability of prior archetypes to MLaaS platform business models.

## 2.1 MLaaS Platforms

AI is the broad field of building smart machines. Machine learning is a part of AI that learns from data. Deep learning is a type of machine learning using multi-layered neural networks to automatically extract meaningful data features. LLMs are deep learning models designed to understand and generate human-like language. Within this technological landscape, MLaaS platforms are gaining relevance as the tools that offer machine learning capabilities via cloud-based services.  As machine learning adoption grows across sectors, defining enabling technologies, like MLaaS platforms, clearly becomes essential for both scholars and business practitioners.

MLaaS platform is a cloud-based digital platform that offers modular on-demand machine learning capabilities, such as data processing, model training, and deployment, through tools and APIs. It enables scalable and collaborative machine learning workflows by connecting diverse users, while abstracting infrastructure complexity through multisided platform design. MLaaS platforms combine the cloud's defining features, including on-demand access, elasticity, and infrastructure abstraction, with the coordination logic of digital platforms that connect diverse user groups (Gawer, 2014; de Reuver et al., 2018; Rochet & Tirole, 2003). They support flexible machine learning workflows through modular design (Baldwin & Clark, 2000), enable ecosystem participation via multisided interactions (Hagiu & Wright, 2015a), and scale efficiently by leveraging data and infrastructure economies (Shapiro & Varian, 1999; Evans & Schmalensee, 2016). This hybrid nature situates them uniquely within the broader AI service platform domain.

The literature (e.g., Lins et al., 2021; Philipp et al., 2020; Geske et al., 2021) consistently conceptualizes MLaaS platforms as cloud-based systems providing modular and scalable machine learning solutions. Key defining features include: (1) cloud-hosted access to AI or machine learning tools without requiring local infrastructure, (2) integrated frameworks for developing, training, and deploying models, (3) streamlined workflows that minimize data and computational demands, and (4) a service-oriented architecture facilitating flexible machine learning service integration and exchange. Taken together, these elements suggest that a robust MLaaS platform definition must go beyond technical functionality. It should capture not only the modular and cloud-based nature of these platforms but also their role as orchestrators of value within AI-driven digital ecosystems (Tiwana, 2014; Cusumano et al., 2019). Their scalable and multisided characteristics distinguishes them from traditional cloud services and overlaps with some characteristics of AI service platforms (Geske et al., 2021).

## 2.2 Business Model

Contemporary literature presents three major interpretations of business models (Massa et al., 2017): (1) as *cognitive or logical schemas* that capture how managers conceptualize value creation; (2) as *real-world attributes* observable in empirical patterns; (3), as *formal conceptual representations* (Amit & Zott, 2001; Osterwalder et al., 2005) that serve as blueprints of a company's logic. Over time, the concept of a business model has also evolved through five key phases (Osterwalder et al., 2005). It began with definitions and classifications (Timmers, 1998; Rappa, 2001), advanced to "*shopping lists*" of components (Chesbrough & Rosenbloom, 2002; Magretta, 2002), then to structured frameworks (Afuah & Tucci, 2000), followed by formal modelling (Osterwalder et al., 2005), and most recently to practical applications in digital contexts.

This evolution reflects a growing emphasis on component-based analysis. As research has progressed, many scholars have focused on breaking down business models into discrete components to facilitate systematic classification and analysis (Morris et al., 2005; Shafer et al., 2005). This component-based view supports the examination of how different configurations affect value creation, delivery, and capture, though persistent ambiguity of definitions highlighting the need for conceptual clarity in complex digital ecosystems like MLaaS platforms. We further identified 13 definitions of business models ranging from *'architecture design*' (Teece, 2010) to 'activity systems' (Zott & Amit, 2010), and include *system-based* views (Afuah & Tucci, 2000) and *abstract representations* (Al-Debei & Avison, 2010). Despite surface differences, they converge on core components: *value creation*, *delivery*, and *capture* (see Table 1), further reinforcing the observed trend.

Among these interpretations, the view of business models as conceptual representations, widely adopted in taxonomy-based research, offers both clarity and function as analytical tool (Osterwalder et al., 2005). Both the component view and conceptual representation view of business model are effectively operationalized in the Business Model Canvas (Osterwalder & Pigneur, 2010) (see Figure 2), which identifies the nine core components that collectively express how a firm creates, delivers, and captures value. These components align with Teece's

(2010) mechanisms and provide the granularity necessary for comparative analysis. Given that eight of the thirteen definitions we identified also emphasize these same value-focused dimensions, we adopt the Business Model Canvas' components and Teece's (2010) value creation, delivery and capture mechanisms concept in the conceptual framework of this study.

*Table 1: Overview of selected business model definitions from literature*

| Author(s) | Year | Definition | Concept |
|---|---|---|---|
| Afuah & Tucci | 2000 | "...a system made up of components, linkages between components, and dynamics." | System-based |
| Al-Debei & Avison | 2010 | "...an abstract representation of an organization's core architectural, operational, and financial arrangements to achieve strategic goals." | Abstract representation |
| Amit & Zott | 2001 | "...the content, structure, and governance of activities between a firm and its partners to create/capture value." | Activity system |
| Chesbrough & Rosenbloom | 2002 | "...a construct mediating between technology development and economic value creation." | Construct |
| Demil & Lecocq | 2010 | "...an articulated or implicit logic for firm activities, enabling value creation and capture." | Business logic |
| Magretta | 2002 | "...a story that explains how an enterprise works." | Narrative |
| Osterwalder et al. | 2005 | "A business model is a conceptual tool containing a set of objects, concepts and their relationships with the objective to express the business logic of a specific firm. " | Conceptual representation |
| Teece | 2010 | "...the design or architecture of the value creation, delivery, and capture mechanisms employed." | Architecture design |

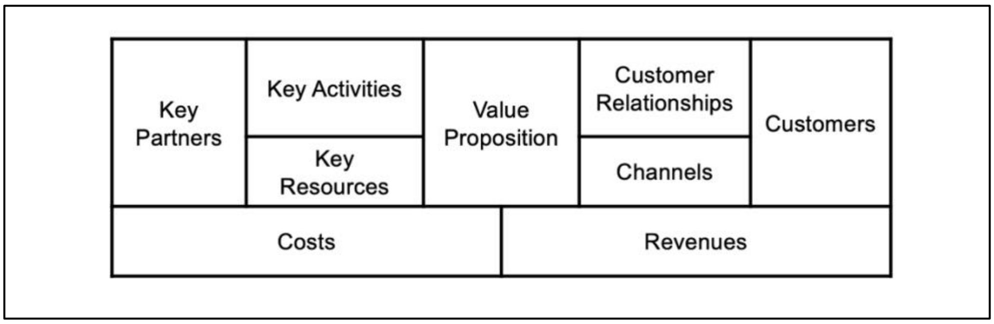


*Figure 2: Business Model Canvas and components (Osterwalder & Pigneur, 2010)*

## 2.3 Value Creation, Delivery and Capture in Platform Business Model

Platforms have emerged as the organizational forms that redefine how value is created, delivered, and captured across industries (de Reuver et al., 2018). Drawing on Teece's (2010) notion of business models as "*the design or architecture of the value creation, delivery, and capture mechanisms employed,*" most platform scholars conceptualize platform-based models around these three core dimensions (Clauss et al., 2017; Clauss et al., 2019; Osterwalder & Pigneur, 2013). Unlike traditional pipeline firms that transform inputs through linear processes and own physical assets (Kortmann & Piller, 2016), platforms orchestrate interactions among independent participants. Their advantage lies not in controlling resources (Barney, 1991), but in intermediating between distinct user groups and scaling through network effects and modular architectures (Parker et al., 2016; van Alstyne et al., 2016).

Platforms redefine value creation by enabling direct exchanges among participants rather than relying on internal resources. MLaaS platforms like Snowflake exemplify this shift, providing tools and services, such as model training tools, that facilitate third-party innovation (Gawer, 2014). As platform's technology partners contribute prebuilt models or datasets, the platform's utility grows, creating "*data network effects*" (Gregory et al., 2021; Haftor et al., 2021). This observation aligns with de Reuver et al.'s (2018) view that platforms excel by harnessing external resources rather than owning them. The traditional resource-based view is inverted, as value stems from orchestrating participation rather than internal research and development (Barney, 1991; Parker et al., 2016; van Alstyne et al., 2016).

In delivering value, platforms standard interactions through modular interfaces (Hagiu & Wright, 2015a), by acting as matchmakers and optimizing cross-side interactions where the activity of one user group enhances the utility for another (Rochet & Tirole, 2003). MLaaS platforms, such as Amazon SageMaker, abstract infrastructure complexities via APIs, allowing developers to focus on model development without managing backend systems. This modularity balances stability and innovation, ensuring core functions remain consistent while complementary services evolve independently (Gawer, 2014). Platforms, like Google Vertex AI, further demonstrate this flexibility by supporting different industries, from healthcare to finance, without requiring customized configurations. As de Reuver et al. (2018) note, such architectures enable platforms to scale across sectors, contrasting with traditional pipeline models that rely on proprietary, inflexible value delivery systems (Cusumano et al., 2019).

The mechanisms for value capture in platforms differ fundamentally from those of pipeline businesses, as revenue is tied to ecosystem activity rather than standalone product sales. Hagiu and Wright (2015a) distinguish between direct monetization, such as transaction fees, and indirect strategies, like leveraging user bases for ancillary revenue, which is the supplementary revenue streams that complement the main platform activities. MLaaS platforms often combine both approaches. For example, while offering free access to model repositories, Amazon SageMaker charges pay-per-token fees for compute resources, which are the computational power and infrastructure required to process data, run applications, and perform calculations in machine learning models. This hybrid model aligns with van Alstyne & Parker’s (2017) insight that platforms maximize profits by scaling user participation at minimal marginal cost. However, governance is important because if the platform takes too much value from partners, it can discourage their participation and limit the growth of the ecosystem (de Reuver et al., 2018). Instead of relying on licensing markups, platforms adopt infrastructure as a service (IaaS) pricing and revenue-sharing models (Tiwana et al., 2010). This also reflects Rochet and Tirole’s (2003) idea that using different pricing strategies for different users can help the platform grow over time.

## 2.4 Business Model Taxonomies

Business model taxonomies serve as the foundational tools for structuring and analysing how firms create, deliver, and capture value. Historically, these taxonomies have evolved from simplistic, deductively derived typologies, such as Timmers’ (1998) early classifications of e-business models, to more sophisticated, empirically grounded frameworks that leverage clustering techniques and multidimensional analysis (Lambert, 2015; Remane et al., 2016).

The shift reflects a growing recognition that business models are complex, dynamic systems requiring granular categorization to account for industry-specific nuances, such as those in digital platforms or AI-based services. However, while existing taxonomies offer valuable insights into business models, they often neglect emerging domains like MLaaS platforms with inherited characteristics from multiple domains such as cloud computing, digital platform and AIaaS. Thus, they demand adapted frameworks to support further analysis. So, this section focuses on critiquing prior taxonomy methodologies and identifying gaps in their application to MLaaS platforms.

The development of business model taxonomies, especially in platform domain, has predominantly followed iterative, mixed-method approaches, blending deductive (conceptual-to-empirical) and inductive (empirical-to-conceptual) strategies. Nickerson et al.’s (2013) method is frequently cited, as seen in Staub et al. (2021), where dimensions are first derived from literature (e.g., *value proposition*, *revenue streams*) and later refined through empirical validation. For instance, Staub et al. (2021) classify digital platforms by combining categories from Gassmann & Frankengerger’s(2014) Business Model Navigator (e.g., *value proposition*, *target customer*) with case studies from the real estate industry. These studies exemplify the iterative refinement of taxonomies, where conceptual dimensions (e.g., "*platform structure*") are tested against real-world data to ensure robustness (Staub et al., 2021). Deductive studies, as in Freichel et al. (2021), prioritize theoretical coherence, extracting dimensions like "*transaction type*", both online or offline, from platform. In contrast, inductive studies, such as Engelbrecht et al. (2016), derive taxonomies from empirical data, such as clustering data-driven business models based on "*technological effort*" and "*target audience.*"

Upon inspection, existing taxonomies exhibit two key limitations when applied to MLaaS platforms. First, they overlook business model dimensions specific to MLaaS platforms. For instance, Weber et al. (2022) and Metelskaia et al. (2018), propose classification for AI startup business model with a broad focus and the level of analysis that remain at the organizational level, but not directly related to MLaaS platforms. Second, platform business model taxonomies prioritize marketplace dynamics (e.g., buyer-seller matchmaking) over the value creation, delivery and capture mechanisms of MLaaS platforms. For example, Freichel et al. (2021) classify platforms by "*transaction type*" (online or offline) and "*market orientation*" (buyer or seller), which are irrelevant to MLaaS platforms like Databricks or DataRobot that use computational infrastructure and model management tools as resources to create value. This misalignment stems from a focus on transactional ecosystems (e.g., Airbnb, Uber) rather than technical platforms where value is captured through computational resource allocation or algorithmic services such as AutoML that automates data preprocessing, algorithm selection, and hyperparameter tuning in machine learning. Ejsmont et al. (2024) highlight this gap, noting that multisided AI platforms require revenue models tailored to data scientists’ workflows, not consumer-marketplace interactions.

Third, digital platform taxonomies often assume a static marketplace structure, ignoring the layered architecture of MLaaS platforms. For instance, Täuscher & Laudien's (2017) taxonomy categorizes platforms by "*transaction content*" (product or service) and "*revenue source*" (buyer or seller), which fails to capture how MLaaS platform providers like Microsoft Azure AI operate across infrastructure (Infrastructure as a service or 'IaaS'), development (PaaS), and application (SaaS) layers where each later has distinct monetization model (e.g., cloud credits for model training, subscriptions for access to APIs). This limitation reflects a broader disconnect: marketplace taxonomies treat platforms as neutral intermediaries, while MLaaS platforms actively shape value creation through technical constraints (e.g., model latency thresholds) and proprietary tooling (e.g., integrations with third party tools).

Our conceptual framework addresses these gaps by establishing a meta-model, using Business Model Canvas' component-based structure (Osterwalder & Pigneur, 2010) with the emphasis on Teece's (2010) value creation, delivery and capture mechanisms (see Chapter 4). This meta-model is then used to scope the dimensions and characteristics relevant to MLaaS platform during the taxonomy development.

## 2.5 Business Model Archetypes

Business model archetypes are generalized, recurring patterns that describe how firms create, deliver, and capture value (Osterwalder & Pigneur, 2010; Remane et al., 2017). Rooted in configurational theory, these archetypes represent ideal-type combinations of business model elements that emerge across industries (Massa et al., 2017). In the context of business model and information systems research, archetypes refer to abstract patterns or ideal types that capture common characteristics observed across multiple instances of a phenomenon (Gregor, 2006). They serve as simplified representations that help researchers and practitioners classify, compare, and analyse complex real-world cases by distilling them into broader categories.

Each of the three papers, Sundberg et al. (2022), Weber et al. (2022), and Geske et al. (2021), derive motives or archetypes for AI service platforms or AI startups. But their approaches differ in their methodology and focus. Sundberg et al. (2022) develop a typology of AI service platforms using a literature review, identifying four archetypes: *analytics*, *learning*, *conversational*, and *distributed* platforms (see Table 2). Their classification is based on three dimensions: *AI definitions* (e.g., systems that think or act like humans), *scope* (content-changing vs. context-changing AI), and *novelty* (low to high). For instance, analytics platforms are characterized by low novelty and content-changing automation, while distributed platforms are high novelty and autonomous systems (Sundberg et al., 2022).

*Table 2: AI service platforms typology (Sundberg et al., 2022)*

| Type | Definition | Novelty and scope |
|---|---|---|
| Analytics | Systems that think rationally and act according to rules and formal procedure. | Low level of novelty. Content changing through automation or amplification. Humans act as controllers or conductors. |
| Learning | Systems that think like humans, through neural networks. | Medium level of novelty. Context-changing through alteration. Humans act as co-creators. |
| Conversational | Systems that act like humans through an interactive interface. | Low level of novelty: Content changing through automation with Humans as controllers. |
| Distributed | Systems that act rationally in multi-actor environments. | High level of novelty: Content changing through autonomous systems kept in check by humans. |

Weber et al. (2022) take a more empirical approach, constructing a taxonomy of AI startup business models using a sample of 100 startups and hierarchical cluster analysis. Their four archetypes, *AI-charged Product/Service Provider*, *AI Development Facilitator*, *Data Analytics Provider*, and *Deep Tech Researcher*, are derived from dimensions such as *core AI value*, *primary AI technology*, and *delivery mode*. For example, *AI-charged Product/Service Providers* embed pre-trained AI models into standardized offerings, whereas *Deep Tech Researchers* focus on cutting-edge AI solutions for niche problems (Weber et al., 2022). The authors emphasize the role of data in value creation and the technical complexity of AI, which influences how startups position themselves in the market.

Geske et al. (2021) focus on AI service platforms, developing a taxonomy through iterative analysis of 31 platforms and expert interviews. Their taxonomy layers highlight how platforms vary in industry focus, service models (e.g., ready-to-use vs. development services), and complementor involvement. While they do not explicitly label archetypes, their discussion reveals three motives: platforms enabling custom AI development, those offering ready-to-use AI applications, and those providing complementary services like data labelling (Geske et al., 2021). Their work highlights the modularity of AI platforms and the spectrum of customization options available to users.

Despite differences in methodology, the papers share key insights about AI-specific business model archetypes. First, the central role of technical complexity and customization in differentiating archetypes are highlighted. Platforms range from low-code, user-friendly solutions such as Sundberg et al.'s (2022) "*lightweight AI*", to highly specialized, developer-centric environments such as Weber et al.'s (2022) *Deep Tech Researcher*. Second, data's role in value creation is a recurring theme. Analytics platforms rely on data for decision support (Sundberg et al., 2022), while *AI Development Facilitators* leverage data for model training (Weber et al., 2022). Third, the degree of third-party integration varies, with some platforms acting as standalone solutions (e.g., Geske et al.'s ready-to-use services) and others acting as ecosystems (e.g., multisided platforms).

When deriving archetypes, the papers suggest several considerations. First, dimensions of classification must be clearly defined, whether *functional* (e.g., AI capabilities), *structural* (e.g., platform layers), or *strategic* (e.g., market focus). Second, empirical validation is critical; Weber et al. (2022) and Geske et al. (2021) use real-world samples to ground their taxonomies, while Sundberg and Holmström (2022) rely on literature synthesis. Third, practical applicability should be prioritized. So, business model archetypes should help researchers or business users navigate choices, such as selecting platforms based on technical needs (Geske et al., 2021) or evaluating startup business models (Weber et al., 2022).

# 3. Methodology

To classify the different business models of MLaaS platforms (**RQ**), our study are guided by two main research objectives: (1) To systematically conceptualize the business model of MLaaS platforms using a business model taxonomy, (2) To analyse and explain the archetypes, observed in the recurring patterns across the different business model characteristics in MLaaS platforms. As established in Chapter 2, this study adopts Osterwalder et al.'s (2005) three-layer hierarchical view to conceptualize the MLaaS platform business model. This chapter explain the taxonomy design, data collection, taxonomy development, taxonomy evaluation and taxonomy analysis applied in this study.

## 3.1 Taxonomy Design Phase

A *taxonomy* is a systematic classification of objects (e.g., MLaaS platforms) based on shared characteristics. It is different from broader *classifications* (general groupings), *frameworks* (theoretical constructs), or *typologies* (conceptually derived categories) (Nickerson et al., 2013). Unlike purely empirical classifications, such as *numerical taxonomy* (Sokal & Sneath, 1963) or *unidimensional typologies* (Bailey, 1994), taxonomies integrate both conceptual foundations (e.g., Teece's value creation, delivery and capture mechanisms) and empirical validation (e.g., platform documentations) to generate actionable insights (Doty & Glick, 1994). Taxonomies are particularly suited for this study because they provide a structured way to organize complex, multidimensional phenomena around the different business models of MLaaS platforms. By defining the dimensions and characteristics of the MLaaS platform business model, the taxonomy can conceptualize the business activities these platforms use to create, deliver, and capture value (Teece, 2010).

The initial taxonomy design followed Möller et al.'s (2022) four meta-dimensions(see Table 3). The object of analysis was the operational MLaaS platforms that were sampled from G2's (G2 Crowd Inc., 2025) listings. Amongst the 24 cases identified, 20 cases were purposively sampled for use in the taxonomy development as the ending conditions was met, with all 24 cases included in the taxonomy analysis. Data collection focused on the publicly available platform documentations, peer-reviewed articles and proceedings from SCOPUS and Google Scholar. In the development phase, we use of Nickerson et al.'s (2013) iterative taxonomy development method, complemented by qualitative techniques whenever needed. We used a morphological box to visualize the business model taxonomy (Szopinski et al., 2020), highlighting the business model dimensions that were either mutually or partially exclusive. Finally, taxonomy analysis derived archetypes from all cases using cross-case analysis.

*Table 3: Business model taxonomy design (Adapted from Möller et al., 2022)*

| Meta-dimension | Dimension | Characteristics |
|---|---|---|
| Data | Object of Analysis | MLaaS platforms |
| | Data Collection | Platform documentations, third party publications, literature |
| | Data Sampling | Theoretical sampling (24 platforms) |
| | Theoretical Lens | Teece's (2010) value creation, delivery and capture mechanisms and Osterwalder et al.'s (2010) Business Model Canvas |
| Development | Development Method | Nickerson et al. (2013) complemented with qualitative methods |
| | Industry focus | AI industry |
| | Technology Scope | All types of MLaaS platforms |
| | Depth of Analysis | Theoretically grounded business model dimensions aligned with Business Model Canvas components (Osterwalder et al., 2010) |
| Representation | Exclusivity | Both mutually or partially exclusive |
| | Visualization | Morphological box (Szopinski et al., 2020) |
| Application | Further Application | Taxonomy analysis to derive archetypes |
| | Clustering Tool | Cross-case analysis |

This study adopted an interpretivist approach, viewing a MLaaS platform business model as socially constructed through platform's interactions with its stakeholders and various contextual factors (Bryman, 2016). The research emphasized an understanding of how meaning emerged from a platform's documentation, branding, and online communications rather than seeking objective classifications. An iterative, context-dependent taxonomy development process examined both observable features and implied meanings in business model characteristics (Schwandt, 1994). Key dimensions (e.g., *value proposition*) were interpreted through close reading of platform narratives rather than rigid categorization. Researcher reflexivity ensured transparency in interpretive judgments, particularly when analysing edge cases or conflicting evidence (Yanow, 2014). While not pursuing generalizable truths, the study grounded its taxonomy in empirical cases while acknowledging the interpretive nature of all classifications. This approach proves particularly valuable for MLaaS platforms, where business models evolve dynamically through technological and social influences.

### 3.1.1 Literature Review

Before beginning the taxonomy development phase, the meta-characteristic and definition of our object of analysis were defined by a two-phase literature review. Due to limitation of resources, both time and researchers available, we selected literature review approaches that would allow us to address two distinct research objectives efficiently. The first objective required broad conceptual understanding of MLaaS platform business models, while the second required identification of specific taxonomy dimensions from existing frameworks. Using two targeted approaches enabled us to achieve both breadth in definitional clarity and depth in identifying specific taxonomy dimensions and characteristics within our resource constraints. We used a scoping review to formulate a working definition of "MLaaS platform business model" using contemporary literature(Arksey & O'Malley, 2005; Tricco et al., 2015). This definition led us to the meta-characteristic of our taxonomy (see Chapter 4): "*the characteristics representing the value creation, delivery, and capture mechanisms of MLaaS platforms.*" We then followed up with a rapid review (Ganann et al., 2010) to gather business model dimensions and characteristics from the recent business model taxonomies literature, for inclusion in the preliminary taxonomy.

#### *3.1.1.1 Scoping Review*

A scoping review is a structured methodology designed to systematically explore the breadth and depth of existing literature on a particular topic to identify key themes and gaps requiring further investigation (Arksey & O'Malley, 2005; Tricco et al., 2015). Given the unclear conceptualization of MLaaS platforms, we followed the guidance from Nickerson et al. (2013) by first adopting an exploratory approach though a scoping review. We broadly followed this method to answer our first sub-question (**SQ1**): "*To what extent can the business models of MLaaS platforms be defined using current literature?*" This phase adopted Arksey and O'Malley's (2005) scoping review framework, guided by several key components (see Table 4).

*Table 4: Overview of methodology components of scoping review*

| Component | Description |
|---|---|
| Objective | 1. Map literature on MLaaS platform, business models, cloud computing, and digital platforms.<br>2. Analyze the relevant definitions to guide our own formulation. |
| Methodology | Arksey and O'Malley's (2005) framework: Research question formulation, database searches, screening, synthesis. |
| Data Sources | SCOPUS queries on the four domains. |
| Inclusion Criteria | Peer-reviewed articles (2015–2024); focus on definitions and frameworks; exclude technical and non-platform-related studies. |
| Outcome | Working definition of MLaaS platform business models and concept map. |

The scoping review process involved several key stages: (1) defining clear research objective, (2) executing a search strategy across SCOPUS database, (3) applying inclusion and exclusion criteria to filter relevant studies, and (4) synthesizing findings thematically to construct a cohesive narrative (Arksey & O'Malley, 2005; Peters et al., 2020). Our research aimed to formulate a working definition for MLaaS platform business models, a concept currently shaped by conflicting assumptions (e.g., "*Is cloud computing a prerequisite?", "How ingrained is platform theory in MLaaS platform?*"). The objectives of this scoping review were: (1) to identify the domains relevant to the formulation of working definition, and (2) to collect and analyse the definitions of relevant concepts. Using this method, we extracted foundational concepts from business model (e.g., Teece, 2010; Osterwalder & Pigneur, 2010) and information systems literature (e.g., Geske, 2021; Lins et al., 2021; Sundberg et al., 2022) to determine the key components that a business model should address. Meanwhile we also make distinctions such as differentiating "*MLaaS platforms*" (Philips et al., 2021) from "*AI service platforms*" (Geske et al., 2021) by understanding their differences in their focus, target user and level of configurability. We then illustrated their links with other relevant domains in a conceptual map (see Figure 3) and a concept matrix (see Table 5). With a bottom-up approach, we then inductively synthesized insights from 19 identified peer-reviewed articles, focusing on establishing cloud computing's role in the emergence of MLaaS platform and the manifestation of platform-based business model characteristics in MLaaS platforms. Through deductive reasoning, we derived a working definition aligned with platform-based business models and cloud computing origins. This evidence-based definition of "*MLaaS platform business model*" allowed us to frame the meta-characteristics of the business model taxonomy later in our study. Detailed application of the methodology, result and analysis are explained in Appendix A1.

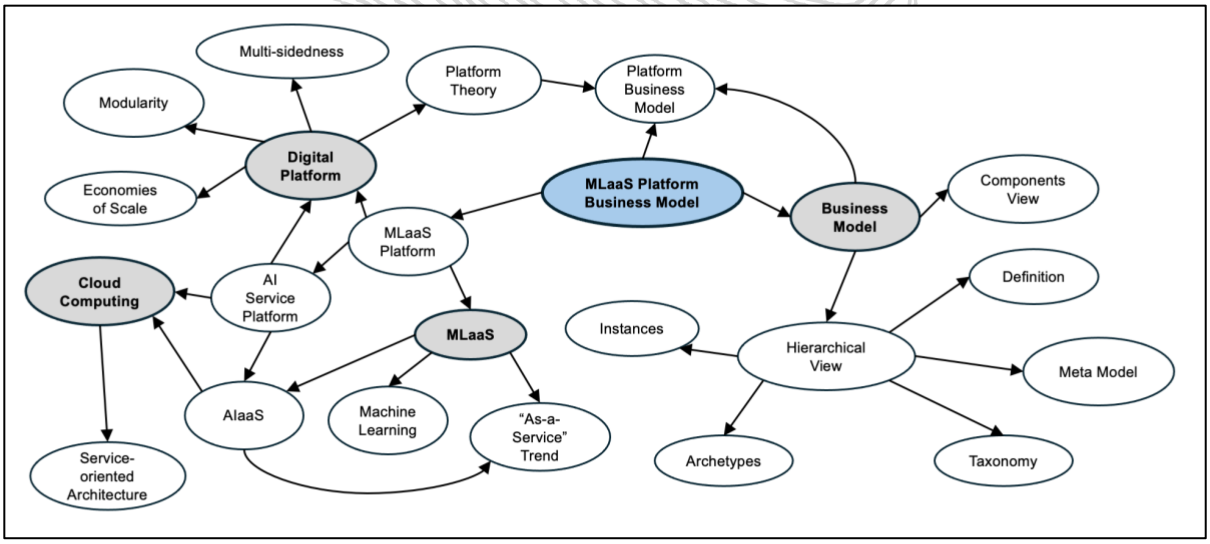


*Figure 3: Concept map*

The scoping review aligned with exploratory theory-building (Edmondson & McManus, 2007), necessary for emerging domains like MLaaS platforms. By systematically mapping definitions and theoretical concepts, the scoping review addressed the lack of conceptual clarity in MLaaS platform research. Its broad scope ensured our definition of the taxonomy's meta-characteristic was grounded in both information systems and business model literature, mitigating bias from isolated disciplinary perspectives.

*Table 5: Concept matrix from scoping review*

| Contribution | Year | Concept | | | |
|---|---|---|---|---|---|
| | | Business Model | MLaaS Platform | Cloud Computing | Digital Platform |
| Belussi F. et al. | 2019 | x | | | |
| Cozzolino A. et al. | 2021 | | | | x |
| De Reuver et al. | 2018 | | | | x |
| Durao et al. | 2014 | | | x | |
| Gawer | 2022 | | | | x |
| Geske et al. | 2021 | | x | | |
| Hagiu & Wright | 2015 | | | | x |
| Kapoor & Teece | 2021 | | | | x |
| Lins et al. | 2021 | | x | | |
| Luitse | 2024 | | x | x | x |
| Osterwalder et al. | 2015 | x | | | |
| Philipp R. et al. | 2020 | | x | | |
| Priem et al. | 2018 | x | | | |
| Ritter & Lettl | 2018 | x | | | |
| Shree et al. | 2021 | | | | x |
| Teece & Linden | 2017 | x | | | |
| van der Vlist et al. | 2024 | | x | x | |
| Veit et al. | 2014 | x | | | |
| Zhao et al. | 2020 | x | | | x |

#### 3.1.1.2 Rapid Review

A rapid review is an accelerated literature synthesis method designed to deliver timely, evidence-based insights (Tricco et al., 2015). Unlike systematic reviews, which prioritize exhaustive coverage, rapid reviews strategically streamline the review process, focusing on highly relevant studies to efficiently identify key trends, gaps, and applicable frameworks (Ganann et al., 2010). We employed a rapid review to address our second sub-question (**SQ2**): *"What key characteristics demonstrated by MLaaS platform business models can be included in the taxonomy?"* The rapid review provided useful insights for the conceptual-to-empirical iteration of our taxonomy development, as it systematically extracted relevant dimensions and characteristics of existing business model taxonomies across MLaaS platform and other related domains like platform-based models and AI-specific taxonomies.

The rapid review pursued two specific objectives: (1) identifying suitable business model taxonomy dimensions and characteristics, and (2) assessing the applicability of those taxonomy dimensions and characteristics to MLaaS platforms. To ensure conceptual alignment, we focused on literature addressing platform as the level of analysis, employing broad search terms including *"business model," "taxonomy," "MLaaS,"* and *"platform."* After adapting Mwilu et al.'s (2015) inclusion criteria for emerging technologies (see Table 6), the review concentrating on non-technical studies in information systems, business, and management domains published within the last decade, while permitting backward searches for seminal works.

*Table 6: Search criteria of rapid review*

| Search Criteria | Scope |
|---|---|
| Content | Focused on MLaaS platforms or related domains, expanding to platform business model taxonomies. Excludes digital marketplaces and non-AI taxonomies without business model dimensions. |
| Language | English publications only. |
| Research Design | Non-technical papers within information systems, business and management domain. |
| Publication Date | Last 10 years, with backward reference searching permitted for foundational sources. |

The rapid review's methodological approach reflects theory-refining research (Edmondson & McManus, 2007), where iterative synthesis of existing taxonomies informs emergent frameworks. Our reproducible search protocol combined predefined SCOPUS queries with backward reference searching and screening of the title, abstract and text of the literature. The protocol reduced the corpus to 12 studies selected for full text analysis. This approach mirrored the scoping review's deductive-inductive synthesis, with digital platform theory (Rochet & Tirole, 2003; Gawer & Cusumano, 2014; de Reuver et al., 2018) anchoring our effort to extract business model dimensions while emergent themes (e.g., configurability in machine learning services) refined the boundaries of each dimension. Detailed search criteria, application of the methodology, result and analysis are explained in Appendix A2.

## 3.2 Data Collection, Sampling and Analysis

This study used desk research as the main data collection method, through systematic gathering and analysis of secondary data from peer-reviewed academic literature, platform documentation, white papers, and third party publications. Within the context of this study, desk research enabled the development of a curated MLaaS platform case database.

### 3.2.1 Constructing the Case Database

In this study, we used a four-step process used to construct the case database (see Figure 4). The sampling frame was derived from G2, a B2B software review platform, widely recognized for its structured categorization and verified user feedback (G2 Crowd Inc., 2025).

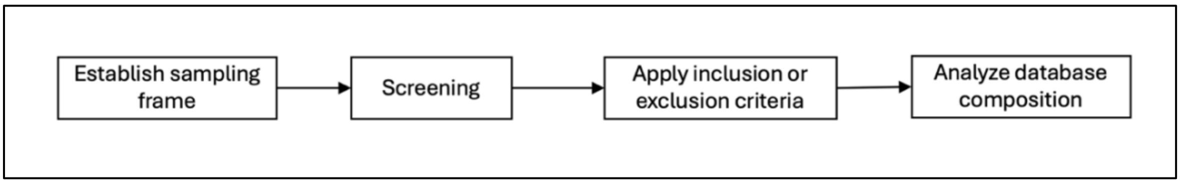


*Figure 4: Case database construction flow diagram*

G2's categorization was leveraged to systematically identify MLaaS platform-relevant offerings within two primary software groups: "*Data Science & machine learning Platforms*" (252 listings) and "*MLOps Platforms*" (159 listings). These categories were chosen based on their alignment with the end-to-end machine learning lifecycle, a defining characteristic of MLaaS platforms. A set of strict inclusion and exclusion criteria (see Table 7) was applied to filter the cases. Platforms were kept only if they showed the core characteristics of MLaaS platform: (1) full or modular lifecycle support (e.g., training, deployment, monitoring), (2) multisided interaction capabilities (e.g., enabling collaboration between users and third party providers through the platform), and (3) platform modularity that supported multiple machine learning services and third-party integration.

*Table 7: Inclusion and exclusion criteria of initial database construction*

| **Category** | **Criteria** | **Objective** |
|---|---|---|
| Inclusion Criteria | | |
| 1. Digital Platform Characteristics | Modular (supports multiple machine learning services) and at least two-sided (serves both technical and non-technical users). | Ensures platforms are flexible and serve diverse user roles. |
| 2. End-to-End machine learning Lifecycle Support | Covers major machine learning activities (e.g., data preparation, model training, deployment). | Excludes tools limited to a single phase (e.g., only training or deployment). |
| 3. Operational Maturity | Excludes startups with unreleased products or recently acquired/merged companies. | Avoids bias from unstable or rapidly evolving features. |
| 4. English Documentation | Sufficient public English materials for analysis. | Ensures consistency in data collection. |
| Exclusion Criteria | | |
| 1. Narrow-Scope Tools | Data labelling (e.g., V7), or deployment-only (e.g., Clarifai). | Lacks full MLaaS platform capabilities. |
| 2. Non-Platform Solutions | Standalone model hosting (e.g., Hugging Face) or governance tools (e.g., Dagshub). | Missing end-to-end platform features. |
| 3. Agentic AI Platforms | Focus on autonomous agents (e.g., Akira.AI), not structured machine learning workflows. | Outside MLaaS platform scope. |

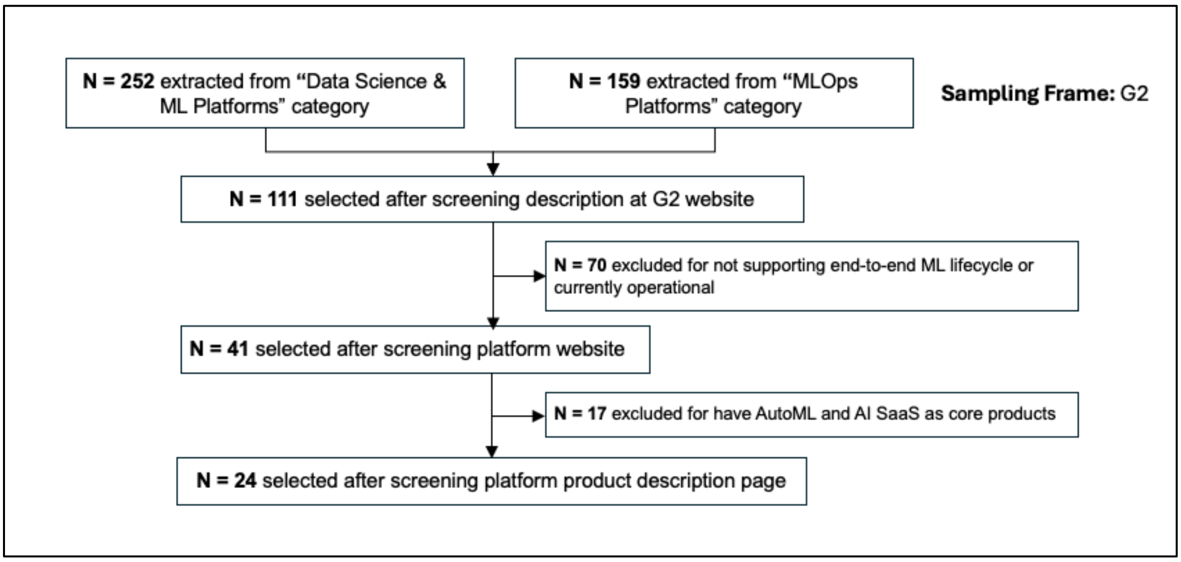


*Figure 5: Case database construction flow diagram*

The first round of screening of product specification page on the platform's website was conducted to verify the four inclusion criteria. We then further examined the product specification page of the cases to exclude platforms that presented AutoML as their core value proposition or were AI SaaS platforms. This deliberate screening process resulted in a final sample of 24 platforms (see Figure 5 above and the case database in Appendix B). We observed that the cases in the database were organized around four dimensions, product category stated on the G2 website, machine learning model scope, industry focus, and service bundling (see Table 8 below). So, we confirmed that the final database of 24 platforms demonstrated diversity by capturing the full spectrum of MLaaS platform characteristics.

*Table 8: Diversity in the case database*

| Dimension | Count | Definition |
|---|---|---|
| *Product Category (Source: G2)* | | Classifies platforms based on their primary focus in the machine learning workflow |
| Data Science & Machine Learning Platform | 8 | General-purpose tools for end-to-end machine learning development, from experimentation to deployment |
| MLOps Platform | 16 | Specialized in machine learning operational aspects |
| *Model Scope* | | Indicates whether platforms cater to specific type of machine learning |
| Either machine learning or LLM | 10 | Designed for singular model types |
| Both machine learning and LLM | 14 | Model-agnostic, supporting various types of machine learning models |
| *Industry focus* | | Reflects the customer segment focus of the platform |
| Broad | 10 | Tailored for SMEs and general-purpose use cases |
| Enterprise | 10 | Tailored for large corporations |
| Enterprise - Industry Specific | 4 | Tailored for specific segments (e.g., healthcare, finance) |
| *Service bundling* | | Indicates where platform is part of larger technology companies' service offerings |
| No | 14 | Independent platforms (e.g., DataRobot). |
| Yes | 10 | Part of service bundling in a software or cloud vendor's product offering (e.g., Amazon SageMaker, Google Vertex AI) |

### 3.2.2 Case Sampling, Data Analysis and Data Sources

In this study, we employed theoretical sampling to select cases that represented extreme, typical, or contrasting configurations (Eisenhardt, 1989). Unlike random sampling, this approach prioritized cases that challenged or confirmed emerging patterns, strengthening the taxonomy's conceptual and empirical validity (Glaser & Strauss, 1967). Crucially, it prevented bias toward dominant market players by deliberately including niche and emerging platforms, thereby enhancing generalizability. Theoretical saturation was achieved when additional cases no longer yielded new insights, confirming the taxonomy's comprehensiveness (Guest et al., 2006).

In the conceptual-to-empirical phase, the preliminary taxonomy was tested against real-world cases. Superb AI Suite, HyperSense, Snowflake, and IBM Watsonx Studio were chosen because they represent distinct configurations across two key dimensions: (1) *Service bundling from existing data analytics or cloud providers*. This tests whether the taxonomy can differentiate between integrated (Snowflake and IBM Watsonx Studio) and standalone (Superb AI and HyperSense) MLaaS platform business models. (2) *Machine learning model scope*. This variation helps assess whether the taxonomy adequately captures specialized (Superb AI) and general-purpose (Snowflake, HyperSense and IBM Watsonx Studio) platforms. In the empirical-to-conceptual phase, theoretical sampling shifted toward selecting cases based on the objective of each iteration.

*Table 9: Guiding questions for categorization*

| Guiding Question for Categorization | Meta Model Second Order Layer |
|---|---|
| What core value does the platform deliver to its users through its services? | Value Proposition |
| What technical or operational processes enable the platform's functionality? | Key Activities |
| What proprietary assets give the platform a competitive advantage? | Key Resources |
| How do collaborations with external entities enhance the platform's offerings? | Key Partnerships |
| What support mechanisms ensure user adoption and retention? | Customer Relationship |
| Through what distribution channels are services delivered to customers? | Channels |
| Who are the target users and what do they have in common? | Customer Segments |
| What pricing models generate revenue for the platform? | Revenue Streams |
| What fixed or variable costs determine the platform's profitability? | Cost Structure |

In both phases, we used open and axial coding to categorize qualitative data into concepts (e.g., '*pay-as-you-go*') and themes (e.g., '*pricing model*') accordingly. The themes from axial codes were systematically evaluated against the definitions of Business Model Canvas components in the meta-model's second order layer using predefined guiding questions (see Table 9), which served as reference dimensions for our business model taxonomy. This  ensured the final taxonomy accurately aligned with the meta-model of MLaaS platform business model.  In this study, multiple data sources were used to ensure triangulation, including platform documentation from official websites and whitepapers that provided insights into features and business models; and external publications from news and industry reports. Further details about the cases in the database were included in Appendix B. We further prepared a Supplement document to include data sources used in taxonomy development, evaluation and analysis phases (Section 1) and codebook for each of the 24 cases (Section 2).

## 3.3 Taxonomy Development Phase

The development of the MLaaS platform business model taxonomy followed Nickerson et al.'s (2013) taxonomy development methodology. The process alternated between conceptual-to-empirical (C2E) iterations and empirical-to-conceptual (E2C) iterations to refine the business model taxonomy until the objective and subjective ending conditions were fully met (see Table 10 and Appendix D). In each iteration, we applied four different taxonomy operations, which included adding, adapting, merging, and deleting the taxonomy dimensions and characteristics (see Appendix C1 for detailed operations in each C2E iterations and Appendix C2 for detailed operations in each E2C iterations).

*Table 10: Objective and subjective ending conditions (Nickerson et al., 2013)*

| Objective | Subjective |
|---|---|
| 1. All objects, or at least a representative sample of objects have been examined. | 1. The taxonomy is concise: the taxonomy dimensions are meaningful, but not overwhelming. |
| 2. No object was merged with a similar object or split into multiple objects in the last iteration. | 2. The taxonomy is robust: the dimensions and characteristics of the taxonomy can sufficiently differentiate the objects of interest. |
| 3. At least one object is classified under every characteristic of every dimension. | 3. The taxonomy is comprehensive: the taxonomy can be used to classify all objects within the research domain. |
| 4. No new dimensions or characteristics were added in the last iteration. | 4. The taxonomy is extendible: new dimensions or characteristics can be easily added. |
| 5. No dimensions or characteristics were merged or split in the last iteration. | 5. The taxonomy is explanatory: the identified dimensions and characteristic can be used to explain an object of interest. |
| 6. Every dimension is unique and not repeated (i.e., there is no dimension duplication). | |
| 7. Every characteristic is unique within its dimension (i.e., there is no characteristic duplication within a dimension). | |
| 8. Each cell (combination of characteristics) is unique and is not repeated (i.e., there is no cell duplication). | |

### 3.3.1 Design Iterations

#### *3.3.1.1 Conceptual-to-empirical iterations (C2E)*

The C2E phase applied a deductive approach, starting with taxonomy dimensions and characteristics derived from literature identified in rapid review. The taxonomy C2E phase went through three iterations, each refining the taxonomy by integrating literature insights while ensuring alignment with meta-characteristic of the taxonomy and meta-model of MLaaS platform business model. Each iteration introduced new dimensions and characteristics from literature after being adapting to the MLaaS platform context (See Appendix C1). The first iteration (C2E-1) drew from AI service platform literature (Geske et al., 2021; Ejsmont et al., 2024; Sundberg et al., 2022), prioritizing literature that demonstrated applicability to MLaaS platform domain (see Appendix C1.1). The second iteration (C2E-2) incorporated AI startup business model literature (Lins et al., 2021; Metelskaia et al., 2018; Weber et al., 2022) (see Appendix C1.2). The third iteration (C2E-3) integrated insights from platform business model and data-driven business model literature (Naous et al., 2017; Freichel et al., 2021) (see Appendix C1.3). The preliminary taxonomy was then systematically tested against four real-world MLaaS platform cases, that each case represented a distinct case segmentation identified during case database construction. The business model characteristics of these cases were extracted using open and axial coding, with the categories and themes compared against the preliminary taxonomy dimensions and characteristics (see Appendix C1.4). The identified gaps led to further refinements, resulting in the final 17 dimensions and 64 characteristics of the preliminary business model taxonomy (see taxonomy explanation in Chapter 5.4 and iteration details in Appendix C1.5).

### 3.3.1.2 Empirical-to-conceptual iteration (E2C)

The E2C phase applied an inductive approach, analysing real-world cases to derive and refine taxonomy dimensions while validating those established in the earlier phase. The taxonomy E2C phase went through three iterations, each refining the taxonomy with insights drawn from the interpretation of empirical observations in case data (see Appendix C2). Each iteration introduced new dimensions and characteristics that were aligned with our meta-model. The first E2C iteration aimed to establish the preliminary taxonomy dimensions by examining cases representing identified segmentations identified in the case database (see Appendix C2.1 for details). Four purposively-selected platforms, covering specialized platform for machine learning and Edge AI (Barbara), general-purpose platform for machine learning (Domino Data Lab), and platform bundled with cloud provider (Amazon SageMaker). This iteration focused on identifying the differences between: (1) specialized and general-purpose platforms, and (2) standalone and integrated business models. The second E2C iteration explored the boundary cases and emerging characteristics to test and expand the taxonomy (see Appendix C2.2 for details). Cases were selected to examine niche-focused platform (LLM-focused *Modular* and pharmaceutical industry-focused *CodeOcean*), platform integrated into an existing data platform ecosystem (*Databricks Mosaic AI*), and platform integrating automated workflows (ZAMS). This iteration paid particular attention to platform governance structures and pricing models that began emerging in the first iteration but required further validation. The third E2C iteration focused on cases that could improve the taxonomy's robustness and completeness (see Appendix C2.3 for details). This iteration emphasized on the taxonomy's ability to classify platform supporting both machine learning and LLM. This included platforms focusing on both machine learning and LLM which are general purpose *Dataloop*, enterprise-focused *Iguazio* and *PrimeHub*, cloud provider's bundled MLaaS platform *Google Vertex AI*. The analysis also paid special attention to the mutual exclusivity between the characteristics of each dimension. Lastly, when the final iteration fulfilled both objective and subjective ending conditions (Nickerson et al.,2013), we stopped the taxonomy development process (see Appendix D for progressive status of ending conditions) and proceeded with taxonomy evaluation phase.

### 3.3.2 Inductive Content Analysis

We employed Elo and Kyngäs' (2008) inductive content analysis in our study because it accommodated both theory development from fragmented knowledge, and interpretivism's emphasis on meaning-making. This precisely addressed our need to understand emerging MLaaS platform phenomena. Unlike Mayring's (2014) structured approach or Krippendorff's (2018) quantitative focus, In each design iteration, we followed a three-phase process (preparation, organizing, reporting) to capture both explicit features and socially constructed meanings within platform ecosystems based on the focus of that iteration(see Figure 6).

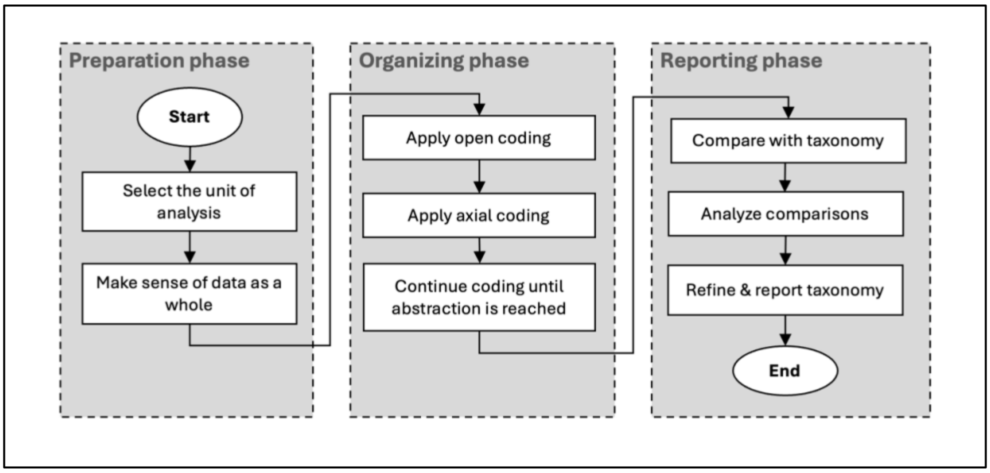


*Figure 6: Inductive content analysis (Adapted from Elo & Kyngäs, 2008)*

Overall, the inductive method proved particularly valuable when analysing edge cases. For example, some platforms that extended MLaaS features with AutoML as secondary value proposition, required careful interpretation to avoid forced categorization. By remaining grounded in the data rather than imposing predefined structures, the analysis captured these edge characteristics organically which was a key advantage over deductive approaches (Elo & Kyngäs, 2008).

## 3.4 Taxonomy Evaluation Phase

The taxonomy evaluation design in this study adopts an internal evaluation approach with predefined criteria to evaluate the MLaaS platform business model taxonomy(see Table 11). This choice was supported by the emergent nature of MLaaS platforms, where the absence of established domain experts made external validation unfeasible (Szopinski et al., 2019). This phase directly addressed the sub-question (**SQ 3**): "*To what* *extent is the business model taxonomy of MLaaS platform structurally valid and practically usable?*" Following the interpretivist principles, we conducted within-case analysis following a four-phase process(see Figure 7). The internal evaluation approach combined a structural validation of business model taxonomy dimensions and characteristics (Szopinski et al., 2019) with usability assessment of practical application of the taxonomy (Miles and Huberman, 1994).

*Table 11: Taxonomy evaluation design (Adapted from Szopinski et al., 2019)*

| Category | Criteria | Description |
|---|---|---|
| WHO?<br>(Subject of evaluation) | Involvement in taxonomy building | Subject has been involved in taxonomy building |
| | Background | Practitioners |
| | Experience | Domain |
| WHAT?<br>(Object of evaluation) | Type | Real-world |
| | Involvement in taxonomy building | Object has not been used in taxonomy building |
| | Coverage | Representative |
| HOW?<br>(Method) | Approach | Within case analysis |
| | Method | Case study |

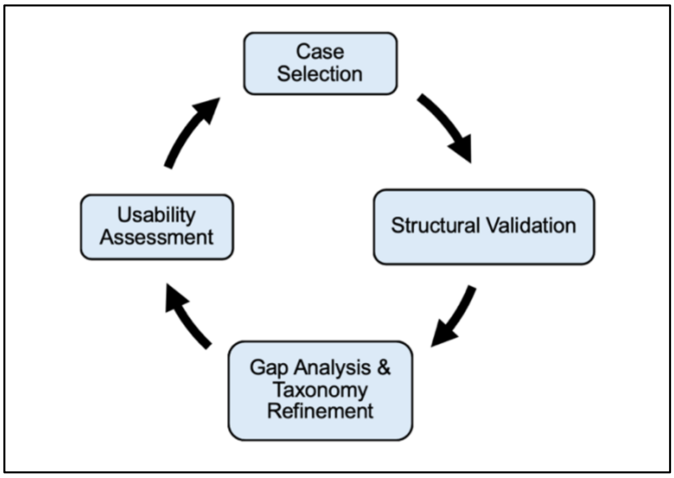


*Figure 7: Taxonomy evaluation process diagram*

We started the process with case selection guided by theoretical sampling principles to maximize variation across MLaaS platform business model characteristics. Based on three differentiating factors: model scope, industry focus, and service bundling, four cases (BigML, DataRobot, Microsoft Azure AI, and VESSL AI) were purposefully selected from our case database. In addition to the finalized business model taxonomy, the process leveraged multiple data sources, including platform documentation, product specifications, and third-party analysis, to ensure validity through triangulation. Next, we continued with structural validation where taxonomy mapping was the main analytical tool. Coding method, used in the earlier taxonomy development phase, was followed for each case, along with codebook documented as an audit trail. This phase applied Szopinski et al.'s (2019) structural validation criteria (see Table 12). We documented our findings into individual case profiles and an attributes matrix comparing all four cases.

*Table 12: Taxonomy evaluation criteria*

| Evaluation purpose | Criteria | Definition |
|---|---|---|
| Structural Validation<br>(Szopinski et al., 2019) | Completeness | All relevant business model dimensions are captured without significant gaps |
| | Mutual Exclusivity | Characteristics within a dimension are non-overlapping and clearly distinct |
| | Clarity | Definitions are unambiguous and empirically observable |
| Usability Assessment<br>(Miles and Huberman, 1994) | Utility | Enables consistent classification without expert interpretation |
| | Robustness | Handles all cases effectively |
| | Theoretical Grounding | Dimensions align with established concepts in literature |

We then proceeded with a gap analysis on the taxonomy-case mapping. The gaps revealed points of misfit between the taxonomy and empirical observations, highlighting ambiguities within the final taxonomy. We addressed the identified gaps through iterative adjustments to the business model taxonomy dimensions and characteristics. The process distinguished between necessary expansions (such as adapting Model Scope to include "*both*" as a characteristic) and clarifications (such as redefining "*key partners*" subcategories). Each modification was cross-validated against previously analysed cases to ensure backward compatibility, a critical step maintaining internal consistency as emphasized by Nickerson et al. (2013). The iterative refinement cycle continued until reaching theoretical saturation, where new cases no longer revealed substantive gaps in the business model taxonomy. Lastly, we conducted usability assessment by following Miles and Huberman's (1994) criteria on utility, robustness and theoretical grounding. This stage reflected on whether the business model taxonomy could be used effectively when being applied on regular cases like BigML and edge cases like VESSL AI. The assessment employed the same case evidence as structural validation but focused on assessing the practical usability criteria, completing the taxonomy evaluation phase as recommended by Szopinski et al. (2019).

The choice of within-case analysis over alternative evaluation methods was driven by three reasons. As Nickerson et al. (2013) observe in their taxonomy development framework, classification systems require validation at the unit level to ensure each category has clear real-world examples or observable features that clearly illustrate what each category represents. Second, the method's emphasis on contextual detail aligned with the interpretivist view that taxonomies should reflect how technologies actually exist in practice, not just how they might be abstractly conceptualized (Yanow, 2014). Finally, by focusing intensively on a few carefully selected cases rather than superficially surveying many, the analysis could uncover the unique combinations of factors that give each platform its distinctive character while still fitting within the broader taxonomy.

## 3.5 Taxonomy Analysis Phase

In the taxonomy analysis phase, we examined how cases were classified across a structured framework to uncover meaningful patterns. We used cross-case analysis to identify how different platforms align across the key dimensions. This phase addressed the sub-question (**SQ 4**): "*What business model archetypes can be derived from the recurring patterns in the MLaaS platform business model taxonomy?*"

Cross-case analysis is a key component of theory-building from case studies  and comparative technique used in qualitative research to identify patterns, similarities, and differences across multiple cases (Eisenhardt, 1989). Our primary objective was to discover meaningful groupings of platforms based on shared configurations across the taxonomy dimensions with discriminatory power. These groupings were then interpreted as archetypes that represented recurring patterns across a given set of cases. The decision to use cross-case analysis was guided by our considerations to interpret the business model taxonomy in a structured manner while remaining aligned with the interpretivist research philosophy of this study. Rather than starting with predefined categories of archetypes, the method allowed for emergent patterns to be discovered through systematic comparison through the research's interpretation and analysis.

The implementation of the cross-case analysis followed a multi-step process (see Figure 8 above). First, each of the 24 sampled cases was mapped against the business model taxonomy dimensions and characteristics. A taxonomy-case mapping was created to visually represent these configurations (see Appendix E). Next, we continued with a frequency analysis to quantify appearance of characteristics by recording their frequency of these recurring characteristics across all platforms. This revealed dominant patterns, such as the prevalence of integrated managed cloud services among *cloud orchestrators* that prioritized scalability or the emphasis on industry specialization in *niche specialists*. The dimensions were considered to have discriminatory power when natural groupings were clearly present and could be clearly identified in mutual exclusive pattern. By tallying these traits, we highlighted which features were archetype-defining with discriminatory power versus peripheral without discriminatory power, allowing initial clusters to emerge. We then refined the clusters with interpretive cluster analysis by qualitatively assessing how platforms orchestrate value creation, delivery, and capture. The findings showed the commonalities and variations across the cases, revealing how platform characteristics aligned or diverged across the MLaaS platforms. Boundary cases, such as Cloudera AI, which blended *aggregator*-like tool curation with data-related infrastructure typical in *data orchestrator*, were flagged for deeper scrutiny. These discriminatory dimensions effectively derived four distinct archetypes (*cloud orchestrator*, *data orchestrator*, *aggregator* and *niche specialist*) across the 24 platform cases while accommodating outliers. Lastly, we visualized taxonomy and the recurring patterns for inclusion in Chapter 8.

A key design choice in this process was the use of qualitative rather than statistical clustering. Unlike quantitative typologies that rely on numerical thresholds, this study emphasized interpretive pattern recognition grounded in domain knowledge. This approach was consistent with the use of archetypes in organizational research, where the focus was on identifying ideal-type configurations rather than mathematically optimal clusters (Doty & Glick, 1994). By grounding these archetypes in case documentations (see Figure 9) and supported by a systematic process ( see Figure 8), our use of cross-case analysis enabled the identification of archetypes that highlighted the recurring patterns across the MLaaS platforms.

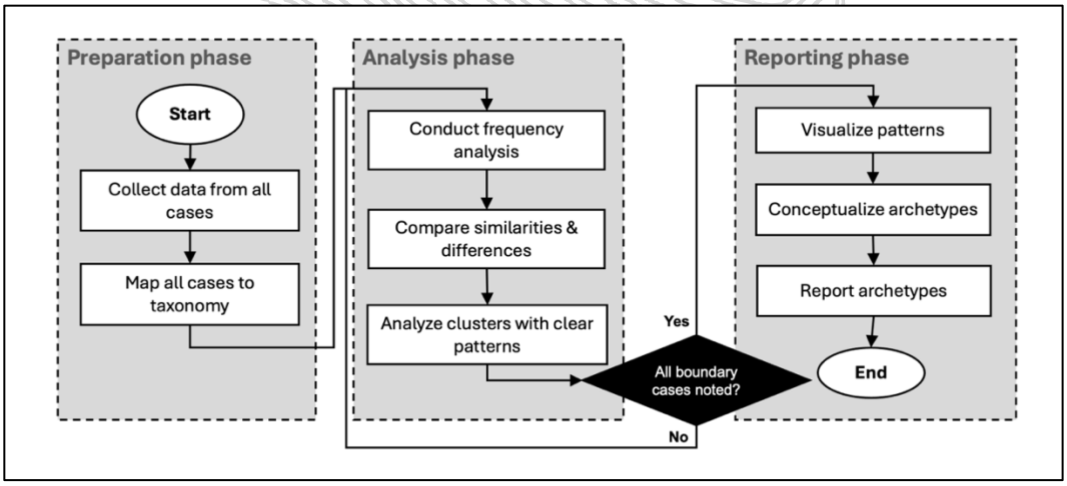


*Figure 8: Cross case analysis process*

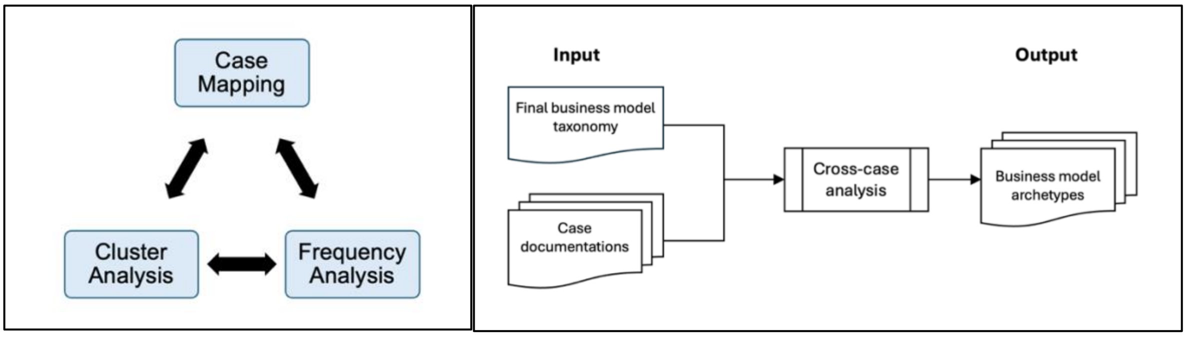


*Figure 9: Cross-case analysis' iterative nature (left) and data flow (right)*

## 3.6 Methodological Limitations

The qualitative research design adopted in this study offers a systematic and qualitative approach to developing a business model taxonomy and identifying archetypes for MLaaS platforms. However, several methodological limitations should be acknowledged.

First, the reliance on secondary data from sources such as G2 and platform websites may introduce bias or incomplete information, as these sources are often curated for marketing purposes and may not fully capture the operational nuances of each platform. To ensure diversity in the cases, the study employed purposive sampling by including platforms that capture different business model characteristics. Another source of bias is from the subjective nature of qualitative content analysis, which can be influenced by the researcher's interpretation. This was addressed by following established frameworks such as Elo and Kyngäs' (2018) inductive content analysis and by iteratively refining the taxonomy using both conceptual and empirical phases, as recommended by Nickerson et al. (2013).

The scope of the study is intentionally limited to MLaaS platforms that support the end-to-end machine learning lifecycle, which means that pre-built AI applications without customization capabilities are excluded. While this focus strengthens the relevance of the taxonomy to business models rather than technical features, it may limit the generalizability of findings to other types of MLaaS platforms. The evaluation of the taxonomy through within-case and cross-case analysis helps to validate its structural validity and practical usability, but the absence of primary data collection, such as interviews with platform providers, super users or domain experts, may restrict deeper insights into the MLaaS platforms.

Finally, the iterative process of taxonomy development and archetype derivation is inherently shaped by the framework and the predefined ending conditions. Although these methods are well-established, alternative research design or additional empirical data could yield different results. Nonetheless, careful documentation, the use of multiple validation steps, and alignment with established theoretical frameworks help to compensate for these limitations.

# 4. Conceptualizing the Business Model of MLaaS Platform

This chapter responds to the first sub-question (**SQ1**): "*To what extent can the business models of MLaaS platforms be conceptualized using current literature?*" Drawing on our scoping review of theoretical frameworks and empirical findings, this chapter presents two key theoretically-grounded propositions that forms the definition of MLaaS platforms, emphasizing the theoretical concepts that are important for the subsequent taxonomy development. Building on these propositions, a conceptual framework is introduced, integrating Teece's (2010) value creation, delivery and capture mechanisms with Osterwalder and Pigneur's (2010) Business Model Canvas through a meta-modelling approach. This meta-model serves as a reference for structuring and framing the dimensions and characteristics extracted from the analysed empirical cases. We then propose a working definition of "*MLaaS platform business model*" to be used in our taxonomy development process.

## 4.1 Conceptualizing the MLaaS Platform

This section presents two propositions that describe the characteristics of MLaaS platform: (1) *MLaaS platform is a partial subset of AI service platform* and (2) *MLaaS platform is a digital platform*. Based on these two propositions, we then formulate our definition of the MLaaS Platform.

### 4.1.1 Proposition 1: MLaaS Platform is a Partial Subset of AI Service Platform

To conceptualize MLaaS platforms, it is essential to situate them within the broader digital ecosystem, differentiate them from AI service platforms, and highlight their defining characteristics. MLaaS platforms share some characteristics of AI service platforms, or AIaaS platforms, that focus exclusively on end-to-end machine learning workflows. These platforms offer cloud-based tools and services that enable users to develop, train, deploy, and manage machine learning models without requiring in-house infrastructure or expertise (Lins et al., 2021; Philipp et al., 2020). Common themes across the definitions identified in our scoping review (see Table 13) highlight the following elements: (1) Cloud-based access to AI or machine learning capabilities without in-house infrastructure. (2) Models and tools for building, training, and deploying machine learning models. (3) Simplified development with reduced data storage and resource requirements. (4) Modular, service-oriented architecture that enables machine learning service exchange.

*Table 13: Overview of definitions in literature*

| Concept | Definition | Source |
|---|---|---|
| AIaaS or MLaaS | "AIaaS offers its users—primarily business customers - access to state-of-the-art AI capabilities, without the need for volumes of training data, expensive computational resources or lengthy development timescales that are generally required by traditional 'in-house' machine learning development." | Lewicki et al. (2023) |
| | "AIaaS refers to pre-trained models provided to customers on a commercial basis." | Cobbe & Singh (2021) |
| | "AIaaS as cloud-based systems providing on demand services to organizations and individuals to deploy, develop, train, and manage AI models." | Lins et al. (2021) |
| | "AI services are – regardless of their specific technological implementation – capable of performing cognitive functions." (Hofmann et al., 2020; Rai et al., 2019). | Hofmann et al. (2020) and Rai et al. (2019), as cited in Geske et al. (2021). |
| | "MLaaS is a paradigm allowing users to utilize machine learning without managing computing resources." | Philipp et al. (2020) |
| AI Service Platform | AI service platforms enable "the provisioning and exchange of AI services between agents," supporting AI application creation/use via modular architectures. | Geske et al. (2021) |
| MLaaS Platform | MLaaSP (MLaaS-Platforms) provide "the ability to build, train, and deploy machine learning models in a single toolset." | Philipp et al. (2020) |

AI service platforms encompass a wide array of cloud-based solutions for deploying and managing AI capabilities, including natural language processing, computer vision, and rule-based systems (Geske et al., 2021). These platforms use modular architecture that abstracts technological complexity, enabling accessibility for both technical and non-technical users. Geske et al. (2021) define them as systems that "*provide organizations with* *access to AI technology... through federating and coordinating constitutive agents*." This positions AI service platforms as intermediaries that create value through scalable, interoperable services.

Within this broad category, MLaaS platforms are distinguished by their deep focus on machine learning workflows (see Figure 10). Unlike general purpose AI service platforms that may also offer pre-built APIs, MLaaS platforms target users who require control over model development and deployment (Philipp et al., 2020). Platforms such as Amazon SageMaker and Google Vertex AI provide tools for data preprocessing, model lifecycle management and operations, enabling the user control to customize their work at granular level. According to Lins et al. (2021), MLaaS platform spans multiple AIaaS layers: *infrastructure* (e.g., GPU clusters), *developer* (e.g., TensorFlow), and *software services* (e.g., pre-trained models). Vertex AI, for example, integrates Jupyter notebooks with Tensor Processing Unit provisioning to offer a vertically integrated solution, unlike general AIaaS platforms like OpenAI's GPT-4, which offer limited configurability (Lins et al., 2021).

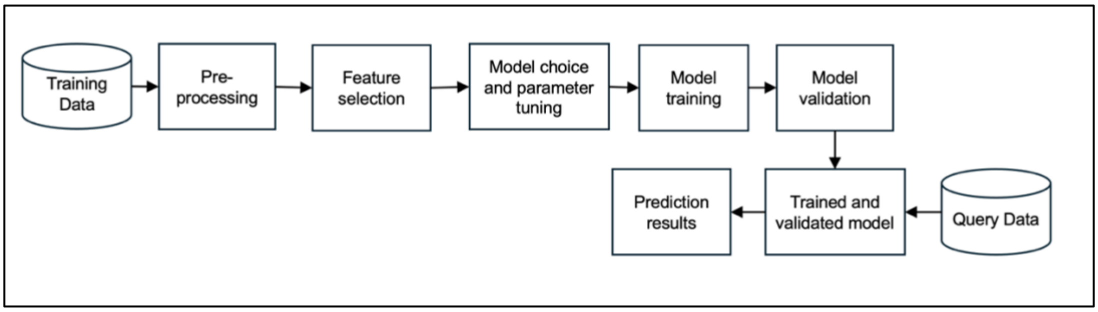


*Figure 10: Machine learning pipeline (Lins et al., 2021)*

A core differentiator is MLaaS platform's hybrid value proposition that combines software tools, infrastructure provisioning, and domain-specific expertise. The overlap and distinctions (see Table 14) between AI service platform and MLaaS platform are echoed across the literature. Philipp et al. (2020) and Lins et al. (2021) emphasize MLaaS platforms' orientation toward customizable machine learning pipelines and developer-centric tooling, whereas Geske et al. (2021) view broader AI service platforms as orchestrators of generalized AI capabilities with lower configurability. While both platform types abstract complexity and promote scalability, MLaaS platforms are uniquely positioned to support end-to-end machine learning lifecycle management. We therefore positions MLaaS platforms as a technically intensive partial subset of AI service platforms, optimized for flexible model development, fine-tuning, and deployment in cloud-integrated environments.

*Table 14: Similarities and differences between both platforms*

| Dimensions | AI Service Platform | MLaaS Platform |
|---|---|---|
| Focus | Broad AI and machine learning services (natural language processing, computer vision, etc.) | Specific machine learning workflows (data ingestion to deployment) |
| Target Users | General and specialist users | Specialist users only (e.g., data scientists, machine learning engineers) |
| Configurability | Low (standardized APIs, pre-built services) | High (granular control over service configuration) |

### 4.1.2 Proposition 2: MLaaS Platform is a Digital Platform

MLaaS platforms extend characteristics inherited from cloud services by operationalizing key digital platform concepts, including *multisidedness*, *economies of scale*, and *modularity*. These platforms support scalable, flexible, and collaborative ecosystems for machine learning development and deployment. Unlike conventional cloud infrastructures, MLaaS platforms orchestrate interactions among multiple user groups, including data scientists, developers, enterprises, and third-party tool providers, creating a dynamic environment for value co-creation (Gawer, 2014; de Reuver et al., 2018). This section places MLaaS platforms within digital platform literature by applying *multisided platform theory*, *economies of scale*, and *modularity* to explain their distinct characteristics.

MLaaS platforms are best understood through *multisided platform theory*, which frames platforms as intermediaries facilitating direct interactions between distinct user groups (Rochet & Tirole, 2003; Hagiu & Wright, 2015a). Hagiu and Wright (2015a) highlight two core features: users engage in direct interactions while retaining control over key terms (e.g., pricing, service quality), and users must make platform-specific investments (e.g., API integration). MLaaS platforms exemplify these features by connecting model developers, third party partners, and users. Snowflake's application, data and model marketplace, for instance, enable developers and providers to provide pre-trained models, data and completed machine learning-based solutions that enterprises can procure and fine-tune for their own use, creating cross-side network effects where value grows with user participation (Rochet & Tirole, 2003). This structure distinguishes MLaaS platforms from vertically integrated SaaS platforms, which typically control the entire value chain. Instead, MLaaS platforms allow third-party contributors to shape service delivery and pricing (Hagiu & Wright, 2015a). Gawer (2014) emphasizes that platforms thrive by fostering third-party innovation within federated ecosystems. However, *multisided platform* characteristics are not universal. Geske et al. (2021) note that while some AI service platforms limit third-party roles, MLaaS platforms often enable integration through modular features like pre-trained models or complementary third party services, reinforcing their multisided architecture.

MLaaS platforms achieve *economies of scale* by centralizing computational infrastructure and aggregating data across users. Cloud providers like Amazon SageMaker and Google Vertex AI lower marginal costs by sharing high-performance resources, such as GPU clusters, across multiple users. This shared infrastructure allows dynamic scaling, stronger security, and consistent service delivery that individual firms would struggle to match. As these platforms grow, some of the platforms, such as Amazon SageMaker and Google Vertex AI, with proprietary machine learning models could collect increasingly diverse datasets, which enhance model training and performance through “*data network effects*” (Gregory et al., 2021). More users mean better data, which in turn strengthens the platform’s models and services. Because machine learning models are information goods with near-zero marginal costs (Shapiro & Varian, 1999), platforms can scale rapidly by offering reusable APIs, pre-trained models, and shared tools. This enables faster innovation and knowledge transfer across the ecosystem. Yet, despite distributed value creation, platform owners capture most of the value, exemplifying what Gawer (2020) calls the paradox of centralized value capture in open digital ecosystems.

Finally, *modularity* enables flexibility and adaptability, allowing platforms to quickly respond to changing user needs and technological advances. MLaaS platforms decompose the machine learning lifecycle into interoperable components (e.g., training, deployment), meaning each part can work seamlessly with others through standardized interfaces (Ribeiro et al., 2015). This interoperability makes it easier for users to substitute or upgrade individual modules. As a result, innovation is fostered, since developers can integrate new tools or services without overhauling entire systems. However, platform openness can be different, indicating a varying level of freedom of how external parties can access, modify, or build on platform components (Geske et al., 2021). For example, Microsoft Azure AI balances modular openness with strong governance, using boundary resources such as APIs, SDKs, and compliance tools to manage interactions between the platform and external developers (Ghazawneh & Henfridsson, 2013). These resources define the platform’s boundary conditions, enabling collaboration while maintaining control and security.

## 4.2 Adopting a Hierarchical view in the Conceptual Framework

This section introduces a hierarchical framework we use in this study to analyse the business models of MLaaS platforms by drawing on established business model theory. First, we use Osterwalder et al.’s (2005) three-level Business Model Conceptual Hierarchy in our conceptual framework design. We then explain how we apply Teece’s (2010) value creation, delivery, and capture mechanisms concept on the Business Model Canvas (Osterwalder & Pigneur, 2010) as a practical lens for classifying the core components of business model. Together, these frameworks support a systematic, theory-driven analysis of MLaaS platform business models using a meta-model. By adopting this conceptual framework with a meta-model, we bridge the theoretical clarity of Teece’s framework with the empirically observable components of the Business Model Canvas, enabling systematic classification of MLaaS platform business models in the subsequent phases.

### 4.2.1 A Hierarchical Conceptual View

Osterwalder et al. (2005) propose a three-level Business Model Conceptual Hierarchy for conceptualizing business models, which serves as a useful framework for structuring our analysis of MLaaS platforms (see Figure 11). At the first and most abstract level, the *business model concept* offers a general meta-model that defines the essential elements of how a firm creates, delivers, and captures value. This level is theoretical and serves to build a shared vocabulary. The Business Model Canvas is a practical example of this meta-model, comprising nine components such as *value proposition*, *channels*, *customer relationships*, *revenue streams*, and *key resources* (Osterwalder & Pigneur, 2010). These components act as the “*building blocks*” that allow researchers to describe and compare diverse business models. For MLaaS platforms, the Business Model Canvas is particularly useful, because it captures the services, infrastructure, and partnerships these MLaaS platforms leverage to deploy and manage machine learning lifecycle. This aligns with Teece’s (2010) framework, enabling a coherent analysis of how MLaaS platforms create, delivery and capture value.

The second level, *business model types*, refers to taxonomies or classifications based on common features across firms. These types help identify patterns, allowing systematic comparison and theory development. For example, earlier research by Timmers (1998) and Rappa (2001) categorized online business models into types like subscription or brokerage. We apply a similar logic to derive archetypes from the taxonomy of MLaaS platforms we develop in this study. The third level, *business model instances*, refers to firm-specific applications of these models, similar to the platform cases in our study. Studying real-world examples like Google Vertex AI, provides practical insights and helps validate the conceptual and typological layers of the framework.

We adopt Osterwalder et al.'s (2005) hierarchical view (see Figure 11) to structure our approach to classifying MLaaS platforms because it ensures clarity in the level of analysis and consistency in conceptualization. Focusing solely on instances (e.g., analysing Amazon SageMaker) might obscure broader patterns applicable across multiple platforms, while an exclusive focus on high-level meta-models might lack practical relevance. Therefore, by distinguishing between abstract concepts, taxonomies, and concrete implementations, we avoid confusion and oversimplification.

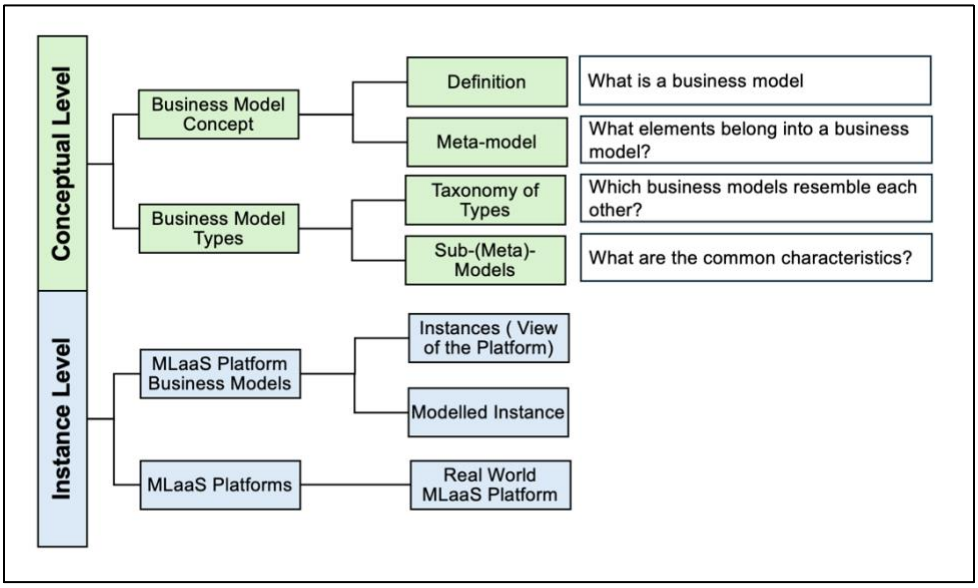


*Figure 11: Business Model Concept Hierarchy (Osterwalder et al., 2005)*

### 4.2.2 A Meta-model of MLaaS Platform Business Model

As proposed in Osterwalder et al.'s (2005) Business Model Concept Hierarchy, we construct a meta-model that inform us the components that belong to a business model (see Table 15). First, we frame Osterwalder and Pigneur's (2010) nine Business Model Canvas components using Teece's (2010) view to understand how the different components of MLaaS platform business model that contribute to the value creation, delivery and capture mechanisms (see Figure 12). This view adds a layer of high-level interpretation to the Business Model Canvas components, providing a structured theoretical lens to analyse the value creation, delivery and capture mechanisms of the MLaaS platforms.

*Table 15: A synthesis for the meta model*

| First order construct | Second order construct | Definition(s) |
|---|---|---|
| *Value creation* | | *Values offered and ways to produce them* |
| | Value Proposition | The core values offered by platform services |
| | Key Activities | The technical and operational processes underpinning the platform, such as data preprocessing, model training, and API maintenance |
| | Key Resources | Proprietary assets that differentiate the platforms |
| | Key Partnerships | Collaborations with external parties to enhance |
| *Value delivery* | | *Ways to reach customer and sustain engagement* |
| | Customer Relationship | Support structures (e.g., developer communities, service level agreements (SLAs)) that ensure adoption and retention. |
| | Channels | Distribution methods. |
| | Customer Segments | Target users (e.g., data scientists, SMEs) and their specific needs, which shape delivery. |
| *Value capture* | | *Ways to monetize and sustain profits* |
| | Revenue Streams | Pricing models and revenue sources |
| | Cost Structure | Fixed/variable costs that determine profitability |

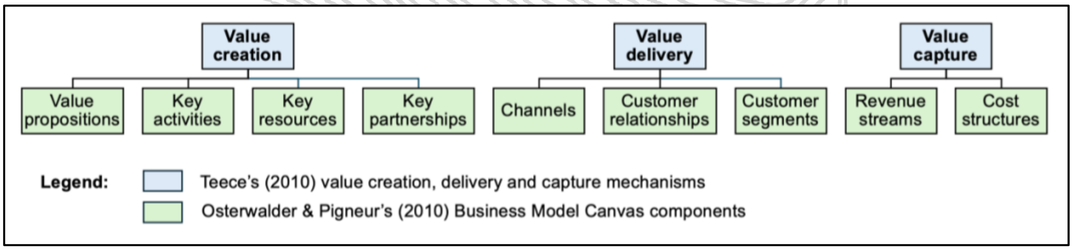


*Figure 12: Mapping Teece's (2010) and Osterwalder & Pigneur's (2010) frameworks*

According to Teece (2010), *value creation* refers to the development of innovative products, services, or business models that meet unmet customer needs or improve existing solutions. Cronk and Fitzgerald (1999) define value as the perceived worth or desirability of an object. Thus, this process of creating value relies on a firm's ability to leverage its resources (e.g., technology, knowledge, and partnerships) to produce unique offerings. That includes the firm's ability to enhance further the services they offer, develop new services, and enter new markets. The four Business Model Canvas components: *key partners*, *key activities*, *key resources*, and *value proposition,* align with value creation mechanism. For MLaaS platforms, value creation involves leveraging platform services, scalable cloud infrastructure, and partnerships to provide value proposition that reduce the need for in-house AI expertise (Lins et al., 2021). Platforms like Google Vertex AI and Microsoft Azure AI create value by offering pre-trained models, automated workflows, and collaborative environments that accelerate AI and machine learning adoption.

*Value delivery* involves the mechanisms through which a firm ensures its offerings reach customers efficiently and effectively (Teece, 2010). For MLaaS platforms, this translates to three Business Model Canvas components: *channels*, *customer relationships*, and *customer segments*. The architecture of these platforms embeds value delivery mechanisms intrinsically. For example: Amazon SageMaker's one-click deployment interface simplifies the operationalization of their value delivery model (Geske et al., 2021), while IBM Watsonx's hybrid cloud strategy, supporting model deployment to both multi-tenant and on-premise private cloud environment, ensures that the industries with data sovereignty requirements can adopt AI and machine learning. Teece (2010) emphasizes effective value delivery hinges on aligning technical capabilities (e.g., low-latency inference) with customer expectations, a challenge addressed through modular service designs that support diverse use cases.

Finally, *value capture* refers to a platform's ability to monetize its innovations and retain a portion of the value created. Teece (2010) highlights that value capture depends on strategic assets (e.g., proprietary algorithms, data), pricing models, and competitive positioning. This mechanism links to the Business Model Canvas' *revenue streams* and *cost structure*. MLaaS platforms employ diverse value capture mechanisms: freemium pricing models create entice user adoption, while ecosystem lock-in due to user's dependency on platform-specific tools and resources (e.g., compute and data storage resources), ensures long-term user retention (Lins et al., 2021).

## 4.3 Towards a Working Definition

Building on the definition of MLaaS platforms as cloud-based digital platforms that facilitate modular, scalable machine learning workflows (Philipp et al., 2020; Lins et al., 2021), we define the "*MLaaS platform business model*" as:

> *"A conceptual representation that describes the components of a MLaaS platform that create, deliver, and capture value by providing on-demand, modular and scalable machine learning services to different user groups through shared cloud infrastructure, and multisided interactions, while monetizing using differentiated pricing models, partnerships, and resources."*

This definition integrates the key components defined in Teece's (2010) value creation, delivery and capture mechanisms and Osterwalder and Pigneur's (2010) Business Model Canvas components, reflecting the meta-model adopted in this study. It aligns with prior literature (Gawer, 2014; Rochet & Tirole, 2003; de Reuver et al., 2018; Hagiu & Wright, 2015a) while emphasizing the unique value propositions that distinguish MLaaS platforms from generic AI services (Geske et al., 2021). This high-level definition sets the stage for the subsequent taxonomy development, where we systematically classify business models based on their value creation, delivery and capture mechanisms.

# 5. Conceptualizing the Preliminary Business Model Taxonomy

This chapter begins the process of addressing the second sub-question (**SQ2**) by applying the literature review findings in the conceptual-to-empirical phase of taxonomy development. We first generate the preliminary business model taxonomy which we refine in the subsequent empirical-to-conceptual phase. The literature review reveals three common research scopes amongst existing literature: (1) studies related to AI service platform (Ejsmont et al., 2024; Geske et al., 2021; Sundberg et al., 2022), (2) AI-related business model studies (Lins et al., 2021; Metelskaia et al., 2018; Weber et al., 2022), and (3) platform and data-oriented business model taxonomies (Bartels et al., 2023; Freichel et al., 2021; Staub & Haki., 2021; Täuscher & Laudien, 2017; Engelbrecht et al., 2016; Naous et al., 2017). In the first iteration, we use the findings from the literature directly related to MLaaS platforms while synthesizing insights from other AI-related literature such as AI startup and AI solutions in the following iteration. In the last iteration, we use platform and data-oriented business model literature to identify alignment or contradiction that can support adjusting the preliminary taxonomy design.

## 5.1 Business Model Dimensions from AI Service Platforms

The three studies directly related to AI service platform business models (Ejsmont et al., 2024; Geske et al., 2021; Sundberg et al., 2022) used different artifacts, which are typology, taxonomy, and business model canvas design, as tools to conceptualize business models of AI service platform. We identify several dimensions that can describe business model characteristics that are applicable for MLaaS platforms and shortlist them for inclusion in preliminary business model taxonomy (see Table 16).

Under *value creation*, the dimensions focus on the value offered by MLaaS platforms to their users and the way value is created. The *Capability* dimension, sourced from *Technology* in Sundberg et al. (2022), distinguishes platforms based on their primary service types: *analytics*, *learning*, *conversational*, and *distributed*. This classification aligns with the different degrees of human involvement required in capturing the value of the final AI solution. Similarly, *Service model*, derived from Geske et al.'s (2021) work, captures the nature of offerings, such as deployment services or ready-to-use solutions, which determine how value is created for consumers. The *core value proposition* (VP), derived from Ejsmont et al.'s (2024) work, represents the primary reason consumers adopt the platform, though its refinement is pending. The *ancillary VP*, from Ejsmont et al.'s (2024) research, extends this by detailing supplementary benefits for supply-side users, such as SME-specific issue assessments or module combinations, which enhance the platform's appeal. We identify *scope* (Sundberg et al., 2022) and rephrase it into *value scope* because it reflects the differentiation amongst the platforms based on the breadth of solutions developed using the machine learning services, from narrow, *content-changing* uses to broad, *context-changing* impacts. We further include *novelty* (Sundberg et al., 2022) and change it to *innovation impact* to highlight its emphasis on the extent of disruption introduced by the user due to the platform. Adapting from *complementary resources* and *integrated complementor offering*(Geske et al., 2021), *complementary offerings* further refine value creation by specifying supplementary resources (e.g., pre-trained models) or services (e.g., data labelling) that support users in deploying AI.

For *value delivery*, the dimensions describe how platforms facilitate interactions and transactions. *Delivery channel* (Ejsmont et al., 2024) identifies interfaces like marketplaces or payment processors, where value exchange takes place. *Platform use* (Geske et al., 2021) reflects the pattern of third party integration to the platform, distinguishing between *continuous*, *on-demand*, or *hybrid* usage patterns. *Platform interaction* (Geske et al., 2021) specifies user interfaces, such as *code-based terminals* or *GUIs*, affecting accessibility and user experience. *Integrated complementors* (Geske et al., 2021) categorize third-party contributors (e.g., *selected complementors* or *crowds*), which determine the platform's ecosystem dynamics. *Industry focus* (Geske et al., 2021) indicates whether platforms cater to specific or broad range of industry sectors, affecting their customization and relevance.

Under *value capture*, the dimensions focus on monetization and innovation. *Pricing model* (Geske et al., 2021) details mechanisms like *one-off payments* or *freemium* structures, which affect affordability and adoption. *Revenue model* (Ejsmont et al., 2024) outlines income streams, such as *subscriptions* or *transaction fees*, critical for platform sustainability.

*Table 16: Dimensions and definitions from AI service platform literature*

| Dimensions | Definition | Source |
|---|---|---|
| Capability | Main type of services offered | Sundberg et al. (2022) |
| Service Model | The nature of the platform's offerings that captures how an AI service platform creates value for the consumer. | Geske et al. (2021) |
| Ancillary VP | The values experienced by supply side that extend the core value proposition for the demand side | Ejsmont et al. (2024) |
| Complementary Offering * | Types of supplementary elements offered for users to develop or deploy AI services | Geske et al. (2021) |
| Core VP | The main reason why consumer wants to use the platform | Ejsmont et al. (2024) |
| Delivery Channel * | Types of interfaces that enables value exchange with users. | Ejsmont et al. (2024) |
| Industry focus * | The extent of which the platform's product or service is bound to a specific industry | Geske et al. (2021) |
| Innovation Impact | The degree of newness and novelty introduced due to the platform capability | Sundberg et al. (2022) |
| Integrated Complementor | The stakeholders that are integrated to the platform | Geske et al. (2021) |
| Platform Interaction ** | The interface where users interact with the platform | Geske et al. (2021) |
| Platform Use | The degree of user's integration with the platform from a temporal perspective. | Geske et al. (2021) |
| Pricing Model | The pricing mechanisms of the service offered through the platform | Geske et al. (2021) |
| Revenue Model | The ways for the platform to generate revenue | Ejsmont et al. (2024) |
| Value Scope * | Breadth of applications or use cases | Sundberg et al. (2022) |

*Note. * indicates a dimension is rephrased from how it appears in the source text. ** indicates a dimension is merged from two or more dimensions that are derived from source texts.*

## 5.2 Business Model Dimensions from AI Solutions, Startups and AIaaS

In this section, we evaluate the business models of AI solutions(Metelskaja et al., 2018), AI startups (Weber et al., 2022), and AIaaS (Lins et al., 2021) through Teece's (2010) perspective on value creation, delivery, and capture mechanisms to uncover findings that can suggest new characteristics or dimensions in the preliminary taxonomy.

In terms of *value creation*, AI solutions leverage advanced capabilities such as machine learning, natural language processing, computer vision, and robotics to offer novel functionalities (Metelskaja et al., 2018). Although MLaaS platforms are based on digital infrastructure, this observation still aligns with the *capability* dimension in the preliminary taxonomy, which distinguishes platforms based on service types like MLaaS, large language model as a service, deep learning as a service. However, Metelskaja et al. (2018) highlight that AI solutions often incorporate continuous learning, where models improve over time through new data—a dynamic aspect not explicitly captured in the current taxonomy. The way data is used to create value differs greatly. Some startups depend on data provided by customers (Weber et al., 2022). This relates to the idea of complementary services, such as pre-trained models, but also shows that the importance to distinguish between different data sourcing strategies.

For *value delivery*, AI startup business models describe the value exchange through software applications in diverse formats (e.g., web, desktop, mobile; on-premise, software-as-a-service), programmable interfaces at the code level (e.g., APIs, SDKs), or simply the base technology without having a software application or programmable interfaces (e.g., codes and specific algorithms) (Weber et al., 2022). With MLaaS platform essentially a digital infrastructure, this observation only resonate with the *delivery channel* dimension (e.g., marketplaces ) if we exclude the development and hardware provision for AI's base technology from the conceptualization of this dimension. Metelskaja et al., (2018) also highlight customization levels, from standardized products to fully customizable solutions, which could refine the *platform interaction* dimension by incorporating flexibility in user interfaces. Furthermore, the widespread of B2B and industry-specific solutions amongst AI startups (Weber et al., 2022) supports the *industry focus* dimension but suggests that an industry-specific focus may influence delivery mechanisms.

Regarding *value capture*, subscription-based and pay-as-you-go models are common in AIaaS (Lins et al., 2021), aligning with the *pricing model* and *revenue model* dimensions for MLaaS platforms. However, Metelskaja et al. (2018) states that cost structures are heavily driven by R&D and talent, a factor not explicitly addressed in the preliminary taxonomy. This could inform future iterations by adding *cost structure* dimension to capture operational expenses unique to AI platforms.

## 5.3 Business Model Dimensions from Other Business Model Taxonomies

### 5.3.1 Platform Business Model Taxonomies

The analysis of digital platform taxonomies (Täuscher & Laudien, 2017; Staub & Haki, 2021; Freichel et al., 2021) reveals the platform-specific concepts to be included in the preliminary business model taxonomy for MLaaS platforms. Staub and Haki (2021) introduce *platform structure* (exchange vs. innovation) as a key differentiator, which aligns with but extends the *AI Service Model* dimension (Geske et al., 2021). Their focus on *platform openness* (closed vs. open ecosystems) further support merging *integrated complementors* (Geske et al., 2021) into *platform openness*, taking a broader perspective on the extent of third parties or complementors operating under restrictive (e.g., proprietary APIs) or self-moderated (e.g., community-driven) rules, a critical factor for platform scalability and innovation. Freichel et al. (2021) provide a comprehensive framework, emphasizing data governance and platform architecture. Their *data dependency* dimension (open vs. proprietary data) is absent in the preliminary taxonomy but essential for MLaaS platforms, suggesting the addition of *data strategy* to differentiate platforms leveraging public datasets (e.g., Snowflakes) from those relying on proprietary data (e.g., IBM Watsonx Studio). This confirms the characteristics observed amongst AI startups as highlighted by Weber et al. (2022).

### 5.3.2 Data-driven Business Model Taxonomies

The review of data-driven business model taxonomies reveals common themes that validate the alignment with the preliminary taxonomy for MLaaS platforms. Since the capabilities of data-driven digital platforms is skewed towards data analytics, we use data-driven business model to offer conceptual validations on the preliminary taxonomy without introducing or changing any dimensions. A key insight from Naous et al. (2017) is the emphasis on service delivery models (public, private, hybrid clouds) and scalability, which directly aligns with the *delivery channel* and *platform use* dimensions in the preliminary taxonomy. Their finding, that cloud-based analytics platforms enable on-demand access to tools, echoes with the importance of temporal perspective on third party integration (*platform use* dimension), that describe the extent of which the complementor remains connected to the platform to create value, and the interface design that impacts the degree of flexibility in user participation (*platform interaction* dimension) in MLaaS platforms. Engelbrecht et al. (2016) further highlight data sourcing (user-generated, third-party, proprietary) and value creation mechanisms (analytics, visualization), which resonate with the *complementary offerings* and *integrated complementors* dimensions. Their focus on interoperability and ecosystem evolution suggests that MLaaS platforms must accommodate data inputs with different contexts (*industry focus* dimension) and support third-party collaboration (*integrated complementors* dimension), as seen in Geske et al. (2021) and Ejsmont et al. (2024). The studies also stress revenue model flexibility, with Naous et al. (2017) noting subscription-based and pay-per-use models, reinforcing the *pricing model* and *revenue model* dimensions. Engelbrecht et al. (2016) add that data-driven models evolve into platform ecosystems, implying that MLaaS platforms may similarly transition from standalone tools to hubs for innovation (*value scope* dimension). This aligns with Sundberg et al.'s (2022) distinction between content-changing and context-changing MLaaS platforms.

## 5.4 Preliminary Business Model Taxonomy

After considering all the insights from each conceptual-to-empirical iterations, the completed validation at the end of conceptual-to-empirical phase led us to a refined preliminary business model taxonomy (see Figure 13). Since the taxonomy dimensions and characteristics are solely derived from academic literature, the business model taxonomy requires further refinement before it is ready for further analysis.

| | Dimension | Characteristics | | | |
|---|---|---|---|---|---|
| **Value Creation** | machine learning service model | model deployment service | model development & deployment service | model development & deployment service with ready to use services | |
| | core VP | accessibility | tailored intelligence | efficiency | adaptability |
| | model hosting | cloud | on premise | hybrid | |
| | model scope | context changing | content changing | | |
| | complementary services | specialized expertise | automation | governance & compliance | |
| | machine learning lifecycle activities | data processing | training | monitoring | |
| | | building | validation | retraining | |
| | complementary resources | tools | ecosystem | data assets | prebuilt models |
| | assets | tangible | intangible | | |
| | data strategy | third party | customer-supplied (on-demand) | customer-supplied (continuous) | |
| | data type | numeric/sensor data | textural/ document data | natural language data | |
| | | visual data | mixed data | | |
| | infrastructure | compute | private connectivity | disaster recovery | cloud storage |
| | partnerships | technology partners | OEM partners | investors | research institution |
| | | independent software vendors | system integrators | data brokers | |
| **Value Delivery** | customer relationships | online customer portals | communities | customer support | |
| | industry focus | industry specific | industry unspecific | | |
| | interface | graphical | APIs | SDKs | |
| **Value Capture** | pricing model | usage-based pricing | tiered pricing | free trial | |
| | revenue streams | subscription | advertising | transaction | |
| | | partnership program | education | cloud infrastructure services | |
| | costs | tangible | Intangible | | |

*Figure 13: Refined preliminary business model taxonomy*

# 6. Evaluating the Business Model Taxonomy

The taxonomy evaluation phase takes place after the taxonomy fulfils the predefined ending conditions. Upon completion of taxonomy development,  we start evaluating the business model taxonomy by analysing four selected cases and iteratively refining the business model taxonomy's dimensions. Following the predefined evaluation criteria from Szopinski et al. (2019) and Miles and Huberman (1994), we use the cases to validate the taxonomy's structural validity and practical usability.

*Table 17: Diversity of selected cases*

| Case | Model scope | Industry focus | Service bundling |
|---|---|---|---|
| BigML | ML only | Broad | No |
| Microsft Azure AI | ML and LLM | Broad | Yes |
| DataRobot | ML and LLM | Enterprise | No |
| VESSL AI | ML and LLM | Regional Enterprise | No |

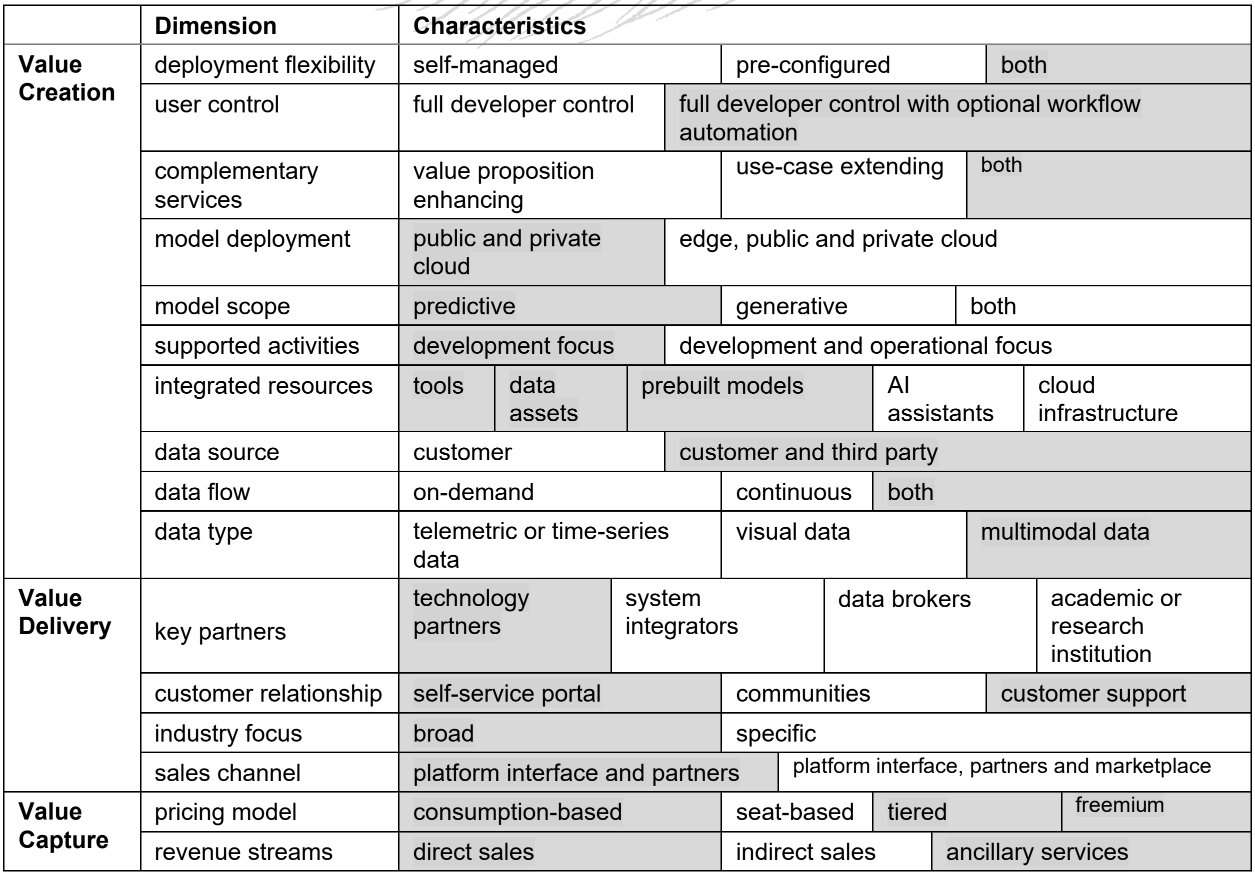
The evaluation employs a within-case analysis, by first building case profile and analysing the selected representative platforms that exhibit variation across three critical criteria: 1) model scope, 2) industry focus, and 3) service bundling.  The selected cases were chosen to reflect diverse business model characteristics within the MLaaS platforms (see Table 17). BigML represents a MLaaS platform catering to a broad customer base without affiliation with a major cloud vendor. Its taxonomy applicability is tested against its modular, user-centric approach, particularly in industries requiring data governance (e.g., healthcare). Microsoft Azure AI exemplifies an MLaaS platform integrated within a cloud ecosystem. Its inclusion tests the taxonomy's handling of platforms leveraging third-party foundational AI models  and hybrid deployment options for enterprise clients. DataRobot serves as an enterprise-focused MLaaS platform with a strong emphasis on automation and governance. Its independence from large cloud vendors allows evaluation of how the taxonomy accommodates platforms balancing predictive and generative AI use cases without tightly-integrated cloud infrastructure. Lastly, VESSL AI represents a specialized MLaaS platform targeting both academic and enterprise users. Its partnerships with cloud vendors and focus on regional customer base in Korea assess the taxonomy's coverage of niche business models.

## 6.1 BigML Case Profile and Mapping

Founded in 2011, BigML enables users to build, deploy, and manage machine learning models with minimal or extensive coding expertise, offering predictive modelling, automated workflows, and hybrid cloud deployment options. Designed for accessibility and scalability, BigML serves diverse industries, from healthcare to finance (S-5i), by combining machine learning services with transparent, interpretable prediction results. Its business model emphasizes flexible pricing, enterprise-grade security, and ecosystem integrations. A taxonomy-case mapping showcases the full manifestation of characteristics after the business model taxonomy is applied (see Figure 14).

BigML's value creation stems from its flexible technical infrastructure and comprehensive services supporting machine learning workflows. The platform supports both public and private cloud deployments, enabling enterprises to choose configurations that meet their security and scalability requirements. For instance, BigML Enterprise allows private installations on corporate infrastructure while maintaining integration with cloud-based services, addressing regulatory concerns in sectors like healthcare and finance(S-5b, see Data Source in Supplement document). This deployment flexibility is complemented by granular user control options, where developers can leverage WhizzML for workflow automation or maintain full manual oversight of the machine learning lifecycle (S-5b). The Scriptify feature exemplifies this balance by converting workflows into reusable scripts with "one-click" functionality, reducing repetitive tasks without sacrificing transparency(S-5b).

The platform's machine learning model scope focuses primarily on predictive analytics, including classification, regression, and time-series forecasting, which dominate enterprise use cases. While generative capabilities are limited, BigML compensates with depth in practical applications like customer churn prediction and inventory optimization. Model deployment options further enhance value creation by supporting hybrid environments where sensitive data can remain on-premises while leveraging cloud scalability for computation-intensive tasks. For example, models can be trained locally and deployed via BigML's cloud API for real-time predictions, blending security with performance (S-5b). Complementary services such as certifications and training programs (S-5e) lower adoption barriers, particularly for non-technical users. The platform's integrated resources create a complete ecosystem, combining an intuitive dashboard for business analysts with REST APIs for developers. Data management capabilities accommodate both customer-uploaded datasets and third-party sources, with support for multimodal data types including structured, time-series, and textual data. The public model gallery demonstrates this integration(S-5g), offering prebuilt assets that accelerate prototyping while maintaining customization options.



| | Dimension | Characteristics | | | | |
|---|---|---|---|---|---|---|
| **Value Creation** | deployment flexibility | self-managed | pre-configured | both | | |
| | user control | full developer control | full developer control with optional workflow automation | | | |
| | complementary services | value proposition enhancing | use-case extending | both | | |
| | model deployment | public and private cloud | edge, public and private cloud | | | |
| | model scope | predictive | generative | both | | |
| | supported activities | development focus | development and operational focus | | | |
| | integrated resources | tools | data assets | prebuilt models | AI assistants | cloud infrastructure |
| | data source | customer | customer and third party | | | |
| | data flow | on-demand | continuous | both | | |
| | data type | telemetric or time-series data | visual data | multimodal data | | |
| **Value Delivery** | key partners | technology partners | system integrators | data brokers | academic or research institution | |
| | customer relationship | self-service portal | communities | customer support | | |
| | industry focus | broad | specific | | | |
| | sales channel | platform interface and partners | platform interface, partners and marketplace | | | |
| **Value Capture** | pricing model | consumption-based | seat-based | tiered | freemium | |
| | revenue streams | direct sales | indirect sales | ancillary services | | |

*Figure 14: Taxonomy-case mapping of BigML*

BigML delivers value through strategic partnerships and multi-channel customer engagement. Technology partnerships with cloud providers and integration tools like Zapier extend platform functionality, while collaborations with academic institutions enhance research credibility (S-5h). These alliances create an ecosystem where users can combine BigML's core capabilities with specialized third-party services, as seen in Node-RED integrations for Internet of Things (IoT) applications. Customer relationships are cultivated through self-service portals, community forums, and dedicated customer support channels (S-5f). Enterprise clients receive additional value through personalized onboarding and SLAs, ensuring smooth adoption at scale. BigML's horizontal market focus enables broad applicability across industries, from retail demand forecasting to financial risk modelling, while sales channels leverage both direct platform access and partner networks for extended reach.

The platform employs a diversified monetization approach that aligns with user needs and value perception. A freemium tier with limited resources serves as an entry point, while consumption-based pricing and tiered subscriptions (Prime/Enterprise) cater to growing usage demands. This structure is evident in BigML's pricing page (S-5d), where plans scale with dataset size and parallel task capacity, allowing businesses to pay for precisely the resources they require. The freemium model with a 14-day trial period allows users to evaluate the platform before committing a subscription tier. Revenue streams combine direct platform subscriptions with high-value ancillary services. Enterprise licenses for private deployments command premium pricing, while certification programs and consulting services create additional monetization opportunities. The absence of indirect sales underscores BigML's focus on maintaining service quality through direct customer relationships, though marketplace features hint at future potential for ecosystem-driven revenue sharing.

## 6.2 Microsoft Azure AI Case Profile and Mapping

Microsoft Corporation, a global technology leader, has expanded its cloud computing offerings through Microsoft Azure, a comprehensive cloud platform providing infrastructure, and software services. Azure AI, a key component of Azure and a MLaaS platform as well, delivers machine learning services, enabling enterprises to build, deploy, and scale machine learning efficiently. This case examines Azure AI's business model through value creation, delivery, and capture mechanisms. A taxonomy-case mapping showcases the full manifestation of characteristics after the business model taxonomy is applied (see Figure 15).

| | Dimension | Characteristics | | | | |
|---|---|---|---|---|---|---|
| **Value Creation** | deployment flexibility | self-managed | pre-configured | both | | |
| | user control | full developer control | full developer control with optional workflow automation | | | |
| | complementary services | value proposition enhancing | use-case extending | both | | |
| | model deployment | public and private cloud | edge, public and private cloud | | | |
| | model scope | predictive | generative | both | | |
| | supported activities | development focus | development and operational focus | | | |
| | integrated resources | tools | data assets | prebuilt models | AI assistants | cloud infrastructure |
| | data source | customer | customer and third party | | | |
| | data flow | on-demand | continuous | both | | |
| | data type | telemetric or time-series data | visual data | multimodal data | | |
| **Value Delivery** | key partners | technology partners | system integrators | data brokers | academic or research institution | |
| | customer relationship | self-service portal | communities | customer support | | |
| | industry focus | broad | specific | | | |
| | sales channel | platform interface and partners | platform interface, partners and marketplace | | | |
| **Value Capture** | pricing model | consumption-based | seat-based | tiered | freemium | |
| | revenue streams | direct sales | indirect sales | ancillary services | | |

*Figure 15: Taxonomy-case mapping of Microsoft Azure AI*

Microsoft Azure AI creates value through a sophisticated combination of technical capabilities and service offerings that address enterprise AI needs. The platform's pre-configured deployment model lowers barriers to entry by providing managed services that abstract infrastructure complexities, while maintaining flexibility through hybrid cloud and edge deployment options. This is exemplified by Azure Arc's ability to extend AI services to on-premises environments and edge devices, crucial for latency-sensitive applications like industrial IoT predictive maintenance (S-11f). The platform delivers substantial value through its user control paradigm, offering full developer control with optional workflow automation. Data scientists leverage Azure ML Studio for end-to-end workflow management, while business analysts utilize AutoML features for streamlined model development. This dual approach is demonstrated in customer implementations where financial institutions use AutoML for fraud detection while maintaining manual control over model refinement.

Azure AI's machine learning *model scope* encompasses both predictive and generative capabilities, creating value across traditional and emerging AI use cases. The integration of Azure OpenAI Service provides access to cutting-edge language models, enabling applications ranging from document processing to conversational AI (S-11d). Retail customers have deployed these capabilities for personalized recommendations, while healthcare organizations use predictive models for patient outcome forecasting. Complementary services significantly enhance Azure AI's value proposition. Azure AI Search enables sophisticated retrieval-augmented generation implementations, as seen in the NBA's fan engagement platform that processes billions of data points (S-11h). Azure AI Content Safety provides critical moderation capabilities for user-generated content, valued by social media platforms operating at scale (S-11f). The platform's integrated resources create value through comprehensive tooling and integrated managed cloud services. Access to 1,700+ foundation models, including partnerships with Hugging Face and Meta, accelerates time-to-value (S-11a). Managed cloud services handle resource provisioning and scaling, while Copilot's AI-assisted development increases productivity.

Azure AI delivers value through an extensive partner ecosystem and multiple customer engagement channels. Technology partners like NVIDIA provide specialized hardware acceleration, while system integrators facilitate enterprise deployments (S-11d). Academic collaborations, such as those with research institutions, foster innovation in areas like responsible AI. Customer relationships are maintained through in a few ways. The self-service Azure AI Studio portal empowers independent developers to troubleshooting their issues on their own, while enterprise customers benefit from dedicated technical support and professional services. Microsoft's Cloud Solution Architect program exemplifies this, providing tailored guidance for complex implementations like PIMCO's financial analytics platform (S-11b). The platform serves both horizontal and vertical markets through adaptable solutions. Industry-agnostic capabilities like computer vision and natural language processing are complemented by specialized offerings for sectors like healthcare and finance. The case study where McDonald's China deployment demonstrates this versatility, demonstrated how Azure AI transformed customer experiences while meeting industry-specific requirements (S-11a). Sales channels combine direct and indirect approaches. The Azure Marketplace extends platform capabilities through third-party solutions, while partner networks provide localized implementation expertise.

Azure AI employs sophisticated pricing models to capture value from its offerings (S-11b). The consumption-based approach aligns costs with actual usage, particularly beneficial for variable workloads. Tiered pricing accommodates users from both SMEs and large enterprises, while freemium options lower adoption barriers for startups and educational institutions. Revenue streams are diversified across direct sales, indirect channels, and ancillary services. Enterprise agreements with organizations like Unity generate substantial direct revenue, while marketplace transactions contribute indirect sales. Ancillary services including migration support and custom model development provide high-margin opportunities, as seen in Airbus's deployment where specialized consulting was required for air-gapped environments (S-11h). The platform's value capture is further enhanced through strategic pricing of premium features. Azure OpenAI Service operates on a separate consumption model that captures value from cutting-edge capabilities. Reserved capacity discounts encourage long-term commitments, optimizing Azure's resource utilization while providing customer cost certainty.

## 6.3 DataRobot Case Profile and Mapping

Founded in 2012, DataRobot is a MLaaS platform that enables organizations to build, deploy, and manage predictive and generative AI models at scale. Serving industries such as finance, healthcare, and manufacturing, DataRobot combines no-code tools with advanced customization options, helping businesses accelerate AI adoption while maintaining governance and compliance. Recognized as a leader in Gartner’s Magic Quadrant for Data Science and Machine Learning Platforms (S-22l), DataRobot supports the full machine learning lifecycle, from data preparation to model monitoring, making AI accessible to both data scientists and business users. A taxonomy-case mapping showcases the platform’s characteristics after the business model taxonomy is applied (see Figure 16).

| | **Dimension** | **Characteristics** | | | | |
|---|---|---|---|---|---|---|
| **Value Creation** | deployment flexibility | self-managed | pre-configured | both | | |
| | user control | full developer control | full developer control with optional workflow automation | | | |
| | complementary services | value proposition enhancing | use-case extending | ~~both~~ | | |
| | model deployment | public and private cloud | edge, public and private cloud | | | |
| | model scope | predictive | generative | both | | |
| | supported activities | development focus | development and operational focus | | | |
| | integrated resources | tools | data assets | prebuilt models | AI assistants | cloud infrastructure |
| | data source | customer | customer and third party | | | |
| | data flow | on-demand | continuous | both | | |
| | data type | telemetric or time-series data | visual data | multimodal data | | |
| **Value Delivery** | key partners | technology partners | system integrators | data brokers | academic or research institution | |
| | customer relationship | self-service portal | communities | customer support | | |
| | industry focus | broad | specific | | | |
| | sales channel | platform interface and partners | platform interface, partners and marketplace | | | |
| **Value Capture** | pricing model | consumption-based | seat-based | tiered | freemium | |
| | revenue streams | direct sales | indirect sales | ancillary services | | |

*Figure 16: Taxonomy-case mapping of DataRobot*

DataRobot creates value through a MLaaS platform that combines deployment flexibility with advanced machine learning capabilities. The platform supports both self-managed and pre-configured deployments, enabling enterprises to choose between full infrastructure control through hybrid cloud options (AWS, Google

Cloud, Azure) or accelerated implementation using pre-built AI agent templates. This flexibility is exemplified by DataRobot's one-click deployment templates for common use cases like time-series forecasting and its Covalent orchestration tool for workload portability across environments (S-22d). User control represents a core value proposition, offering full developer control with optional workflow automation. DataRobot enables granular management of the machine learning lifecycle while incorporating AutoML for feature engineering, as seen in its automated feature detection and reduction capabilities (S-22h). The platform's ability to process multimodal data - including time-series, textual, and visual inputs - supports diverse applications from predictive maintenance to generative AI workflows. For instance, its geospatial data prep features transform coordinates into actionable vectors, while time-aware data partitioning ensures accurate temporal modelling(S-22h).

Complementary services significantly enhance DataRobot's value creation. The platform integrates NVIDIA NeMo guardrails for real-time risk mitigation in generative AI applications and provides automated compliance documentation for regulations like the EU AI Act (S-22e). These features address critical enterprise concerns about model governance, as demonstrated by FordDirect's use of DataRobot to maintain responsible AI deployment standards (S-22a). The platform's development and operational focus covers the complete machine learning lifecycle, from data healing and duplication tools to production monitoring of custom metrics like prediction latency and data drift. Integrated resources form another key value driver, combining tools (GUI builders, Jupyter notebooks), prebuilt models (NVIDIA NeMo, time-series templates), and data assets (feature stores, synthetic data generation) (S-22h). DataRobot's AI assistant provides model explainability through SHAP values and LIME interpretations, while its managed cloud services offer auto-scaling GPU clusters for compute-intensive tasks like LLM fine-tuning (S-22h). The platform's mixed data sourcing capability allows enterprises to combine proprietary data with third-party resources, such as Hugging Face models or vector databases, creating a comprehensive analytics environment.

DataRobot delivers its value through strategic partnerships and multiple customer engagement channels. Key technology partners like NVIDIA and SAP enable hardware acceleration and enterprise system integration, while collaborations with academic institutions support open-source contributions to frameworks like Covalent (S-22g). These partnerships extend the platform's capabilities, as seen in the NVIDIA Enterprise AI Factory integration that accelerates secure agent deployment (S-22g). Customer relationships are cultivated through self-service portals for independent experimentation, user communities for knowledge sharing, and dedicated support for enterprise clients. The platform's free trials (S-22f) and open-source tools (S-22d) lower adoption barriers, while enterprise-grade support ensures production readiness. The platform primarily targets vertical customer segments with industry-specific solutions, evidenced by specialized offerings for financial services (e.g., Freddie Mac's automated underwriting models) and healthcare (e.g., Baptist Health's clinical decision support). However, its modular architecture retains broad applicability for general machine learning use cases. Sales channels combine direct platform access with partner-enabled deployments, creating multiple entry points for different customer types and use case complexities.

DataRobot employs a diversified pricing strategy combining consumption-based, tiered, and freemium models. GPU-intensive tasks utilize pay-per-use billing, while subscription plans offer predictable costs for core platform access. The freemium approach includes 14-day trials (S-22f). Open-source tools like Covalent, which according to case documents can be installed in under a minute, provide low-friction evaluation opportunities. Revenue streams are strategically diversified across direct sales (enterprise licensing), indirect channels (partner-led implementations), and ancillary services (custom application development). Professional services like compliance consulting and model governance further contribute to revenue while enhancing customer stickiness. The platform's value capture is reinforced by its partner ecosystem, where collaborations with system integrators like SAP create additional monetization opportunities.

## 6.4 VESSL AI Case Profile and Mapping

Founded in 2020, VESSL AI is an MLaaS platform that combines managed cloud infrastructure with developer control, while being headquartered in South Korea. Its platform serves enterprises, startups, and academic researchers, offering scalable, cost-efficient machine learning model development from training to production. A taxonomy-case mapping showcases the platform’s characteristics after the business model taxonomy is applied (see Figure 17).

VESSL AI creates value through its pre-configured deployment model that eliminates infrastructure management burdens while maintaining full developer control over the machine learning lifecycle. The platform's hybrid architecture enables deployment across both public clouds (AWS, Google Cloud) and private infrastructure (S-23a). This flexibility is particularly valuable for enterprises with strict data governance requirements or existing hardware investments. The platform specializes in generative AI applications, offering optimized workflows for large language models like Llama 3.2 and diffusion models such as Stable Diffusion (S-22a). VESSL Hub provides numerous ready-to-use templates that reduces setup time from days to hours(S-22a). Complementary services enhance core functionality, including academic programs that provide free high performance computing access and job scheduling that simplifies cluster management for research institutions (S-23c). VESSL AI supports the complete machine learning lifecycle from development to operations through integrated tools. Data handling capabilities focus on customer-provided sources with third-party integrations, supporting multimodal data types crucial for generative AI applications.

| | **Dimension** | **Characteristics** | | | | |
|---|---|---|---|---|---|---|
| **Value Creation** | deployment flexibility | self-managed | [✓ pre-configured] | both | | |
| | user control | [✓ full developer control] | full developer control with optional workflow automation | | | |
| | complementary services | value proposition enhancing | [✓ use-case extending] | both | | |
| | model deployment | [✓ public and private cloud] | edge, public and private cloud | | | |
| | model scope | predictive | [✓ generative] | both | | |
| | supported activities | development focus | [✓ development and operational focus] | | | |
| | integrated resources | [✓ tools] | data assets | [✓ prebuilt models] | AI assistants | cloud infrastructure |
| | data source | [✓ customer] | customer and third party | | | |
| | data flow | on-demand | [✓ continuous] | both | | |
| | data type | telemetric or time-series data | visual data | [✓ multimodal data] | | |
| **Value Delivery** | key partners | [✓ technology partners] | system integrators | data brokers | [✓ academic or research institution] | |
| | customer relationship | [✓ self-service portal] | [✓ communities] | customer support | | |
| | industry focus | Broad | [✓ specific] | | | |
| | sales channel | [✓ platform interface and partners] | platform interface, partners and marketplace | | | |
| **Value Capture** | pricing model | [✓ consumption-based] | seat-based | [✓ tiered] | [✓ freemium] | |
| | revenue streams | [✓ direct sales] | indirect sales | ancillary services | | |

*Figure 17: Taxonomy-case mapping of VESSL AI*

Next, VESSL AI delivers value through strategic partnerships that extend its technical capabilities and market reach. Technology partners like NVIDIA and Google Cloud provide hardware optimizations and specialized services, while academic collaborations with institutions such as Seoul National University foster innovation and talent pipeline development (S-23a). The platform serves clients with regional presence in a horizontal market across industries, with demonstrated use cases ranging from automotive (Hyundai's autonomous driving models) to generative AI startups (Ketolab's LLM fine-tuning). Customer relationships employ multiple engagement channels tailored to different segments. Self-service tools including the web console and CLI cater to technical users, while the VESSL community on GitHub provides peer support. Enterprise clients receive dedicated assistance, as evidenced by the custom onboarding and SLA guarantees mentioned in enterprise plan details (S-23b; S-23d). Sales channels combine the platform interface with partner networks, creating multiple touchpoints for customer acquisition and retention.

In terms of value capture, VESSL AI employs a tiered pricing model with freemium elements to capture value across market segments(S-23b; S-23d). The academic program serves as both a community investment and lead generation tool, while consumption-based billing for GPU usage aligns costs with actual value received (S-23c). Enterprise plans feature custom pricing for reserved compute instances and dedicated customer support, creating predictable revenue streams from established clients. Revenue generation focuses on direct sales supplemented by ancillary services. The case documents highlight custom invoicing options and professional services like tailored MLOps consulting as additional income sources.

## 6.5 Evaluation Result

The business model taxonomy's structural validity is assessed based on three key criteria: *comprehensiveness*, *mutual exclusivity*, and *clarity* (Szopinski, 2019). During taxonomy evaluation, the taxonomy demonstrated completeness in capturing core MLaaS platform business model characteristics across all cases. *Comprehensiveness* refers to whether the taxonomy captures all relevant dimensions of MLaaS business models. A comparison with the case profiles confirms that these 16 dimensions sufficiently describe the business models of BigML, Microsoft Azure AI, DataRobot, and VESSL AI. For instance, BigML's hybrid cloud deployment aligns with the *deployment flexibility* dimension, while Azure AI's combination of predictive and generative capabilities fits within the *model scope* dimension. No major omissions were identified, suggesting the taxonomy is comprehensive. *Mutual exclusivity* ensures that dimensions and characteristics do not overlap. The business model taxonomy largely maintains distinct boundaries between dimensions. For example, integrated resources (*tools*, *prebuilt models*, *data assets*) and complementary services (*value proposition enhancing*, *use case extending*) are logically separate, as seen in DataRobot's differentiation between its prebuilt NVIDIA NeMo models (*integrated resources*) and its automated compliance documentation (*complementary service*). Lastly, *clarity* assesses whether the dimensions and characteristics are unambiguous and well-defined. The taxonomy uses precise terminology, such as *consumption-based* versus *tiered pricing*, which are clearly distinguished in the case profiles (e.g., Azure AI's pay-per-use model vs. DataRobot's subscription tiers).

The usability of the MLaaS business model taxonomy is evaluated using predefined criteria of *utility*, *robustness*, and *theoretical grounding,* ensuring its practical applicability in analyzing real-world cases (Miles & Huberman, 1994). *Utility* refers to the taxonomy's effectiveness in describing and comparing MLaaS business models. The taxonomy demonstrates high utility by capturing key variations across platforms, such as *deployment flexibility* (*self-managed* vs. *pre-configured*) and machine learning *model scope* (*predictive* vs. *generative*). For instance, it distinguishes Azure AI's hybrid cloud deployment from VESSL AI's pre-configured model, enabling precise comparisons. Additionally, the taxonomy helps in identifying strategic differentiators, such as DataRobot's *narrow* versus BigML's *broad* industry focus, proving its value for both academic and business analysis purpose. Secondly, *robustness* assesses whether the taxonomy remains consistent across diverse cases. The framework holds well when applied to platforms with varying business strategies, from BigML's freemium model to Azure AI's tiered enterprise pricing. However, minor inconsistencies arise in overlapping dimensions, such as *data flow* and *deployment flexibility*, where continuous data processing often aligns with pre-configured services. Despite this, the taxonomy maintains strong discriminatory power, reliably classifying business model attributes without significant ambiguity. Lastly, *theoretical grounding* examines whether the taxonomy aligns with established business model literature. The dimensions are aligned with the meta model, drawing from recognized frameworks, such as Osterwalder and Pigneur's (2010) value proposition and revenue stream components, while incorporating MLaaS platform-specific elements (e.g., *model deployment* and characteristics of *integrated resources* dimension).

| | **Dimension** | **Characteristics** | | | | |
|---|---|---|---|---|---|---|
| **Value Creation** | deployment flexibility | self-managed | pre-configured | both | | |
| | user control | full developer control | full developer control with optional workflow automation | | | |
| | complementary services | value proposition enhancing | use-case extending | both | | |
| | model deployment | public and private cloud | edge, public and private cloud | | | |
| | model scope | predictive | generative | both | | |
| | supported activities | development focus | development and operational focus | | | |
| | integrated resources | tools | data assets | prebuilt models | AI assistants | cloud infrastructure |
| | data source | customer | customer and third party | | | |
| | data flow | on-demand | continuous | both | | |
| | data type | telemetric or time-series data | visual data | multimodal data | | |
| **Value Delivery** | key partners | technology partners | system integrators | data brokers | academic or research institution | |
| | customer relationship | self-service portal | communities | customer support | | |
| | industry focus | broad | specific | | | |
| | sales channel | platform interface and partners | platform interface, partners and marketplace | | | |
| **Value Capture** | pricing model | consumption-based | seat-based | tiered | freemium | |
| | revenue streams | direct sales | indirect sales | ancillary services | | |

*Figure 18: Final business model taxonomy of MLaaS Platforms*

Overall, the taxonomy satisfies the predefined structural validity (Szopinski et al., 2019) and usability criteria (Miles & Huberman, 1994). Minor refinements in dimension boundaries could further enhance robustness, but its utility and theoretical foundation remain well-supported. Thus, this leads to the conclusion of the evaluation of our final business model taxonomy (see Figure 18).

# 7. Classifying the Business Models of MLaaS Platforms

The classification of business models for MLaaS platforms requires a structured approach to capture the value creation, delivery and capture mechanisms of these platforms. This chapter explains the finalized business model taxonomy, developed through Nickerson et al.'s (2013) iterative methodology. As illustrated in Figure 19, the taxonomy organizes 16 dimensions and 46 characteristics into value creation, delivery and capture mechanisms. This chapter walks through each dimension and characteristic of the final MLaaS platform business model taxonomy primarily using Teece's (2010) role as the meta-model's first-order construct as our main theoretical lens in this study.

## 7.1 Value Creation

Value creation lies at the heart of MLaaS platform business models, encompassing the strategic orchestration of resources, activities, and partnerships to deliver machine learning services. Grounded in Teece's (2010) assertion that value creation stems from "*innovative resource orchestration to address unmet needs,*" this section examines how platforms apply their resources to create differentiated offerings for the customers. The taxonomy identifies 10 dimensions under Value Creation.

### 7.1.1 Deployment Flexibility

*Deployment flexibility* describes the range of options available to users in determining where and how they deploy the platform's infrastructure, as well as the complexity of the deployment process. This dimension captures whether the platform requires minimal setup, or demands more involved self-installation. Platforms with *pre-configured* deployment minimize setup complexity by offering fully managed, vendor-hosted infrastructure. These platforms typically integrate seamlessly with the platform owner's own cloud ecosystem or have built-in connector programmed into the platform, so users require little to no manual configuration to deploy their models. For instance, Amazon SageMaker primarily operates exclusively within AWS, allowing users to launch machine learning environments with minimal effort but restricting deployment to non-AWS cloud services. In contrast, *self-managed* deployment grants users full control over deployment options but requires manual setup of the process using APIs. Some platforms, including Dataiku, provide both deployment models, accommodating varying user needs. These hybrid solutions allow organizations to choose between a *pre-configured* deployment to fully managed cloud environment or a *self-managed* deployment to environment not fully-integrated to the platform. *Pre-configured* options lower the barrier to entry, while *self-managed* deployments cater to users with specific infrastructure requirements. By offering flexibility in deployment complexity and location, such platforms appeal to a broader range of users, from startups and SMEs requiring quick onboarding to large enterprises with stringent IT policies.

### 7.1.2 User Control

*User control* defines the extent to which users, like data scientists and developers, can customize their machine learning workflows. *Full developer control* grants users complete autonomy over machine learning lifecycle workflow at code level, from data preprocessing to deployment, typically through command line interfaces or integrated development environment. This is ideal for advanced practitioners who require fine-grained adjustments, as seen in platforms like VESSL AI, where users script every step. A hybrid approach, *full developer control with optional workflow automation*, combines flexibility with optional use of AutoML services for efficiency. AutoML in MLaaS platforms automates data preprocessing, feature engineering, model selection, tuning, or deployment, thereby enabling rapid, accessible machine learning. Dataiku exemplifies this by offering both manual coding and drag-and-drop AutoML tools, allowing users to switch between control levels based on task complexity. This spectrum of control ensures platforms can serve novices and experts alike regardless of the organization sizes.

### 7.1.3 Complementary Services

*Complementary services* dimension are services that supplement the core functionalities of MLaaS platforms, addressing adjacent user needs beyond basic model development and deployment. *Value proposition enhancing* services amplify the platform's primary value proposition, such as security features that bolster trust in model outputs or governance tools that ensure regulatory compliance. For instance, IBM Watsonx's AI fairness checker strengthens reliability for healthcare applications by embedding ethical safeguards directly into the workflow. *Use-case extending* services broaden the platform's applicability, enabling new ways of using the platform through capabilities like data labelling or integrated storage. Amazon SageMaker exemplifies this through SageMaker Ground Truth, which expands the platform's utility by providing in-house data annotation tools, eliminating dependencies on external vendors.

### 7.1.4 Model Deployment

*Model deployment* defines where trained models are hosted and executed. Platforms supporting hybrid approaches enable deployment to both *edge, public and private cloud* environment. Edge deployment enable low-latency inference by deploying models directly on hardware like sensors or smartphones, crucial for applications like fraud detection in industrial IoT use cases. *Public and private cloud* approaches, like Azure ML's support for both public and on-premise private cloud deployment to user's own data centre. The choice depends on the specific use-case requirements, such as latency tolerance or data sovereignty, of the SME or large enterprise users. Our analysis reveals a consistent pattern in cloud deployment support across MLaaS platforms: those offering cloud-based deployment universally provide both public and private cloud options, while standalone deployment types (exclusively public or private) are never observed in isolation.

### 7.1.5 Model Scope

The *model scope* dimension categorizes platforms by the types of machine learning tasks they support. *Predictive* MLaaS platforms, the most common scope (e.g., Dataiku's time-series forecasting tools), dominated early MLaaS platform offerings (Philipp et al., 2020). These platforms focus on traditional supervised learning, such as fraud detection (e.g., H2O.ai). The rise of generative AI has spurred MLaaS platforms like Google Vertex AI to integrate *generative* capabilities by supporting the development and deployment of foundational LLMs. These platforms specialize in creating new content, like text or images, using models such as GPT or Stable Diffusion (e.g., OpenAI's API). Platforms supporting both, like Databricks, provide tools for predictive and generative models, appealing to organizations with diverse AI needs.

### 7.1.6 Supported activities

The *supported activities* dimension classifies platforms by the extent their service coverage of the machine learning workflow whether it focuses on only development activities or also operational activities as well. Platforms may emphasize *development focus*, offering robust tools for model training but limited deployment features. Some platforms could have *both development and operational focus*, additionally integrating end-to-end MLOps capabilities, like model monitoring and retraining, into the platform. Services from platforms like Google Vertex AI and Dataiku cover both of the end-to-end machine learning lifecycle and subsequent operations, spanning data preprocessing, model training, inference and further activities such as monitoring, retraining and governance.

### 7.1.7 Integrated Resources

*Integrated resources* refer to built-in tools and assets that enhance productivity within a MLaaS platform. These include development *tools* like IDEs and version control systems, exemplified by SageMaker Studio's notebook interface and Git integration. *Data assets*, such as prelabelled datasets or feature stores (e.g., Databricks Feature Store), accelerate model training by providing ready-to-use data. *Prebuilt models* offer pretrained models for tasks such as natural language processing or image recognition, reducing time-to-value for common use cases. While *AI assistants* automate coding tasks, platforms like Google Vertex AI, Amazon SageMaker, Microsoft Azure AI, IBM Watsonxx, that own the *cloud infrastructure*, abstract infrastructure management through integrated managed cloud services. Other platforms that are not bundled with any cloud vendor's existing ecosystem would need to use connectors and APIs to access the compute and storage services of the cloud vendors. These resources cater to different user needs with tools and data assets supporting customization, while prebuilt models and cloud infrastructure prioritize convenience. All cases provide at least a single type of integrated resource with some providing multiple resources that support both the core value propositions and complementary services.

### 7.1.8 Data Source

The *data source* dimension distinguishes how platforms manage input data, a foundational element in machine learning workflows. Platforms may rely exclusively on *customer-provided data* or support a hybrid model integrating *both customer and third-party data* sources. *Customer data* consists of user-submitted datasets, which require strict preprocessing and governance to ensure quality and compliance. In contrast, platforms that incorporate *third-party data*, such as datasets from public data repository (e.g., Kaggle) or data marketplaces, allow users to supplement proprietary customer data with external sources. For instance, Snowflake facilitates hybrid approaches by combining internal transaction records with third-party demographic data, enhancing analytical depth for use cases like fraud detection, where real-time customer data may be enriched with external threat intelligence. Notably, platforms that integrate third-party data always support customer data, but the reverse is not observed.

### 7.1.9 Data Flow

*Data flow* distinguishes platforms based on how data is processed. *On-demand* data flow processes job in batch, where data is ingested and processed at scheduled intervals (e.g., monthly sales prediction). This approach suits scenarios with predictable workloads and no real-time requirements, such as historical trend analysis. *Continuous* data flow supports real-time inference of the machine learning models, allowing streaming data to be processed instantaneously, which are essential for use cases like fraud detection or IoT sensor monitoring. For example, Iguazio’s Intelligence Edge feature emphasized continuous customer data streams for real-time inference, whereas research-focused platforms (e.g., CodeOcean) relied on on-demand uploads.

### 7.1.10 Data Type

The *data type* dimension captures the primary forms of data a platform is designed to process, which fundamentally shapes its industry applicability and technical capabilities. Platforms typically specialize in *telemetric or time-series data*, *visual data*, or *multimodal data*, each serving distinct user needs from different industries. Platforms optimized for *telemetric or time-series data*, such as Superb AI, focus on structured quantitative inputs like IoT instrumentation data or sequential time-stamped data, enabling predictive analytics in industrial and scientific domains. Those specializing in *visual data* provide dedicated tools for image and video processing, supporting computer vision tasks such as object detection through convolutional neural networks. However, these specialized platforms are relatively rare, with only isolated cases observed in our analysis. Most MLaaS platforms adopt a *multimodal* approach, capable of processing diverse data types, including text, documents, natural language, sensor readings, and visual inputs. For instance, Databricks facilitates models with multimodal needs by concurrently analyzing combined data forms, enabling advanced applications like contextual image captioning or integrated analytics. Notably, while platforms with support for multimodal data inherently accepts textual data format, no platforms exclusively specialize in textual data processing alone. This distribution reflects user’s need for versatile data format support that balance specialization with flexibility, ensuring platforms can address complex, real-world problems while maintaining scalability and performance.

## 7.2 Value Delivery

Value delivery ensures MLaaS platforms’ offerings reach and engage intended users effectively. This section analyzes how platforms structure customer interactions, segmentation, and distribution channels.

### 7.2.1 Key Partners

*Key partners* describe the external collaborations that extend a platform’s capabilities or market reach. *Technology partners* are external entities, such as independent software vendors, innovation partners, that collaborate with MLaaS platforms to strengthen their technical capabilities, particularly in hardware or software integration. For example, Amazon SageMaker partners with NVIDIA to leverage GPU acceleration, enhancing the platform’s computational efficiency for demanding machine learning workloads. *System integrators*, such as Deloitte and Accenture, assist with deployment and customization, bridging gaps between platform features and enterprise needs. *Data brokers*, including those from Snowflake’s data marketplace, supply curated datasets, while *research or academic institutions* drive innovation via research partnerships (e.g., CodeOcean’s collaborations with Allen Institute in open-source innovations).

### 7.2.2 Customer Relationships

The *customer relationships* dimension captures the primary mechanisms through which MLaaS platforms engage with and support their users, playing a crucial role in user adoption, retention, and satisfaction. This dimension reflects how platforms balance scalability and personalized assistance, often aligning with their target audience and business model. *Self-service portals* are a foundational support structure, providing users with documentation, tutorials, and troubleshooting guides to resolve issues independently. Platforms like Dataloop exemplify this approach, offering extensive knowledge bases that enable data scientists to quickly find solutions without direct intervention. *Communities*, such as Amazon SageMaker’s user forums, create peer-driven ecosystems where users share knowledge, best practices, and workarounds. Dedicated *customer support*, including SLAs, ticketing systems, and account managers, is a hallmark of commercial platforms targeting enterprise users. For example, Dataiku assigns dedicated support teams to large clients, ensuring rapid resolution of technical issues and alignment with business objectives.

### 7.2.3 Industry Focus

The *industry focus* dimension categorizes MLaaS platforms based on the specialization and orientation of their target markets, distinguishing between narrow-focused solutions vertically-integrated for specific industries and *broad*-focused platforms horizontally-oriented for cross-domain applications. Vertically-oriented platforms, such as CodeOcean, concentrate on niche sectors like biotech, pharmaceutical or retail, offering pre-configured workflows and domain-specific model templates that accelerate deployment for specialized use cases.

### 7.2.4 Sales Channel

The *sales channel* dimension describes the primary methods through which the platform delivers its machine learning services to customers. It focuses on the main avenues, such as direct *platform interfaces*, *marketplaces* for model, data or services, or third party *partners*, by which users discover, access, and transact for the platform’s offerings. This dimension is critical because it determines the platform’s market accessibility, user onboarding efficiency, and scalability. The first characteristic, *platform interface and partners*, describes how the platform relies on its own user interface (web portal, APIs and SDKs) and partnerships with third-party organizations (e.g., cloud providers, system integrators, resellers, professional service providers) to sell and distribute its services. For example, Dataiku operates primarily through its web-based interface, enabling users to build machine learning workflows, while also collaborating with system integrators like Accenture to deploy solutions for large corporations. The second characteristic, *platform interface, partners, and marketplace*, extends this strategy by incorporating a digital marketplace where third-party vendors can offer complementary tools, models, or datasets. This creates a multisided platform ecosystem, enhancing value for users by providing a one-stop shop for machine learning-related services. Amazon SageMaker exemplifies this approach where users interact with SageMaker's interface to train models, leverage AWS's partner network for consulting, and access the AWS Marketplace to deploy prebuilt machine learning services from independent vendors. While only few platforms have their own marketplace, all platforms in our case database rely only platform interface and a directory of authorized partners to enable users' discovery and procurement of the required services.

## 7.3 Value Capture

*Value capture* mechanisms define how MLaaS platforms monetize their services and sustain profitability. This section examines the financial structures of MLaaS platforms through the lens of pricing models, and revenue streams. While cost structures provide valuable insights into operational efficiency, their analysis often requires access to proprietary or confidential financial data, which is not available to this study. Given these constraints, our study focuses exclusively on pricing models and revenue streams to assess value capture mechanisms.

### 7.3.1 Pricing Model

The *pricing model* dimension captures the different approaches MLaaS platforms use to monetize their services, reflecting the diverse and often unpredictable computational demands throughout the machine learning lifecycle. *Consumption-based* pricing, as seen with Amazon SageMaker, charges users based on actual usage but breaks down costs according to different stages of the lifecycle. For example, users pay for GPU hours during model training, fees per API call when running inference, and charges for data storage and transfer within pipelines. This detailed breakdown ensures that costs align closely with the varying resource intensity required at each phase, making pricing more transparent and fair. *Tiered* pricing models, such as those offered by Dataiku, group features into hierarchical plans where higher tiers grant access to more advanced lifecycle resources, like support services with higher SLA, increased capacity for training jobs or simultaneous inference processes. *Seat-based* pricing, used by only Domino Data Lab out of all the cases, links costs to the number of users, often with restrictions on what each user can do within the machine learning lifecycle, such as limiting some seats to training-only roles. This model supports team-based scaling while managing permissions according to responsibilities. *Freemium* pricing models divide offerings either by time or functionality to attract users. For instance, Google Vertex AI provides a 90-day free trial, while Amazon SageMaker offers a free tier that excludes costly activities like GPU-based training. These models encourage initial adoption by lowering entry barriers but restrict access to resource-intensive lifecycle stages to motivate users to upgrade. Across all cases, we noticed that many MLaaS platforms use hybrid models that blend fixed and variable components by combining both tiered and consumption-based models. This approach is likely due to the complexity in pricing the platform services as computational demands vary across the machine learning lifecycle.

### 7.3.2 Revenue Streams

*Revenue streams* in MLaaS platforms are the diverse mechanisms platforms use to generate income from core and complementary services. *Direct sales* refer to the revenue obtained through primary platform offerings, mainly through two ways: subscription models and transaction fees. Subscription-based revenue, as demonstrated by IBM Watsonx's annual licensing structure, provides predictable recurring income. Transaction-based revenue, exemplified by Amazon SageMaker's per-inference and pay-per-use charges for real-time inference APIs. *Indirect sales* come from third-party collaborations and platform ecosystem participation. These include marketplace advertising revenues, where third-party vendors pay for enhanced visibility in platform-hosted marketplace, as observed in Snowflake's data asset listings on their own marketplace. Partnership revenue represents another indirect stream, occurring when platforms share income with technology partners who co-develop specialized services for the platform's user base. *Ancillary services* generate supplementary revenue through value-added offerings that complement core platform functionalities. Training and certification programs, such as Databricks' accredited paid courses, create monetization opportunities by offering formal credentialing pathways.

# 8. Analysing MLaaS Platform Business Model Taxonomy

This chapter presents the findings from the cross-case analysis of 24 MLaaS platforms using the finalized taxonomy. By examining recurring patterns across the characteristics of all these cases, we identify four distinct business model archetypes: *cloud orchestrators*, *data orchestrators*, *aggregators*, and *niche specialists*. The following sections illustrate the taxonomy patterns focusing on the dominant characteristics of each archetype, findings of the taxonomy analysis and the details of each of the four business model archetypes.

## 8.1 Taxonomy Patterns across MLaaS Platforms

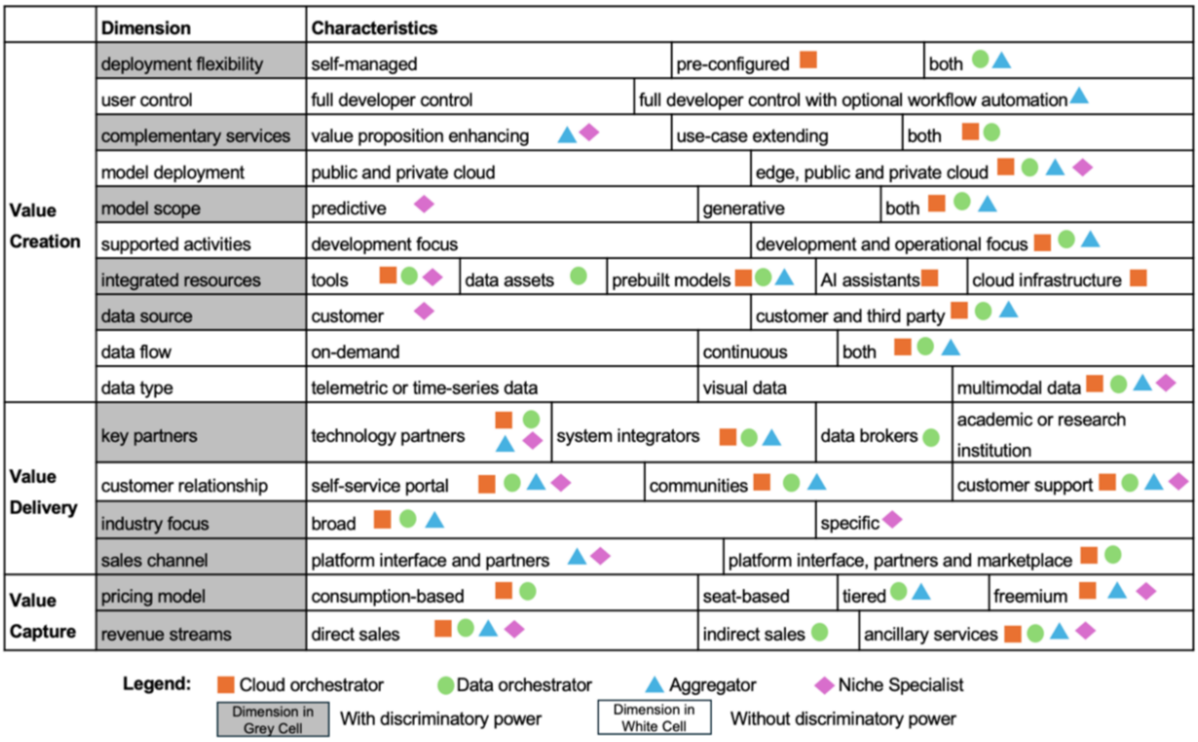
First, *deployment flexibility* dimension reveals a spectrum rather than a simple choice between self-managed and pre-configured approaches. We found that some platforms, like Barbara Edge AI and PrimeHub, prioritize user control by enabling enterprises to deploy models within their own environments. This capability can be essential for sectors with strict regulatory or security requirements, such as healthcare or defence, and demonstrates a commitment to developer autonomy and granular customization. In contrast, fully managed platforms such as Amazon SageMaker and Google Vertex AI abstract away infrastructure concerns, appealing to organizations seeking scalable, rapid deployment in multi-tenant cloud environments. We also found hybrid models, as in Snowflake and Databricks, accommodate both approaches, supporting deployment across public or private cloud as well as on-premises settings. These intermediate cases reflect the industry's response to varied customer needs, emphasizing deployment as a strategic lever for aligning technical flexibility with organizational priorities.

A second prominent pattern involves *industry focus* dimension and the resulting differentiation level of specialization across MLaaS platforms. Broadly focused players like Google Vertex AI and Amazon SageMaker appear to target wide applicability across sectors, favouring modularity and scalability. Their standardized features encourage user adoption in diverse fields, from retail to healthcare. In contrast, we noticed platforms such as Barbara Edge AI and Superb AI with focus on specific industries, building domain-specific features, like advanced edge deployments or proprietary annotation tools, into their value proposition. This specialization allows them to address complex, niche problems that general purpose MLaaS platforms may not support well. However, there is also a hybrid group that occupies the middle ground by maintaining broad usability while developing bespoke solutions for high-value verticals, as seen with DataRobot's sector-targeted offerings. Here, we observed that industry focus became a primary influence in the varying approaches in resource allocation, platform features, and ultimately, in the market segments that platforms are in.

*Integrated resources* dimension represent the focus of the third key pattern. Platforms across all 24 cases vary in how comprehensively they bundle proprietary tools, third-party integrations, and workflow automation. Vertically integrated platforms, often offered by major cloud providers, tend to tightly couple machine learning services with proprietary infrastructure and complementary services, minimizing the need for external tools and streamlining the user experience. Some platforms offer curated marketplaces and plugin ecosystems that empower users to assemble custom workflows, reflecting their focus on interoperability and flexibility. Some platforms, like Snowflake, blend both strategies, controlling core architectural components while selectively opening up to external machine learning tools and datasets via integrated marketplaces. The approach to resource integration influences not only the technical extensibility of platforms but also factors such as data governance and interoperability, each of which is critical in enterprise AI adoption.

Finally, value capture strategies in MLaaS platforms reflect significant differences tied to their business models and customer segments. Consumption-based pricing is prevalent among platforms, aligning with the scale and unpredictability of platform usage. This pricing model allows for elastic adoption but result in variable costs that allow SMEs to scale their commitment to the platform following the increasing returns of machine learning. Subscription and tiered pricing models, common among vertically oriented platforms with specialized focus, prioritize predictable billing and often bundle premium support or features. These niche MLaaS platforms tend to opt for custom quotations, reflecting high-touch, domain-specific consulting and service models. Entry-level freemium tiers can be seen in some MLaaS platforms where these tiers can be effective in encouraging initial adoption. Upselling to enterprise features occurs as the user's organizations grow. In some cases, these platforms generate indirect revenue through marketplace commissions, especially where data sharing or referral of professional services offered by partners are key components to the marketplace. Notably, a pattern emerges between industry focus and pricing transparency: horizontally focused platforms typically publish standard pricing, while vertical specialists use bespoke quotation.

Despite significant differences across these patterns (see Figure 19), there are notable areas of convergence. Most platforms, regardless of archetype, now offer support for both predictive and generative AI models, responding to market demand for versatile machine learning capabilities. Hybrid deployment, supporting both cloud and edge computing, is becoming the standard rather than a unique differentiator. The greatest distinction among platforms lies not in the presence of specific technical features but in the way they combine these features, emphasize different business model dimensions (see Figure 19), and align offerings with the unique requirements of their chosen markets. The result of frequency analysis that shows the percentage of cases for each dominant characteristic in each archetype is available in Appendix E4.



| | Dimension | Characteristics | | | | |
|---|---|---|---|---|---|---|
| **Value Creation** | deployment flexibility | self-managed | pre-configured ■ | both ● ▲ | | |
| | user control | full developer control | full developer control with optional workflow automation ▲ | | | |
| | complementary services | value proposition enhancing ▲ ◆ | use-case extending | both ■ ● | | |
| | model deployment | public and private cloud | edge, public and private cloud ■ ● ▲ ◆ | | | |
| | model scope | predictive ◆ | generative | both ■ ● ▲ | | |
| | supported activities | development focus | development and operational focus ■ ● ▲ | | | |
| | integrated resources | tools ■ ● ◆ | data assets ● | prebuilt models ■ ● ▲ | AI assistants ■ | cloud infrastructure ■ |
| | data source | customer ◆ | customer and third party ■ ● ▲ | | | |
| | data flow | on-demand | continuous | both ■ ● ▲ | | |
| | data type | telemetric or time-series data | visual data | multimodal data ■ ● ▲ ◆ | | |
| **Value Delivery** | key partners | technology partners ■ ● ▲ ◆ | system integrators ■ ● ▲ | data brokers ● | academic or research institution | |
| | customer relationship | self-service portal ■ ● ▲ ◆ | communities ■ ● ▲ | customer support ■ ● ▲ ◆ | | |
| | industry focus | broad ■ ● ▲ | specific ◆ | | | |
| | sales channel | platform interface and partners ▲ ◆ | platform interface, partners and marketplace ■ ● | | | |
| **Value Capture** | pricing model | consumption-based ■ ● | seat-based | tiered ● ▲ | freemium ■ ▲ ◆ | |
| | revenue streams | direct sales ■ ● ▲ ◆ | indirect sales ● | ancillary services ■ ● ▲ ◆ | | |

**Legend:** ■ Cloud orchestrator ● Data orchestrator ▲ Aggregator ◆ Niche Specialist
Dimension in Grey Cell With discriminatory power Dimension in White Cell Without discriminatory power

*Figure 19: Taxonomy Patterns and Business Model Archetypes*

*Note. Only characteristics present in at least 70% of the cases of each archetype were visualized.*

## 8.2 Business Model Archetypes

By iteratively synthesizing the patterns we recognized during taxonomy analysis phase, we observed four clusters of cases across the case database where each cluster was presented with distinct characteristics. We thus derived business model archetypes from these four clusters. This section walks through the four business model archetypes (*cloud orchestrator*, *data orchestrator*, *aggregator* and *niche specialist*) and explains the grouping of the four boundary cases into their corresponding business model archetypes (see Table 18).

*Table 18: Cases and Boundary Cases of Business Model Archetypes*

| Archetypes | Cloud orchestrator (n=5) | Data orchestrator (n=4) | Aggregator (n=8) | Niche specialist (n=7) |
|---|---|---|---|---|
| Cases | Amazon SageMaker, Google Vertex AI, IBM Watsonxx, Microsoft Azure AI | Databricks Mosaic AI, Dataloop, Domino Data Lab, Snowflake | BigML, Dataiku, DataRobot, H2O.ai, Iguazio, Jfrog ML, Modular | Barbara, HyperSense, Superb AI, ZAMS |
| Boundary Cases | None | None | Cloudera AI | PrimeHub, CodeOcean, VESSL AI |

### 8.2.1 Cloud Orchestrator

*Cloud orchestrators* create value by democratizing access to advanced machine learning capabilities through fully managed infrastructure from their hyperscale data centres. As Teece (2010) notes in his analysis of digital business models, these MLaaS platforms leverage economies of scale to dramatically lower the technical barriers to AI adoption. The value proposition focuses on abstracting hardware provisioning and optimization, allowing data scientists to focus on model development rather than infrastructure management. Amazon SageMaker embodies this approach through its one-click model training and deployment features, automatically provisioning and optimizing GPU clusters.

Value delivery in *cloud orchestrators* is achieved through deep integration with services that are built specifically for their own cloud infrastructure. Google Vertex AI demonstrates this through its seamless compatibility with BigQuery, enabling SQL-based data preprocessing that significantly reduces time-to-insight. These MLaaS platforms enhance their value delivery through network effects, hosting marketplaces for third-party specialized tools. Microsoft Azure AI extends this further by incorporating responsible AI dashboards for bias detection, addressing growing concerns about ethical AI implementation.

Value capture follows a consumption-based pricing model that aligns with the "*razor-and-blades*" strategy identified by Eisenmann et al. (2006), where platform sells a core product, like compute, data storage and access to a wide selection of machine learning models for the case of MLaaS platforms, cheaply to drive ongoing profit from the required higher-margin replacements. In this context of MLaaS platforms, users pay for compute resources and API calls, creating predictable recurring revenue streams through subscriptions or transactions based on each resource consumption. The model encourages platform adoption through low initial costs while generating long-term revenue through scaled usage with tiered pricing model. Ancillary services, like premium support, training, certification and custom model development through project work, provide additional revenue opportunities, completing the value capture mechanism.

### 8.2.2 Data Orchestrator

The *data orchestrator* platform archetype creates value by solving the fundamental challenge of data fragmentation in machine learning implementation within large enterprises. These platforms establish streamlined data ecosystems that bridge the gap between fragmented data sources and machine learning workflows, addressing the coordination costs in technological platforms (Gawer & Cusumano, 2014). Value creation stems from their ability to integrate governance, lineage tracking, and hybrid deployment capabilities into a well-integrated system for users from enterprises with big data. Snowflake exemplifies this approach by enabling direct machine learning model training on data warehouse contents, effectively eliminating data silos between analytics and machine learning teams. The value proposition focuses on reducing the operational complexity of large-scale machine learning implementations while ensuring regulatory compliance through its ancillary services.

Value delivery in *data orchestrator* platforms occurs through interoperable architectures supporting multi-tenant public cloud and on-premise private cloud deployments. Without native integration to cloud infrastructure like *cloud orchestrators*, *data orchestrators* custom-build tools, such as connectors and APIs, that allow users to deploy the machine learning models to access compute and storage resources in the cloud infrastructure. Some *data orchestrators*, like Snowflake, also secured dedicated compute and storage resources from the cloud providers, which they could then offer to their users as integrated cloud infrastructure resources. On top of a marketplace curated by the platform, these platforms expand their value delivery through the development of an ecosystem. They provide APIs and SDKs that allow third-party developers to create complementary tools, while moderating such interactions within their platform ecosystem (Gawer, 2014).

The value capture mechanism for *data orchestrators* combines subscriptions with tiered-based pricing on their platform and consumption-based pricing on the usage on platform services requiring compute resources. Ancillary services such as consulting and compliance audits are also provided. This approach allows platforms to monetize both core infrastructure and specialized supplementary features. Their value proposition lies in reducing the complexity of data governance, making them indispensable for regulated industries like healthcare and finance.

### 8.2.3 Aggregator

*Aggregator* platforms create value by reducing the fragmentation that characterizes many enterprise machine learning implementations. These platforms curate modular tools and open-source integrations into their web-based and programmatic environments that support the complete machine learning lifecycle. *Aggregator* platform's value proposition focuses on providing flexibility and variety of complementary services that accommodates diverse technical requirements from users that require different degree of control on the machine learning. Value delivery emphasizes customization and interoperability. Unlike cloud orchestrators with tightly integrated infrastructure or data orchestrators with rich data ecosystems, aggregator platforms deliver value by curating partner networks and integrating diverse resources. This approach is evident in BigML's public gallery of reusable model templates and Dataiku's plugin library, which offers connectors for various data platforms. The value capture mechanism of *aggregator* platform often combines freemium models with tiered subscriptions and ancillary services. The core offerings attract users with free basic features while enterprise tiers unlock advanced governance and scalability features. This approach creates multiple entry points to the platform while ensuring revenue growth through upselling. Consulting services for custom model development provide high-margin ancillary revenue, completing the value capture model.

Cloudera AI, as a boundary case, defies straightforward classification as an *aggregator* due to its orchestrator-like control over data infrastructure. While it curates third-party tools and integrates with partner ecosystems like other *aggregator* platforms, it also enforces strict governance via its unified data lakehouse, a feature typical of data orchestrators like Snowflake. However, Cloudera AI is not equipped with data marketplace and network of data brokers like the *data orchestrator*, thus making it classified as *aggregator*.

### 8.2.4 Niche Specialist

*Niche specialists* create value through vertical specialization. These platforms offer preconfigured solutions that outperform general purpose alternatives in specific applications, from edge computing to industry-specific machine learning implementations. Barbara exemplifies this approach with industrial-focused machine learning models for predictive maintenance in utilities, complete with preconfigured support for IoT sensor data. Value delivery focuses on self-managed deployment. Superb AI demonstrates this through its synthetic data tools that streamline computer vision workflows, reducing labelling costs significantly for autonomous vehicle developers. Unlike horizontal platforms, *niche specialists* prioritize depth over breadth in their solutions, as seen in CodeOcean's specialized support for genomic analysis in biotech and pharmaceutical. *Niche specialists*' value capture mechanism relies on sales models where result-based pricing is employed by some of the platforms. Rather than consumption-based metrics, some platforms may employ fixed-fee structures like Barbara's €1,750/month industrial packages or contracts linked to return of investment rate. This approach aligns with some of their focus on delivering measurable business impact rather than charging for compute resources for model training or inference. Many of the *niche specialist* platforms, unfortunately, do not publicly share information regarding pricing and service specification their platform websites, indicating possible custom quotation needed for each customer.

There are three boundary cases for *niche specialist* platform. First, CodeOcean combines niche research focus in biotech and pharmaceutical industry with marketplace-like features, allowing researchers to share reproducible workflows, a trait more common in aggregators. Its pre-configured deployment and consumption-based pricing resemble *cloud orchestrators*, yet its narrow industry focus prevents it from fitting neatly into that category. PrimeHub targets academic and enterprise MLOps but stands out by offering both self-service portals and white-label solutions, both characteristics rare in most *niche specialists*. Lastly, VESSL AI bridges *niche specialist* as MLaaS platform for edge AI and broad-focused cloud platforms. While its edge deployment aligns with *niche specialists* like Barbara, its support for public cloud integrations and freemium pricing mirrors *cloud* and *data orchestrators* with wide range of complementary services. This hybrid approach allows it to serve both specialized edge deployments and broader cloud-based workflows, defying strict archetype boundaries. Despite their boundary characteristics, we chose to classify these three cases as *niche specialist*, not just due to its industry focus but for its special platform use cases and characteristics. This indicates that each *niche specialist* may configure and adapt its business model on an ongoing basis to fulfil the needs of its target user group.

# 9. Discussion

The discussion chapter synthesizes findings from this study, addressing main research question and sub-questions while explaining the findings from prior chapters. The structure follows a logical progression from specific findings to broader implications, using the cases analysed and the respective results to support our arguments.

## 9.1 Answering Main Research Question and Sub-questions

The main research question (**RQ**) asked, “How can the business models of MLaaS platforms be systematically classified into a taxonomy, and what common archetypes emerge from the combination of their key dimensions and characteristics?” This main research question was addressed by demonstrating that the business models of MLaaS platforms can be effectively conceptualized using a hierarchical conceptual framework. To sufficiently address the main research question, we further answered the sub-questions in the following order.

The first sub-question (**SQ1**) inquired, “*How can the business models of MLaaS platforms be conceptualized using current literature?*” To address this, the study adopted a three-level hierarchical approach based on Osterwalder et al.’s (2005) Business Model Concept Hierarchy, first constructing a meta-model to define the essential dimensions of MLaaS platform business models. The meta-model, integrated Teece’s (2010) value creation, delivery, and capture mechanisms with the components of Osterwalder and Pigneur’s (2010) Business Model Canvas. We discovered that Teece's value creation, delivery and capture mechanisms (2010) and Osterwalder's Business Model Canvas components (2010) aligned clearly when being applied to conceptualize MLaaS platforms. To formulate the definition required to kick off the taxonomy development process, the study employed two literature reviews involving scoping review and rapid review. While the scoping review led us to formulate the definition of our object of analysis as dictated by Nickerson et al.’s (2013) methodology, the rapid review systematically examined business model taxonomy literature to extract key dimensions and characteristics to be considered in the construction of preliminary business model taxonomy. However, there is a lack of literature directly related to MLaaS platforms. Because of this gap, the review first turned to literature on AI service platforms, an adjacent domain indirectly and closest related to MLaaS platforms. Much of this literature developed archetypes and taxonomies but did not specifically analyse business models. To address this, we extracted taxonomy dimensions related to business models from these sources, selecting those applicable to MLaaS platforms. Subsequently, we expanded the review to include literature on AI-related business models, focusing specifically on AI startups, AI solutions, and AIaaS. This step aimed to identify additional relevant business model dimensions and characteristics from contexts closely related to MLaaS platform. Finally, we broadened the scope to encompass the wider domains of platform business models and data-oriented business models, where a substantial corpus of established business model taxonomies exists. This progression allowed us to incorporate a more comprehensive set of dimensions drawn from mature literatures, which were essential for constructing a robust and well-grounded MLaaS business model taxonomy. Through the conceptual-to-empirical iterations of taxonomy development process, the key dimensions and characteristics that were extracted during rapid review were then synthesized to construct a preliminary business model taxonomy for MLaaS platforms. Despite the lack of directly relevant literature, the preliminary taxonomy demonstrated that adapting the findings from existing literature from the adjacent domains was sufficient during the conceptual-to-empirical phase.

The second sub-question (**SQ2**) examined, “What are the key characteristics of MLaaS platform business models that are systematically captured and structured within a business model taxonomy?” This study developed a business model taxonomy that captured the key characteristics of MLaaS platform business model through its dimensions and characteristics. The taxonomy systematically illustrated how MLaaS platforms created, delivered, and captured value through the identification of their business model characteristics, with the meta-model using Business Model Canvas components as definitional guidance. For value creation, deployment flexibility, user control, model scope, model deployment, complementary services, supported activities, integrated resources, data source, data flow, and data type were identified as business model taxonomy dimensions. These dimensions revealed how platforms differed in the varying level of automation provided by each platform, types of resources integrated with the platform, and the degree of flexibility and control in the value proposition offered to the users. For value delivery, the taxonomy included key partners, customer relationships, industry focus, and sales channels dimensions, highlighting the different ways platforms engaged with users and different partners within their platform ecosystem based on their varying focus on industry verticals. Unlike the customer segment dimension in the meta-model, the taxonomy incorporates the industry focus dimension from Geske et al.’s (2018) taxonomy to move beyond a simple descriptive characteristic and instead capture the different verticalization characteristics in the customer segment focus of MLaaS platforms. This approach reflects how niche specialist MLaaS platforms of concentrate on serving specific industry verticals while the others (cloud orchestrator, data orchestrators, and aggregator MLaaS platforms) treat customer segments as broad, generic groups. Notably, these MLaaS platforms primarily target business-to-business customers, including both SMEs and large organizations, tailoring their offerings to meet the specialized needs across different sectors that fall in their chosen industry focus. For value capture, the taxonomy captured pricing models and revenue streams dimensions, showing the varying monetization models used by different MLaaS platforms. The taxonomy did not sufficiently characterize the cost structure of MLaaS platforms because cost structure characteristics (tangible vs intangible) that were initially identified in conceptual-to-empirical phase were not consistently available in public platform documentations and third party publications.

The third sub-question (**SQ3**) assessed, "*To what extent is the business model taxonomy of MLaaS platforms structurally valid and practically usable?*" To answer this question, this study applied predefined evaluation criteria to assess the taxonomy's structural validity and practical usability. Following the evaluation approach using Szopinski et al.'s (2020) guidelines, the taxonomy demonstrated the structural validity, which indicated that it was *comprehensive* in covering all relevant aspects, *mutually exclusive* without having any overlap between dimensions or characteristics, and showed *clarity* through the use of precise terminology. Practical usability was also assessed using the criteria set out by Miles and Huberman (1994), focusing on the taxonomy's *robustness*, *theoretical grounding*, and *utility* for both researchers and practitioners. After testing the taxonomy by applying the taxonomy on four MLaaS platforms, (BigML, Microsoft Azure AI, DataRobot, and VeSSL AI), the taxonomy evaluation confirmed that the taxonomy could accurately classify these platforms without redundancy or gaps, demonstrating its comprehensive coverage and functional utility. However, due to time constraints, the taxonomy and archetypes were not directly validated with practitioners, such as MLaaS platform operators or consultants familiar with the domain, to receive feedback on real-world application of the taxonomy. This indicates a possible gap between this study's evaluation that uses theoretical criteria and real-world applicability.

The fourth sub-question (**SQ4**) explored, "*What business model archetypes can be derived from recurring patterns in the MLaaS platform business model taxonomy?*" To answer this sub-question, we identified four distinct archetypes following the cross-case analysis of all 24 cases. The *cloud orchestrator* archetype represented platforms with integrated cloud infrastructure that achieve horizontal scale and broad applicability, such as Amazon SageMaker, Google Vertex AI, and Microsoft Azure AI. *Data orchestrators* focused on a data-centric value proposition by integrating and managing diverse data resources, such as data assets from data brokers, through marketplaces integrated with their platforms. This includes MLaaS platforms such as Snowflake, Databricks Mosaic AI, Dataloop, and Domino Data Lab. *Aggregator* platforms, such as DataRobot, BigML, Dataiku, H2O.ai, Iguazio, Jfrog ML, Modular and Cloudera AI, stood out for having a partner ecosystem and their use of tiered pricing model. *Niche specialist* platforms, like HyperSense, Superb AI, ZAMS, PrimeHub, CodeOcean and Barbara, demonstrated deep specialization by targeting specific industries or use cases with tailored offerings. Boundary cases were noted for both *aggregator* and *niche specialist* platforms, indicating that the clustering process and criteria need to be further refined and reorientated so that there is mutual exclusivity between all the cases for each archetype.

## 9.2 Comparing Research Outcomes with Findings from Prior Work

In this research, we observe that a key distinction between MLaaS platforms and their common archetypes lies in their degree of scalability and specialization. Geske et al. (2021) identified three motives for AI service platforms based on varying levels of development effort and design freedom, but did not explicitly link these to scalability or industry specialization. Unlike Geske et al., who focused on classifying AI service platforms based on functional motives (e.g., development effort vs. design freedom), this research observes that the real-world business models of MLaaS platforms, especially their value creation, depend not only on AI functionality but also on cloud infrastructure and machine learning workflow orchestration.

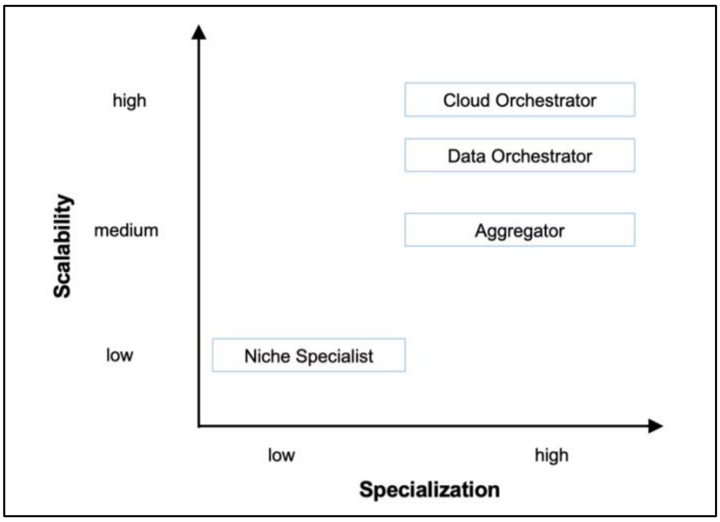


*Figure 20: Business model archetypes across different levels of scalability and specialization*

By adapting Geske et al.’s (2021) motives based on “*development effort*” and “*design freedom*,” the taxonomy captures how MLaaS platforms, especially their business model archetypes, demonstrate different degrees of scalability and choice of specialization through their different characteristics in the “*deployment flexibility*” and “*user control*” dimensions (see Figure 20). While Geske et al. (2021) classify platforms by "*AI service model*" (e.g., deployment vs. development services), this study reveals that *deployment flexibility*, whether a platform supports preconfigured cloud or self-managed edge setups specific for industrial IoT sector, is equally critical. *Cloud orchestrators* like Microsoft Azure AI exemplify this by tightly integrating with its proprietarily-owned cloud architecture, a technical-oriented characteristic overlooked in prior taxonomies. This shows that the value creation of MLaaS platforms is shaped by their varying levels of dependency on technical architectures. Therefore, these MLaaS platforms should not be treated as homogenous entities. Prior taxonomies, including Sundberg et al.’s (2022) typology, tended to overlook the substantial impact of cloud infrastructure dependencies. This oversight fails to account for the significant variation in how technical architectures shape the business models and value creation mechanisms of MLaaS platforms.

This study further reframes these two dimensions within the context of MLaaS platforms into scalability constrained by the level of control over technical infrastructure (e.g., providers vs. users of cloud services) and specialization following the level of domain adaptation of the value proposition (e.g., industry-specific workflows). This reveals a trade-off: platforms optimized for scalability, such as hyperscale *cloud orchestrators*, typically offer lower *specialization*, while those emphasizing specialization, such as *niche specialists*, sacrifice broad *scalability*. For example, *cloud orchestrators* maximize scalability through fully managed, multi-tenant cloud infrastructure, aligning with Teece’s (2010) observation that digital platforms leverage economies of scale to reduce barriers to adoption. These platforms abstract infrastructure complexity, enabling rapid deployment at scale but limiting customization for vertical use cases. In contrast, *niche specialists* prioritize deep domain integration, such as preconfigured edge deployment for predictive maintenance, but lack the horizontal scalability of *cloud orchestrator* and *data orchestrator* MLaaS platforms. This dichotomy extends Sundberg et al.’s (2022) typology, which classified AI platforms by functional capabilities (e.g., conversational vs. analytics) and delve deeper on how *deployment flexibility* (preconfigured vs. self-managed) intersects with *industry focus*.

When evaluated together, both “*model scope*” (predictive vs. generative) and “*model deployment*” (hybrid cloud and edge) in the business model taxonomy reveal that machine learning workflow orchestration is a key characteristic of value proposition and should be considered when analysing the business model of MLaaS platforms. Although cloud-based service delivery is core to AIaaS, and, by extension, MLaaS platforms that inherit some AIaaS characteristics (Lins et al., 2021), many MLaaS platforms also deliver services beyond the cloud by supporting deployment to edge devices or hybrid cloud environments. These options are now pivotal for latency-sensitive applications like industrial IoT and autonomous driving. This is demonstrated by the consistent observation in this research where value creation manifests differently across business model archetypes, from *niche specialist* Barbara’s edge-focused predictive maintenance models to *data orchestrator* Databricks’ hybrid cloud-data lakehouse integrations for both predictive and generative machine learning models.

As indicated by Naous et al. (2017), deployment options provided by data-driven platform business models are linked to value creation. This further reflects the need to consider “*model scope”* and “*model deployment*” dimension when conceptualizing the business model of MLaaS platforms. Therefore, we propose that the business model taxonomy from this research reflect the realities faced by MLaaS platforms, where value creation depends on both platform functionality and how users interact with the platform’s underlying cloud-dependent technical infrastructure. Furthermore, the value creation of MLaaS platforms depends on different levels of scalability and specialization. Platforms offering a high level of scalability in their value proposition tend to focus less on specialization, while platforms focusing on specialization through domain-specific consultation or project implementation services tend not to prioritize scalability.

While value creation in MLaaS platforms is distinct, the mechanisms for value delivery and capture closely mirror those of traditional platform business models. Business model dimensions such as “*sales channel”* and “*pricing model”* in the taxonomy presented by this study resonate those dimensions contributing to value delivery and capture that were found in established taxonomies of digital platform business model (Staub et al., 2021; Täuscher & Laudien, 2017). This indicates that MLaaS platforms adopt proven platform strategies for monetization and customer engagement despite the fact that their value creation mechanisms depend heavily on cloud infrastructure and differentiation through varying support for the machine learning lifecycle. The roles of key partners and customer relationships also align with multisided platform strategies, where ecosystem collaboration is essential for scalability (Rochet & Tirole, 2003). One notable divergence is generative AI support, such as large language model deployment and coding assistants, which introduces new value capture mechanisms such as pay-per-token pricing, a form of consumption-based pricing that differs from the typical subscription models of traditional digital platforms.

Many MLaaS platforms, such as those classified as *cloud orchestrators*, *data orchestrators*, and *aggregators*, often serve both industry-specific and generic use cases, reflecting the need for both specialized machine learning services and general-purpose tools. Compared to traditional digital platforms and AI service platforms, the complementary services of MLaaS platforms are more oriented towards technically-skilled developers and data scientists, focusing on features such as model explainability and AutoML rather than just the marketplace add-ons typical of conventional platforms. A notable finding is the prominence of AutoML and AI assistants across "*user control*", "*complementary services*", and "*integrated resources*" dimensions of MLaaS platform business models. The presence of AutoML features across the MLaaS platforms analysed illustrates how MLaaS platforms prioritize balancing automation with flexibility and accessibility, making advanced machine learning accessible to non-experts while supporting more technically informed users. So, the inclusion of AutoML as workflow automation in MLaaS platforms indicates its current and anticipated growing importance as a key value differentiator.

## 9.3 The Influence of Platform Envelopment on Value Proposition

Our findings from taxonomy analysis suggests that *cloud orchestrators* such as Amazon SageMaker and Google Vertex AI seem to leverage their dominant market position through the bundling of MLaaS functionalities into their existing platforms. We suggest that this phenomenon aligned with "*platform envelopment*" that occurs when dominant platforms expand into adjacent markets by bundling new functionalities with existing services, leveraging their user base and infrastructure to absorb competitors (Eisenmann et al., 2011). By integrating MLaaS capabilities, these platforms simultaneously reduce adoption friction for existing userbase and create dependencies that discourage switching to a different MLaaS platform—a form of lock-in that generates economies of scale (Teece, 2010). In this case, envelopment in MLaaS platforms specifically exploits infrastructure ownership as a competitive advantage, forcing competitors to differentiate through specialization or ecosystem partnerships.

*Data orchestrators* like Snowflake and Databricks, while lacking direct control and ownership over compute infrastructure, demonstrate a comparable form of envelopment based on their integration of data governance and machine learning pipelines within a single environment. This supports the notion presented by Jacobides et al. (2018) that platforms can achieve ecosystem lock-in by orchestrating relationships and coordinating resources beyond mere ownership of hardware. Our analysis suggests that these platforms leverage their deep integration of MLaaS capability with their existing data and analytics functions, creating a value proposition that relies on enhancing interoperability and process efficiency for enterprise users with large amount of complex datasets. This complements Hagiu and Wright's (2015a) insights on multisided platforms, where the platform creates value by facilitating complex interactions among distinct participant groups—in this case, data brokers, data scientists, developers and these type of enterprise users from both SMEs and large organizations.

*Aggregators* and *niche specialists* present a contrasting picture. *Aggregators* emphasize openness and modularity, connecting users with third-party tools without direct control over core cloud or data infrastructure. This approach creates a third-party ecosystem allowing these platforms to create value primarily through partnerships rather than vertical integration. However, the limited ability of *aggregators* to leverage complementary assets or infrastructure ownership restricts their capacity to address larger market size and broader segments. *Niche specialists* appear to deliberately sidestep direct competition with MLaaS platforms from hyperscale cloud providers or data platforms by focusing on specialized industry sectors (such as pharmaceutical, biotech, manufacturing) requiring domain-specific expertise or unique deployment models, such as edge computing for industrial IoT. This observation points to the potential resilience of niche specialist MLaaS platforms despite Big Tech's overarching dominance. Specialization can enable *niche specialist* MLaaS platforms to coexist with other archetypes by serving distinct market needs inadequately addressed by generalist platforms. As revealed by the taxonomy, these *niche specialists* often have proprietary data integrations, and verticalized workflow optimizations, which could be the tactical defence mechanisms against the envelopment efforts from *cloud* and *data orchestrators*. So, the sustainability of this *niche specialist's* position may depend on the degree to which *cloud and data orchestrator* platforms find such segments strategically worthwhile to capture.

One question arises whether these *niche specialist* platforms represent a temporary phenomenon that will eventually be subsumed by dominant players, or if they possess enduring competitive advantages that allow survival or even growth alongside Big Tech. On one hand, Eisenmann et al. (2011) suggests that dominant platforms can strategically bundle adjacent services to pre-empt or absorb competitors, raising the possibility that hyperscale *cloud orchestrator* platforms may eventually extend to include more customized offerings. This propensity is reinforced by the enormous resources and innovation capabilities of Big Tech firms (van der Vlist et al., 2024). On the other hand, highly specialized or domain-specific capabilities such as human capital, proprietary data, unique workflow integrations, can form isolating barriers that are costly or complex for large platforms to replicate authentically (Barney, 1991). In environments where trust, regulatory compliance, latency constraints, or customization are critical, *niche specialists* may maintain its stable market position in the near future (Jacobides et al., 2018).

The findings from this study support and extend existing literature on the platformization and industrialization trend in AI industry by illustrating the phenomenon leading to MLaaS platform envelopment. Consistent with Luitse's (2024) analysis, the integration and bundling activities observed among *cloud* and *data orchestrators* go beyond mere technical capabilities. These efforts reflect a form of vertical integration wherein dominant platforms not only embed machine learning capabilities into their broader cloud or data infrastructures but also exert control over the entire AI value chain. Through this deeply integrated ecosystem approach, the Big Tech operators behind these MLaaS platforms govern access to AI functionalities and cultivate user dependencies, reinforcing their market positions. Similarly, van der Vlist et al. (2024) further articulates that Big Tech's dominance arises from their control of critical computational infrastructure and data resources, which generates structural dependencies across the AI industry. This infrastructural dominance facilitates industrial-scale AI deployment, with these firms creating unified platform environments that both consolidate market power and impose significant barriers to entry. The phenomenon highlighted by van der Vlist et al. (2024) resonates with the envelopment phenomenon detected across MLaaS platforms, where bundled services increase switching costs and lock users into the ecosystem.

Taken together, these perspectives enrich the interpretation of the taxonomy developed in this research by situating MLaaS platform business models within broader AI industry and ecosystem. The coexistence of *cloud orchestrators*, *data orchestrators*, *aggregators*, and *niche specialists* reveals how the economies of scale is driven by infrastructure ownership, and the economies of scope and specialization are relied upon by niche actors. While dominant MLaaS platforms leverage extensive integration and infrastructural control to extend market reach and create lock-in effects, niche specialists sustain their competitive advantages through domain-specific expertise, proprietary data, and customized workflows that are difficult for large players to replicate effectively.

Therefore, the diversity identified in MLaaS platform business model archetypes supports both Luitse's (2024) and van der Vlist et al.'s (2024) arguments about the strategic orchestration and control inherent to AI platform ecosystems. However, this complexity also suggests that despite the powerful envelopment strategies of Big Tech, there remains space for specialized platforms capable of serving unique or underserved market segments.

## 9.4 The Role of Marketplaces and Referral Mechanism in Value Delivery

The findings from this study highlight the central importance of referral mechanisms and marketplaces in shaping the value delivery of MLaaS platforms. This research demonstrates that MLaaS platforms are highly interdependent with partner directories and referral systems. These systems serve as formalized networks that connect them to technology, consulting, and systems integration partners. These structured referral mechanisms frequently precede the development of fully operational marketplaces, especially within the more prominent *data orchestrator* and *cloud orchestrator* MLaaS platforms (Eisenmann et al., 2011). This contrast highlights a possible sequential logic in platform development, where building trust and establishing reliable partner relationships lays the groundwork for the subsequent extension and scaling of the platform ecosystem through marketplaces.

While *aggregator* platforms often focus on creating a directory or referral system to connect users to specialized third-party providers, dominant *data* and *cloud orchestrator* platforms tend to integrate both partner directories and marketplaces within their ecosystems. Leveraging two-sided network effects and complementary assets allows MLaaS platforms to increase user stickiness and rapidly expand the breadth of their service offerings (Hagiu & Wright, 2015a; Rochet & Tirole, 2003). By providing structured entry points for both users and partners, these mechanisms reduce sales and marketing costs while enhancing platform's competitiveness without requiring internal development of every specialized service. Meanwhile, the concept of the marketplace, which is central to digital platform theory, is defined as a managed environment where buyers and sellers interact under rules set by the platform owner (Hagiu & Wright, 2015b). For MLaaS platforms, this theoretical ideal is realized as curated hubs, such as Amazon SageMaker's model zoo, where end users can discover, evaluate, and directly access third-party tools, datasets, and machine learning models. The use of such marketplaces directly aligns with the *integrated resources* dimension established in the business model taxonomy presented in this research, illustrating how MLaaS platforms aggregate resources and connect diverse actors within their ecosystem.

This aggregation has clear strategic implications. By assembling and curating a diverse set of external resources and services, MLaaS platforms deepen user dependence on their ecosystems (Teece, 1986). When third-party resources and services are fully integrated into the platform's own environment, users have little incentive to switch to competitor platforms, especially once access to external innovations is embedded into the user' critical business workflows. This mutual reinforcement of platform value and user dependency echoes the classic notion of demand-side economies of scope: as third-party participation increases, the variety of available solutions expands, directly increasing overall platform value (Teece, 2010). For instance, when Amazon SageMaker users gain immediate access to the machine learning resources and services through the AWS Marketplace, their workflow becomes more tightly coupled to AWS's compute and storage infrastructure. This encourages both platform users and partners to sustain and expand their engagement with the platform, reinforcing the symbiotic relationship between platforms and partners. (Parker et al., 2016). This creates a virtuous cycle of adoption, innovation, and revenue sharing.

In addition to marketplaces, the structured use of referral mechanisms through partner directories plays a key role in the MLaaS platform ecosystems. These directories not only signal credibility and trust but also act as portal to certified technology providers and consulting firms (Biyalogorsky et al., 2001). For the MLaaS platform business model, these directories are critical to the "key partners" dimension by formalizing relationships that support necessary implementation and customization. This is especially important for enterprise clients with complex, high-stakes deployment needs. MLaaS platforms such as Databricks and Snowflake demonstrate this by maintaining extensive, thoroughly vetted partner networks, including global system integrators such as Deloitte and Accenture, that provide specialized industry expertise and tailored professional services. In these cases, partner directories are not just listings but curated access points to a robust ecosystem of trusted service providers whose reputations are interconnected with that of the MLaaS platform itself. These referral mechanisms offers mutual benefits. Platforms benefit from reduced in-house support burdens and can scale professional services flexibly, while partners receive steady deal flow and gain credibility from their association with leading MLaaS platforms. For users, especially those in regulated or complex domains, such as biotech and pharmaceutical, these curated relationships provide assurance regarding the reliability, competence, and risk management associated with model deployment and ongoing support. The interdependency between platforms, and users exemplifies two-sided network effects, where increasing user adoption makes the platform more attractive to new partners, and vice versa (Rochet & Tirole, 2003).

However, these ecosystem-driven strategies are not without challenges or risks. A core issue arises when MLaaS platforms become heavily reliant on external partners for delivering critical functionalities. Over time, platform dependency can increase, exposing the platform owner to risks if a major partner withdraws their solution, alters their product roadmap, or fails to uphold agreed-upon service standards (Eisenmann et al., 2011). Furthermore, as third-party offerings become essential components of user workflows, the task of ongoing partner curation and quality assurance grows in both complexity and significance. Inconsistencies in service quality or misaligned incentives among partners can undermine user trust or damage the platform’s reputation, especially in enterprise deployments with stringent requirements and significant risk exposure (Jacobides et al., 2018). The challenge is compounded by considerations of scale. As a MLaaS platform’s ecosystem expands, the effort required to monitor, audit, and continuously update partner directories and marketplace participants intensifies. Ensuring quality control and strategic alignment among a growing roster of partners becomes a potential bottleneck that can impact the integrity and perceived value of the core platform. This points to the crucial need for robust partner management frameworks, continuous quality monitoring, and a governance model that supports both partner growth and user trust.

# 10. Conclusion

This research aimed to systematically conceptualize the business models of MLaaS platforms by developing a business model taxonomy and identifying common archetypes based on their key business model dimensions and characteristics. The central research question guiding this work was: "*How can the business models of MLaaS platforms be systematically classified into a taxonomy, and what common archetypes emerge from the combination of their key dimensions and characteristics?*"

The study demonstrates that MLaaS platform business models can be effectively conceptualized using a hierarchical conceptual framework based on Osterwalder et al.'s (2005) Business Model Concept Hierarchy. The meta-model successfully integrated Teece's (2010) value creation, delivery, and capture mechanisms with the components of Osterwalder and Pigneur's (2010) Business Model Canvas. Despite the lack of literature directly related to MLaaS platforms, the preliminary taxonomy demonstrated that adapting findings from existing literature from adjacent domains was sufficient during the conceptual-to-empirical phase (Nickerson et al., 2013). The study employed a scoping review to formulate the working definition of "*MLaaS platform business model*" (Arksey & O'Malley, 2005; Tricco et al., 2015), followed by a rapid review (Ganann et al., 2010) to extract key dimensions and characteristics from business model taxonomies literature. The developed taxonomy systematically illustrated how MLaaS platforms created, delivered, and captured value through the identification of their business model characteristics, with the meta-model using Business Model Canvas components as definitional guidance. The taxonomy captured key characteristics of MLaaS platform business models through its dimensions and characteristics. For value creation, deployment flexibility, user control, model scope, model deployment, complementary services, supported activities, integrated resources, data source, data flow, and data type were identified as business model taxonomy dimensions. These dimensions revealed how platforms differed in the varying level of automation provided by each platform, types of resources integrated with the platform, and the degree of flexibility and control in the value proposition offered to users. For value delivery, the taxonomy included key partners, customer relationships, industry focus, and sales channels dimensions. For value capture, the taxonomy captured pricing models and revenue streams dimensions, showing the varying monetization models used by different MLaaS platforms.

Following the evaluation approach using Szopinski et al.'s (2020) guidelines, the taxonomy demonstrated structural validity, indicating comprehensive coverage of relevant aspects, mutual exclusivity without overlap between dimensions or characteristics, and clarity through precise terminology. Practical usability was assessed using criteria set out by Miles and Huberman (1994), focusing on the taxonomy's robustness, theoretical grounding, and utility for both researchers and practitioners. The taxonomy evaluation confirmed that the taxonomy could accurately classify MLaaS platforms without redundancy or gaps, demonstrating comprehensive coverage and functional utility. Four distinct archetypes emerged following cross-case analysis of all 24 cases. The *cloud orchestrator* archetype represented platforms with integrated cloud infrastructure that achieve horizontal scale and broad applicability, such as Amazon SageMaker, Google Vertex AI, and Microsoft Azure AI. *Data orchestrators* focused on a data-centric value proposition by integrating and managing diverse data resources, such as data assets from data brokers, through marketplaces integrated with their platforms, including MLaaS platforms such as Snowflake, Databricks Mosaic AI, Dataloop, and Domino Data Lab. Aggregator platforms, such as DataRobot, BigML, Dataiku, H2O.ai, stood out for having a partner ecosystem and their use of tiered pricing model. Niche specialist platforms, like HyperSense, Superb AI, ZAMS, PrimeHub, CodeOcean and Barbara, demonstrated deep specialization by targeting specific industries or use cases with tailored offerings.

The research process acknowledged several methodological constraints that affected both case selection and taxonomy development. The taxonomy evaluation relied on four primary cases, enabling depth in within-case analysis but limiting generalizability. A desk research approach was employed because business model data is often proprietary, which limited access to deeper strategic insights such as partner ecosystems and pricing strategies (Miles & Huberman, 1994). The interpretivist nature of this research, heavily reliant on a single researcher's perspective, introduces potential researcher bias. However, due to time constraints, the taxonomy and archetypes were not directly validated with practitioners such as MLaaS platform operators or consultants familiar with the domain, indicating a possible gap between this study's evaluation using theoretical criteria and real-world applicability. Future research should address these limitations through several directions that can be categorized into *macro*-level investigations of MLaaS platform ecosystems and *micro*-level analyses of value creation mechanisms. At the *macro* level, future research could examine MLaaS platforms within the broader context of platform ecosystems and value networks, particularly how cloud providers' envelopment strategies reshape the business models of smaller players over time (Eisenmann et al., 2011). At the *micro* level, deeper exploration of value creation mechanisms could refine the taxonomy's granularity, including assessment of customers' perceived value and value co-creation in marketplaces curated by MLaaS platforms. A follow-up qualitative study using expert interviews and case studies could validate the practical usability and improve the taxonomy, while quantitative methods such as cluster analysis could test the prevalence of archetypes across larger samples, enhancing generalizability.

This thesis makes significant contributions to academic research by developing a meta-model, taxonomy, and archetypes that systematically conceptualize the business models of MLaaS platforms. Unlike prior studies that conflate MLaaS with broader AIaaS models (Lins et al., 2021; Geske et al., 2021), this research delineates MLaaS platforms as distinct entities that facilitate end-to-end machine learning workflows while operating within multisided ecosystems. The taxonomy addresses a major gap in prior research, which has largely focused on technical functionalities (Zou et al., 2024; Mei et al., 2025) while neglecting how these platforms monetize infrastructure and tools (Lins et al., 2021). The derived archetypes advance theoretical understanding by illustrating how MLaaS platforms combine and prioritize specific dimensions to remain competitive, particularly in relation to platform envelopment and Big Tech dominance. While this research confirms dependency on cloud infrastructure largely controlled by Big Tech, it challenges the deterministic view that "*there is no AI without Big Tech*" (van der Vlist et al., 2024) by identifying specific business model configurations that enable coexistence of different archetypes. Also, the practical implications extend to multiple stakeholder groups. For MLaaS platform operators, particularly startups and new market entrants, the taxonomy serves as a diagnostic tool to identify their strategic positioning and design their business models. The archetypes provide prescriptive guidance for value creation, demonstrating that platforms adopting the aggregator model achieve differentiation through integration with modular tools from technology partners rather than infrastructure ownership. For established platforms, the archetypes serve as benchmarking tools, helping assess progression along integrated resources dimensions and anticipating knock-on effects across other dimensions when transitioning between archetypes. The findings further reveal broader implications for understanding the platformization of AI and market concentration dynamics. The research demonstrates that platform envelopment represents a dominant business strategy among hyperscale providers, where *cloud orchestrators* leverage their dominant market position through bundling MLaaS functionalities into existing platforms (Eisenmann et al., 2011). However, *niche specialists* demonstrate sustainable competitive positioning through isolating barriers that are costly or complex for large platforms to replicate authentically (Barney, 1991). This finding enriches platform theory by showing that envelopment is not inevitable, but rather mediated by business model choices that prioritize either scalability or vertical specialization.

As MLaaS platforms become critical infrastructure for AI innovation, similar to operating systems in software development, understanding the business model dimensions and characteristics of MLaaS platforms becomes essential for predicting how their technological, economic, and governance structures collectively shape the AI landscape. The business model taxonomy and archetypes provide foundational tools for future research examining platform strategy, AI governance, and market dynamics in the era of platformized AI. This research establishes the necessary framework for systematic analysis of MLaaS platform business models and contributes essential understanding of how these platforms influence AI adoption, innovation, and control across the industries.

## 10.1 Declaration of AI Usage

In conducting this thesis, several AI tools were used to enhance the writing quality in this report. Perplexity was used primarily to identify original academic sources related to key theoretical concepts and to verify whether specific claims were supported by scholarly literature. For the glossary section, Perplexity was also used to rephrase definitions, ensuring each entry was limited to fewer than ten words for clarity and conciseness. Grammarly served as an essential tool to edit and correct grammatical issues throughout the text, helping to maintain a high standard of academic writing. Collectively, these AI tools supplemented but did not replace the researcher’s analytical judgment or independent academic research. Their use is declared here to maintain transparency regarding research integrity and the methodology underpinning the thesis writing process.

# References

[1] Afuah, A., & Tucci, C. L. (2000). Internet business models and strategies: Text and cases. Boston: *McGraw-Hill Higher Education*.

[2] Al-Debei, M. M., & Avison, D. (2010). Developing a unified framework of the business model concept. *European Journal of Information Systems*, 19(3), 359-376.

[3] Amit, R., & Zott, C. (2001). Value creation in e-business. *Strategic Management Journal*, 22(6-7), 493–520. https://doi.org/10.1002/smj.187

[4] Arksey H, O'Malley L. Scoping studies: towards a methodological framework. *Int J Soc Res Methodol*. 2005;8:19–32. doi:10.1080/1364557032000119616.

[5] Bailey, K. D. (1994). Typologies and Taxonomies: An Introduction to Classification Techniques. *SAGE Publications*.

[6] Baldwin, C.Y., and Clark, K.B. (2000). Design Rules, Vol. 1: The Power of Modularity, Cambridge, MA: *MIT Press*.

[7] Barney, J. (1991). Firm resources and sustained competitive advantage. *Journal of Management*, 17(1), 99–120.

[8] Bartels, N., Koch, M., & Gordijn, J. (2023). Developing a Taxonomy for Revenue Models of Platform Business Models. *36th Bled eConference – Digital Economy and Society: The Balancing Act for Digital Innovation in Times of Instability*: June 25 – 28, 2023, Bled, Slovenia, Conference Proceedings, 1–18. https://doi.org/10.18690/um.fov.6.2023.1

[9] Belussi, F., Orsi, L., & Savarese, M. (2019). Mapping business model research: A document bibliometric analysis. *Scandinavian Journal of Management*, 35(3), 101048.

[10] Biyalogorsky, E., Gerstner, E., & Libai, B. (2001). Customer referral management: Optimal reward programs. Marketing Science, 20(1), 82-95.

[11] Bryman, A. (2016). Social research methods (5th ed.). *Oxford University Press*.

[12] Chesbrough, H., & Rosenbloom, R. S. (2002). The role of the business model in capturing value from innovation. *Industrial and Corporate Change*, 11(3), 529–555.

[13] Clauss, T., 2017. Measuring business model innovation: conceptualization, scale development, and proof of performance. *R&D Manag.* 47 (3), 385–403.

[14] Clauss, T., Harengel, P., Hock, M., 2019. The perception of value of platform-based business models in the sharing economy: determining the drivers of user loyalty. *Rev. Manag. Sci.* 13, 605–634.

[15] Cobbe, J., & Singh, J. (2024). Accounting for context in AI technologies. In Handbook on Public Policy and Artificial Intelligence (pp. 94-108). *Edward Elgar Publishing*.

[16] Cozzolino, A., Corbo, L., & Aversa, P. (2021). Digital platform-based ecosystems: The evolution of collaboration and competition between incumbent producers and entrant platforms. *Journal of Business Research*, 126, 385-400.

[17] Cronk, M. C., & Fitzgerald, E. P. (1999). Understanding "IS business value": derivation of dimensions. *Logistics Information Management*, 12(1/2), 40-49.

[18] Cusumano, M. A., Gawer, A., & Yoffie, D. B. (2019). The business of platforms: Strategy in the age of digital competition, innovation, and power (Vol. 320). New York: *Harper Business*.

[19] Demil, B., & Lecocq, X. (2010). Business model evolution: in search of dynamic consistency. *Long Range Planning*, 43(2), 227–246.

[20] De Reuver, M., Sørensen, C., & Basole, R. (2018). The digital platform: A research agenda. *Journal of Information Technology*, 33(2), 124–135.

[21] Doty, D. H., & Glick, W. H. (1994). Typologies as a unique form of theory building: Toward improved understanding and modeling. *Academy of Management Review*, 19(2), 230-251.

[22] Durao, F., Carvalho, J. F. S., Fonseka, A., & Garcia, V. C. (2014). A systematic review on cloud computing. *The Journal of Supercomputing*, 68(3), 1321-1346.

[23] Edmondson, A. C., & McManus, S. E. (2007). Methodological fit in management field research. *Academy of Management Review*, 32(4), 1246-1264. https://doi.org/10.5465/amr.2007.26586086

[24] Ejsmont, K., Gladysz, B., Roczon, N., Bettoni, A., Barut, Z. M., Haber, R., & Minisci, E. (2024). Multisided Business Model for Platform Offering AI Services. In Artificial Intelligence in Manufacturing (pp. 105–120). *Springer*, Cham. https://doi.org/10.1007/978-3-031-46452-2_7

[25] Eisenmann, T., Parker, G., & Van Alstyne, M. (2006). Strategies for two-sided markets. *Harvard Business Review, 84*(10), 92–101.

[26] Eisenmann, T., Parker, G., & Van Alstyne, M. (2011). Platform envelopment. *Strategic management journal*, 32(12), 1270-1285.

[27] Eisenhardt, K. M. (1989). Building theories from case study research. *Academy of Management Review,* 14(4), 532-550.

[28] Elo, S., & Kyngäs, H. (2008). The qualitative content analysis process. *Journal of advanced nursing*, 62(1), 107-115.

[29] Engelbrecht, A., Gerlach, J., & Widjaja, T. (2016). Understanding the anatomy of data-driven business models–towards an empirical taxonomy.

[30] Evans, D. S., & Schmalensee, R. (2013). The antitrust analysis of multisided platform businesses. *NBER Working Paper No. 18783*.

[31] Freichel, C., Fieger, J., & Winkelmann, A. (2021). Developing a taxonomy for digital platforms – A conceptual approach. *Proceedings of the 54th Hawaii International Conference on System Sciences*, 5779-5788.

[32] G2 Crowd Inc. (2025, March 25). G2. Retrieved April 11, 2025, from https://www.g2.com/categories/artificial-intelligence

[33] Ganann, R., Ciliska, D., & Thomas, H. (2010). Expediting systematic reviews: Methods and implications of rapid reviews. *Implementation Science*, 5(1), 56. https://doi.org/10.1186/1748-5908-5-56

[34] Gassmann, O., & Frankenberger, K. (2014). The Business Model Navigator: 55 Models That Will Revolutionise Your Business, Pearson Education Limited, *Harlow*, 2014.

[35] Gawer, A., & Cusumano, M. A. (2014). Industry platforms and ecosystem innovation. *Journal of Product Innovation Management*, 31(3), 417–433.

[36] Gawer, A. (2014). Bridging differing perspectives on technological platforms: Toward an integrative framework. *Research Policy, 43*(7), 1239-1249.

[37] Gawer, A. (2020). Digital platforms' boundaries: The interplay of firm scope, platform sides, and digital interfaces. *Long Range Planning*, 53(4), 101890.

[38] Geske, F., Hofmann, P., Lämmermann, L., Schlatt, V., & Urbach, N. (2021). Gateways to artificial intelligence: developing a taxonomy for AI service platforms.

[39] Ghazawneh, A., & Henfridsson, O. (2013). Balancing platform control and external contribution in third-party development: The boundary resources model. *Information Systems Journal*, 23(2), 173–192. https://doi.org/10.1111/j.1365-2575.2012.00406.x

[40] Glaser, B. G., & Strauss, A. L. (1967). The discovery of grounded theory: Strategies for qualitative research. *Aldine*.

[41] Gregor, S. (2006). The nature of theory in information systems. *MIS quarterly*, 611-642.

[42] Gregory, R. W., Henfridsson, O., Kaganer, E., & Kyriakou, H. (2021). The role of artificial intelligence and data network effects for creating user value. *Academy of management review*, *46*(3), 534-551.

[43] Guest, G., Bunce, A., & Johnson, L. (2006). How many interviews are enough? Field Methods, 18(1), 59-82.

[44] Haftor, D. M., Climent, R. C., & Lundström, J. E. (2021). How machine learning activates data network effects in business models: Theory advancement through an industrial case of promoting ecological sustainability. *Journal of Business Research*, *131*, 196-205.

[45] Hagiu, A., & Wright, J. (2015a). Multi-sided platforms. International Journal of Industrial Organization, 43, 162–174. https://doi.org/10.1016/j.ijindorg.2015.03.003

[46] Hagiu, A., & Wright, J. (2015b). Marketplace or reseller?. *Management Science*, 61(1), 184-203.

[47] Helmond, A. (2015). The platformization of the web: Making web data platform ready. Social media+ society, 1(2), 2056305115603080.

[48] Hofmann, P., Rückel, T., & Urbach, N. (2021). Innovating with artificial intelligence: capturing the constructive functional capabilities of deep generative learning.

[49] Jacobides, M. G., Cennamo, C., & Gawer, A. (2018). Towards a theory of ecosystems. *Strategic Management Journal*, 39(8), 2255–2276.

[50] Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245)

[51] Kapoor, R., & Teece, D. J. (2021). Three faces of technology’s value creation: Emerging, enabling, embedding. *Strategy Science*, 6(1), 1-4., 255–260.

[52] Kortmann, S., & Piller, F. (2016). Open business models and closed-loop value chains: Redefining the firm-consumer relationship. *California Management Review*, *58*(3), 88-108.

[53] Krippendorff, K. (2018). Content analysis: An introduction to its methodology (4th ed.). *SAGE*.

[54] Lambert, S. (2015). The importance of classification to business model research. *Journal of Business Models*, 3(1).

[55] Lewicki, K., Lee, M. S. A., Cobbe, J., & Singh, J. (2023, April). Out of context: Investigating the bias and fairness concerns of “artificial intelligence as a service”. *In Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (pp. 1-17).

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

[57] Luitse, D. (2024). Platform power in AI: The evolution of cloud infrastructures in the political economy of artificial intelligence. *Internet Policy Review*, 13(2).

[58] Ma, Y., Shen, Z., & Shen, J. (2024). Cloud Computing and Hyperscale Data Centers: A Comparative Study of Usage Patterns. *Journal of Theory and Practice of Engineering Science*, 4(06), 11-19.

[59] Magretta J (2002) Why business models matter. *HBR* 80:86–92

[60] Makarius, E. E., Mukherjee, D., Fox, J. D., & Fox, A. K. (2020). Rising with the machines: A sociotechnical framework for bringing artificial intelligence into the organization. *Journal of business research*, 120, 262-273.

[61] Massa, L., Tucci, C. L., & Afuah, A. (2017). A critical assessment of business model research. *Academy of Management annals*, *11*(1), 73-104.

[62] Mayring, P. (2014). Qualitative content analysis: Theoretical foundation, basic procedures and software solution.

[63] Mei, H., Zhang, Y., Peng, N. et al. Machine learning as a service system for particle accelerator and its application in CSNS. *Radiat Detect Technol Methods* (2025). https://doi-org.tudelft.idm.oclc.org/10.1007/s41605-025-00527-7

[64] Metelskaia, I., Ignatyeva, O., Denef, S., & Samsonowa, T. (2018). A business model template for AI solutions. *Proceedings of the International Conference on Intelligent Science and Technology*, 35–41. https://doi.org/10.1145/3233740.3233750

[65] Miceli, M., Posada, J., & Yang, T. (2022). Studying Up Machine Learning Data: Why Talk About Bias When We Mean Power? *Proceedings of the ACM on Human-Computer Interaction*, *6*(GROUP), 1-14.

[66] Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Sage.

[67] Möller, F., Stachon, M., Azkan, C., Schoormann, T., & Otto, B. (2022). Designing business model taxonomies – synthesis and guidance from information systems research. *Electronic Markets*, *32*(2), 701–726. Scopus. https://doi.org/10.1007/s12525-021-00507-x

[68] Morris, M., Schindehutte, M., & Allen, J. (2005). The Entrepreneur’s Business Model: Toward a Unified Perspective. *Journal of Business Research*, 58(6), 726–735. https://doi.org/10.1016/j.jbusr es.2003.11.001

[69] Mwilu, O. S., Prat, N., & Comyn-Wattiau, I. (2015). Taxonomy development for complex emerging technologies–The case of Business Intelligence and Analytics On the cloud. In *PACIS 2015*.

[70] Mucha, T., & Seppälä, T. (2020). Artificial intelligence platforms: A new research agenda for digital platform economy. *ETLA Working Papers*, 76.

*[71]* Naous, D., Schwarz, J., & Legner, C. (2017). Analytics as a service: Cloud computing and the trans-formation of business analytics business models and ecosystems. In *Proceedings of the 25th european conference on inform*

[72] Ng, A. (2025, March 25). Andrew Ng: Why Agentic AI is the smart bet for most enterprises. *Insight Partners.* https://www.insightpartners.com/ideas/andrew-ng-why-agentic-ai-is-the-smart-bet-for-most-enterprises/

[73] Nickerson, R. C., Varshney, U., & Muntermann, J. (2013). A method for taxonomy development and its application in information systems. *European Journal of Information Systems*, 22(3), 336–359.

[74] Osterwalder, A., Pigneur, Y., & Tucci, C. L. (2005). Clarifying Business Models: Origins, Present, and Future of the Concept. *Communications of the Association for Information Systems*, *16*(1). https://doi.org/10.17705/1CAIS.01601

[75] Osterwalder, A., & Pigneur, Y. (2010). Business model generation: A handbook for visionaries, game changers, and challengers. *Wiley*.

[76] Osterwalder, A., & Pigneur, Y. (2013). Designing business models and similar strategic objects: The contribution of IS. *Journal of the Association for information systems*, *14*(5).

[77] Parker, G. G., Van Alstyne, M. W., & Choudary, S. P. (2016). Platform revolution: How networked markets are transforming the economy—and how to make them work for you. *WW Norton & Company*.

[78] Peters, M. D., Marnie, C., Tricco, A. C., Pollock, D., Munn, Z., Alexander, L., ... & Khalil, H. (2020). Updated methodological guidance for the conduct of scoping reviews. *JBI evidence synthesis*, 18(10), 2119-2126.

[79] Philipp, R., Mladenow, A., Strauss, C., & Völz, A. (2020). Machine Learning as a Service: Challenges in Research and Applications. *Proceedings of the 22nd International Conference on Information Integration and Web-Based Applications & Services*, 396–406. https://doi.org/10.1145/3428757.3429152

[80] Priem, R. L., Wenzel, M., & Koch, J. (2018). Demand-side strategy and business models: Putting value creation for consumers center stage. *Long range planning*, 51(1), 22-31.

[81] Rai, A., Constantinides, P., & Sarker, S. (2019). Next generation digital platforms: Toward human-AI hybrids. *MIS Quarterly*, 43(1), Article 1.

[82] Rappa, M. (2001). Business models on the web. *North Carolina State University*. ecommerce.ncsu.edu.

[83] Remane, G., Nickerson, R., Hanelt, A., Tesch, J. F., & Kolbe, L. M. (2016, December). A Taxonomy of Carsharing Business Models. *In ICIS* (pp. 1-17).

[84] Remane, G., Hanelt, A., Nickerson, R. C., & Kolbe, L. M. (2017). Discovering digital business models in traditional industries. *Journal of Business Strategy*, 38(2), 41-51.

[85] Ribeiro, M., Grolinger, K., & Capretz, M. A. M. (2015). MLaaS: Machine Learning as a Service. *In 2015 IEEE 14th International Conference on Machine Learning and Applications* (ICMLA) (pp. 896–902). IEEE. https://doi.org/10.1109/ICMLA.2015.151

[86] Ritter, T., & Lettl, C. (2018). The wider implications of business-model research. *Long range planning*, 51(1), 1-8.

[87] Rochet, J.-C., & Tirole, J. (2003). Platform competition in two-sided markets. *Journal of the European Economic Association, 1*(4), 990–1029.

[88] Shafer, S. M., Smith, H. J., & Linder, J. C. (2005). The power of business models. *Business horizons*, 48(3), 199-207.

[89] Shapiro, C., & Varian, H. R. (1998). *Information rules: A strategic guide to the network economy*. Harvard Business Press.

[90] Shree, D., Singh, R. K., Paul, J., Hao, A., & Xu, S. (2021). Digital platforms for business-to-business markets: A systematic review and future research agenda. *Journal of Business Research*, 137, 354-365.

[91] Schwandt, T. A. (1994). Constructivist, interpretivist approaches to human inquiry. *In N. K. Denzin & Y.*

[92] Sokal, R. & Sneath P. H. A. (1963) Principles of Numerical Taxonomy. *W.H. Freeman and Company*, San Francisco, CA.

[93] Staub, N., Haki, K., Aier, S., & Winter, R. (2021). Taxonomy of Digital Platforms: A Business Model Perspective. *Hawaii International Conference on System Sciences.* https://doi.org/10.24251/HICSS.2021.744

[94] Sundberg, L. T., Holmström, J., & CCiittaattiioonn, Rr. (2022). Towards 'Lightweight' Artificial Intelligence: A Typology of AI Service Platforms. Artificial Intelligence.

[95] Syed, N., Anwar, A., Baig, Z., & Zeadally, S. (2025). Artificial Intelligence as a Service (AIaaS) for Cloud, Fog and the Edge: State-of-the-Art Practices. *ACM Comput. Surv.* https://doi.org/10.1145/3712016

[96] Szopinski, D., Kundisch, D., & Schoormann, T. (2020). Visualize Different: Towards Researching the Fit Between Taxonomy Visualizations and Taxonomy Tasks. In *Wirtschaftsinformatik (Zentrale Tracks)* (pp. 1255-1261).

[97] Szopinski, D., Schoormann, T., & Kundisch, D. (2019). Because your taxonomy is worth it: Towards a framework for taxonomy evaluation. *Proceedings of the Twenty-Seventh European Conference on Information Systems*, Uppsala, Sweden.

[98] Täuscher, K., & Laudien, S. (2017). Uncovering the nature of platform-based business models: An empirical taxonomy. *Proceedings of the 50th Hawaii International Conference on System Sciences*, 5285-5294.

[99] Teece, D. J. (1986). Profiting from technological innovation: Implications for integration, collaboration, licensing and public policy. *Research Policy*, 15(6), 285-305.

[100] Teece, D. J. (2010). Business models, business strategy and innovation. *Long Range Planning*, 43(2–3), 172–194.

[101] Teece, D. J., & Linden, G. (2017). Business models, value capture, and the digital enterprise. *Journal of Organization Design*, 6(1). Scopus. https://doi.org/10.1186/s41469-017-0018-x

[102] Timmers, P. (1998). Business models for electronic markets. *Electronic Markets*, 8(2), 3–8.

[103] Tiwana, A. (2014). Platform ecosystems: Aligning architecture, governance, and strategy. *Morgan Kaufmann*.

[104] Tricco, A. C., Antony, J., Zarin, W., Strifler, L., Ghassemi, M., Ivory, J., ... & Straus, S. E. (2015). A scoping review of rapid review methods. *BMC Medicine*, 13(1), 224. https://doi.org/10.1186/s12916-015-0465-6

[105] Van Alstyne, M. W., Parker, G. G., & Choudary, S. P. (2016). Pipelines, platforms, and the new rules of strategy. *Harvard business review*, 94(4), 54-62.

[106] Van Der Vlist, F., Helmond, A., & Ferrari, F. (2024). Big AI: Cloud infrastructure dependence and the industrialisation of artificial intelligence. *Big Data & Society*, *11*(1), 20539517241232630.

[107] Veit, D., Clemons, E., Benlian, A., Buxmann, P., Hess, T., Kundisch, D., ... & Spann, M. (2014). Business models: An information systems research agenda. *Business & Information Systems Engineering*, *6*, 45-53.

[108] Weber, M., Beutter, M., Weking, J., Böhm, M., & Krcmar, H. (2022). AI Startup Business Models. *Business & Information Systems Engineering*, *64*(1), 91–109. https://doi.org/10.1007/s12599-021-00732-w

[109] Whittaker, M. (2021). "The Steep Cost of Capture." *Interactive Workshop on the Politics of AI*.

[110] Yang, Y & Gokturk, B (2023) Google Cloud brings generative AI to developers, businesses, and governments. In*: Google Cloud Blog*, 14 March 2023. Available at: https://cloud.google.com/blog/products/ai-machine-learning/generative-ai-for-businesses-and-governments

[111] Yanow, D. (2014). Interpretive analysis and comparative research. In E. Goodin (Ed.), The Oxford handbook of contextual political analysis (pp. 159-174). *Oxford University Press.*

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

[113] Zhao, Y., Von Delft, S., Morgan-Thomas, A., & Buck, T. (2020). The evolution of platform business models: Exploring competitive battles in the world of platforms. *Long Range Planning*, 53(4), 101892.

[114] Zou, Q., Deng, Y., Zhu, Y., Zhou, Y., Cai, J., & He, S. (2024). Dissecting I/O Burstiness in Machine Learning Cloud Platform: A Case Study on Alibaba's MLaaS.

[115] Zott, C., & Amit, R. (2010). Business model design: An activity system perspective. *Long Range Planning*, 43(2–3), 216–226.

# Appendix

## A. Literature Review Findings

### A1 Scoping Review Application, Result and Analysis

Given the fragmented nature of existing research, we decompose the problem into three pillars: cloud computing as the technical backbone, MLaaS as the delivery mechanism, and digital platforms as the business architecture. This follows the researcher's existing understanding of interrelation between these domains. We searched Scopus using a comprehensive search strategy (see Table 19) across four key areas: business models, MlaaS platform, cloud computing, and digital platform. The findings from scoping review have been integrated into Chapter 2 and 4 of this report.

*Table 19: Overview of keywords used*

| Topic | Search term | Limit |
|---|---|---|
| Business models | (( "business model" ) AND ( "definition" OR "concept" OR "platform" )) | > 45 citations |
| MlaaS Platform | ("AI service platform" OR "AI-as-a-Service" OR "AIaaS" OR “AI Platform”) OR ("machine learning platform" OR "MLaaS" OR "machine learning as a service") | > 10 citations |
| Cloud Computing | ( "cloud " ) AND ( "AI" )<br>( "cloud" ) AND ( "agenda" OR "survey" OR "classification" OR "taxonomy" ) | > 45 citations |
| Digital Platform | ( "digital platform") AND ( "agenda" OR "survey" OR "classification" OR "taxonomy" ) | > 45 citations |

The inclusion criteria targeted peer-reviewed articles (2015–2024) explicitly discussing business model definition, focusing on foundational theories and concepts focused on IS domains, while excluding purely technical or non-platform MLaaS studies. From an initial pre-screened pool of 1,570 records, screening of title, abstract and article, over around 60 hours, retained 33 papers. The final 19 papers were kept after full text analysis (over around 15 hours) , leading to exclusion of papers focusing on topics like circular business models, technical proofs-of-concept, cloud security architectures, or industry-specific platform applications.

The search scope was deliberately constrained to works, focusing on literature reviews, surveys, published between 2014-2025 within key disciplines including Computer Science, Business Management, Economics, and Social Sciences. The initial pre-screened search result revealed significant concentration in MLaaS and MLaaS Platform compared to business model literature. MLaaS-related studies showed strong representation in both technical integration and survey-focused papers, while digital platform studies yielded fewer but more focused results. Following full text analysis, the scoping review (N=19) reveals a strong focus on business models (7) and digital platforms (8), with most studies published between 2018-2021 (11) (see Table 20). While MLaaS-specific research remains limited (5), definitions are clearly established in 16 studies. The dominance of journal articles (18) indicates a mature, peer-reviewed knowledge base, contrasting with minimal conference contributions (1).

*Table 20: Summary of study characteristics*

| Study Characteristics | Count (N=19) |
|---|---|
| 2014 – 2017 | 5 |
| 2018 – 2021 | 11 |
| 2022 – 2025 | 3 |
| Definitions | |
| Yes | 16 |
| No | 3 |
| Business model | 7 |
| Cloud computing | 3 |
| MLaaS | 5 |
| Digital platform | 8 |
| Publication type | |
| Journal article | 18 |
| Conference paper | 1 |

## A2 Rapid Review Application, Result and Analysis

The main purpose of this rapid review is to assess the existing business model taxonomies within the MLaaS platform and wider domains such as platform business models and AI-specific business models. We apply this method during conceptual-to-empirical iteration in our taxonomy development process because we rely on the existing taxonomies to inform the dimensions that our preliminary taxonomy should include. In line with our meta-characteristic's focus on Teece's value creation, delivery and capture mechanisms, this objective frames the specific questions that the rapid review aims to address: "*what are the business model components that can be used as dimensions in our taxonomy?*" and "*what are the dimensions identified in the existing business model taxonomies from the literature that also applies for MLaaS platform business model taxonomy?*"

*Table 21: Search criteria*

| Search Criteria | Scope |
|---|---|
| Content | Focused on MLaaS platforms or related domains, expanding to platform business model taxonomies. Excludes digital marketplaces and non-AI taxonomies without business model dimensions. |
| Language | English publications only. |
| Research Design | Non-technical papers within IS, business and management domain. |
| Publication Date | Last 10 years, with backward reference searching permitted for foundational sources. |

We focus our search strategy on the literature with the level of analysis that can apply at platform-level and is aligned with our meta-model (see Chapter 4). Our goal is to identify the dimensions that should be included in our taxonomy. Based on the insights from the scoping review, we use different combinations of relevant search terms that describe business model and ML-related domains, including "*business model*", "*taxonomy*", "*classification*", "*MLaaS*" *and* "*platform*". We then adapted Mwilu et al.'s (2015) proposed inclusion criteria for developing taxonomies of emerging technologies to narrow down our search scope (see Table 21). We spent an estimated 10 hours to conduct preliminary screening of the title, abstract, and article and identified 26 sources for full reading and analysis. This led to the final 12 papers for inclusion in our rapid review. After reading the selected literature (estimated of 9 hours), common themes amongst all the sources were identified(see Table 22).

*Table 22: Literature related to the MLaaS platform business models*

| Theme | Author (Year) | Title |
|---|---|---|
| AI service platform archetypes | Sundberg et al. (2022) | Towards 'Lightweight' Artificial Intelligence: A Typology of AI Service Platforms |
| AI service platform taxonomy & archetypes | Geske et al. (2021) | Gateways To Artificial Intelligence: Developing A Taxonomy For AI Service Platforms |
| MLaaS platform ecosystem | Lins et al. (2021) | Artificial Intelligence as a Service: Classification and Research Directions |
| Revenue model of platform business model | Ejsmont et al. (2024) | Multisided Business Model for Platform Offering AI Services |
| | Bartels et al. (2023) | Developing a Taxonomy for Revenue Models of Platform Business Models |
| AI-specific business model dimensions | Metelskaia et al. (2018) | A business model template for AI solutions |
| | Weber et al. (2022) | AI Startup Business Models |
| Taxonomy of data-driven business model | Engelbrecht et al. (2016) | Understanding The Anatomy of Data-Driven Business Models – Towards An Empirical Taxonomy |
| | Naous et al. (2017) | Analytics as a Service: Cloud Computing And The Transformation Of Business Analytics Business Models And Ecosystems |
| Taxonomy of platform business model | Freichel et al. (2021) | Developing a Taxonomy for Digital Platforms – A Conceptual Approach |
| | Staub et al. (2021) | Taxonomy of Digital Platforms: A Business Model Perspective |
| | Täuscher & Laudien (2017) | Uncovering the Nature of Platform-based Business Models: An Empirical Taxonomy |

The review (N=12) shows growing post-2018 interest in MLaaS platform domain, though only 2 studies directly address it (see Table 23). Most research focuses on platform business models (5) or taxonomies (7), primarily published as conference papers (7), indicating an emerging field . Limited journal articles (1) suggest need for deeper validation.

*Table 23: Summary of study characteristics*

| Study Characteristics | Count (N=12) |
|---|---|
| 2014 – 2017 | 3 |
| 2018 – 2021 | 4 |
| 2022 – 2025 | 4 |
| Directly related | 2 |
| Adjacent (broader scope) | 4 |
| MLaaS platform ecosystem | 2 |
| Platform business model | 5 |
| MLaaS-related archetypes | 3 |
| Data-driven business model | 2 |
| Taxonomy | 7 |
| Business model taxonomy | 5 |
| Typology | 1 |
| Classification | 1 |
| *Publication type* | |
| Book chapter | 1 |
| Conference paper | 7 |
| Journal article | 1 |
| Research paper | 3 |

Although we found no existing comprehensive business model taxonomy specifically for MLaaS platforms, our rapid review identified relevant business model dimensions from two sources: (1) AI-specific and AI service platform taxonomies and classifications (Geske et al., 2021; Sundberg et al., 2022) that provide dimensions applicable to MLaaS platform business model, and (2) platform and data-driven business model taxonomies that provide platform-specific dimensions.

## B. Case Database

*Table 24: Overview of case database, and segmentation criteria*

| ID | Platform Name | ML Approach | Customer Segment | Extended Offering | Sampling Purpose |
|---|---|---|---|---|---|
| **1** | Barbara | ML(Edge AI) | Enterprise | No | E2C-1 |
| **2** | Superb AI Suite | LLM | Enterprise - Specific | No | C2E-1 |
| **3** | Modular | LLM | All | No | E2C-2 |
| **4** | dataloop | ML | All | No | E2C-3 |
| **5** | BigML | ML | All | No | E2C-4 |
| **6** | Alibaba ML Platform for AI | ML | Enterprise | Yes | Taxonomy analysis |
| **7** | PrimeHub AI Platform | ML | Enterprise | No | E2C-3 |
| **8** | HyperSense | ML | Enterprise | No | C2E-1 |
| **9** | CodeOcean | ML | Enterprise - Specific | No | E2C-2 |
| **10** | Amazon SageMaker | ML/LLM | All | Yes | E2C-1 |
| **11** | Microsft Azure AI | ML/LLM | All | Yes | E2C-4 |
| **12** | Jfrog ML | ML/LLM | All | Yes | Taxonomy analysis |
| **13** | Snowflake | ML/LLM | All | Yes | C2E-1 |
| **14** | Domino Data Lab | ML/LLM | All | No | E2C-1 |
| **15** | H2O | ML/LLM | All | No | Taxonomy analysis |
| **16** | Iguazio | ML/LLM | Enterprise | No | E2C-3 |
| **17** | Databricks Mosaic AI | ML/LLM | Enterprise | Yes | E2C-2 |
| **18** | Dataiku | ML/LLM | Enterprise | Yes | E2C-1 |
| **19** | Google Vertex AI | ML/LLM | Enterprise | Yes | E2C-3 |
| **20** | IBM Watsonx Studio | ML/LLM | Enterprise | Yes | C2E-1 |
| **21** | Cloudera AI | ML/LLM | Enterprise | Yes | Taxonomy analysis |
| **22** | DataRobot | ML/LLM | Enterprise | No | E2C-4 |
| **23** | VessL | LLM | Enterprise - Specific | No | E2C-4 |
| **24** | ZAM | ML/LLM/Agentic AI | All | No | E2C-2 |

# C. Taxonomy Development Process

## C1 Conceptual-to-Empirical Iterations (C2E)

### C1.1 Iteration 1: Conceptual-to-Empirical Iteration (C2E-1)

After reviewing the dimensions from directly relevant literature against our meta model, we identified the following dimensions to be relevant for MLaaS platforms. We then decided on the operations required to transfer these dimensions and their characteristics into MLaaS platforms. All of these dimensions and characteristics are directly sourced from the literature with some adaptations required to fit the context of MLaaS platform (see Table 25). Table 26 presents the operations that were performed during first iteration of our taxonomy development. The taxonomy in Figure 21 reflects the dimensions and characteristics identified at the end of the first iteration.

*Table 25: Overview of dimensions and their sources (C2E-1)*

| Meta-model | | Dimension | Definition | Source |
|---|---|---|---|---|
| First order construct | Second order construct | | | |
| Value Creation | Value Proposition | AI service model | Nature of service offerings | (Geske et al., 2021) |
| | | core VP | Core value for platform operation | (Ejsmont et al., 2022) |
| | | scope | Breadth of applications or use cases | (Sundberg et al., 2022) |
| | | complementary services | Additional tasks that support core VP. | (Geske et al., 2021) |
| | Key Resources | complementary resources | resources that support core VP | (Geske et al., 2021) |
| | | assets | Core components that platform uses | (Ejsmont et al., 2024) |
| | Key Partnerships | integrated complementor | Type of third-party contributors integrated to the platform | (Geske et al., 2021) |
| Value Delivery | Customer Segments | industry focus | The extent of platform's catering towards specific customer segment | (Geske et al., 2021) |
| | Channels | channels | Platform's interface with users. | (Ejsmont et al., 2024) |
| | | graphical interface | The interface mode between the customer and the platform | (Geske et al., 2021) |
| Value Capture | Revenue Streams | pricing model | Pricing mechanisms of the platform | (Geske et al., 2021, Ejsmont et al., 2024) |
| | | revenue streams | Ways to generate revenue | (Ejsmont et al., 2024) |

*Table 26: Inductive operations overview (C2E-1)*

| Operation type | Description |
|---|---|
| Add | Added 12 dimensions from the broader AI service platform literatures |
| Adapt | Adapted “*AI service model*” dimension to “*ML service model*” and characteristic to reflect “*model*” development and deployment instead of application. |
| | Adapted “*Scope*” dimension to “*Model scope*” to add clarity. |
| | Adapted “*free of charge*” characteristic to “*free trial*” |
| Merge | Merged “*graphical interface*” and “*code-based interface*” dimensions to improve conciseness of the dimension. |
| | Merge “*hybrid service*” and “*ready-to-use service*” characteristics to reflect “ready to use” services being the complementary services to main services. |
| Delete | Removed “*ready-to-use service*” characteristic because it's out of our scope. |

| | **Dimension** | **Characteristics** | | | |
|---|---|---|---|---|---|
| **Value Creation** | ML service model | model deployment service | model development & deployment service | hybrid service with ready to use services | |
| | core VP | to be refined | | | |
| | Model scope | context changing | content changing | | |
| | complementary services | yes | no | | |
| | Model scope | context changing | content changing | | |
| | complementary resources | yes | no | | |
| | assets | tangible | intangible | | |
| | integrated complementor | selected complementors | crowd | selected complementors and crowd | none |
| **Value Delivery** | industry focus | industry specific | industry unspecific | | |
| | channel | to be refined | | | |
| | interface | graphical | code-based | none | |
| **Value Capture** | revenue streams | subscriptions | advertising | transaction | profit-sharing |
| | pricing model | one-off payment | regularly based on variable rates | regularly based on fixed rates, | |
| | | freemium | free trial | | |

*Figure 21: Business model taxonomy at the end of first iteration (C2E-1)*

### C1.2 Iteration 2: Conceptual-to-Empirical Iteration (C2E-2)

In the second iteration, we review the dimensions from AI-related business model literature against our meta model. After identifying the following dimensions to be relevant for MLaaS platforms, we decided on the operations required to transfer these dimensions and their characteristics into MLaaS platforms. Based on the works of Lins et al. (2021) and Weber et al. (2022), we identified additional characteristics to refine "*core VP*" dimension (see Table 27). While not directly specified, Lins et al. (2021) discussed the elements of hardware, software, infrastructure, risks, governance and compliance for AIaaS that also applied for MLaaS platform. We aggregated them into tangible (hardware, software, infrastructure) and intangible (risks, governance and compliance) sources for conciseness and clarity. The remaining three dimensions, "*ML lifecycle activities*", "*data strategy*" and "*data type*" were directly sourced from the literature. We did not identify directly relevant dimensions from Metelskaia et al.'s (2021) work. Table 28 shows the operations that were performed during first iteration of our taxonomy development process. The taxonomy in Figure 22 reflects the dimensions and characteristics identified at the end of the second iteration.

*Table 27: Overview of new dimensions identified and their sources (C2E-2)*

| Meta-model | | Dimension (in Source) | Definition (in Source) | Source |
|---|---|---|---|---|
| First order construct | Second order construct | | | |
| Value Creation | Key Activities | ML lifecycle activities | Activities of the full machine learning pipeline supported by the platform | (Lins et al., 2021) |
| | | data strategy | Data sources supported by the platform | (Weber et al., 2022) |
| | | data type | format of data that the platform core VP can support | (Weber et al., 2022) |
| Value Delivery | Customer Segments | customer segments | type of customers the platform serves or targets | (Weber et al., 2022) |
| Value Capture | Cost Structure | costs | The expenses type of the platform | (Lins et al., 2021) |

*Table 28: Inductive operations overview (C2E-2)*

| Operation type | Description |
|---|---|
| Add | Added "automation", "optimization", "insight generation", "continuous learning", "complexity abstraction", "customizability" characteristics to Core VP.<br>Added "*ML lifecycle activities*" dimension and its characteristics<br>Added "*data strategy*" dimension and its characteristics<br>Added "*costs*" dimension and its aggregated characteristics for hardware, software, infrastructure, risks, governance and compliance.<br>Added "*data type*" and its characteristics<br>Added "*customer segment*" and its characteristics |

| | **Dimension** | **Characteristics** | | | |
|---|---|---|---|---|---|
| **Value Creation** | ML service model | model deployment service | model development & deployment service | hybrid service with ready to use services | |
| | core VP | complexity abstraction | optimization | insight generation | |
| | | customizability | automation | continuous learning | |
| | complementary services | yes | no | | |
| | Model scope | context changing | content changing | | |
| | ML lifecycle activities | data processing | validation | retraining | |
| | | building/training | monitoring | inference | |
| | complementary resources | yes | no | | |
| | assets | tangible | intangible | | |
| | data strategy | proprietary | crowd-sourced | public | |
| | | customer-supplied (on-demand) | customer-supplied (continuous) | | |
| | data type | visual data | mixed data | natural language data | |
| | | numeric/sensor data | textural/ document data | | |
| | integrated complementor | selected complementors | crowd | selected complementors and crowd | none |
| **Value Delivery** | industry focus | industry specific | industry unspecific | | |
| | customer segments | B2B | B2C | other | |
| | interface | graphical | code-based | none | |
| **Value Capture** | pricing model | one-off payment | regularly based on variable rates | regularly based on fixed rates | |
| | | freemium | free trial | | |
| | revenue streams | subscription | transaction | advertising | profit-sharing |
| | costs | tangible | intangible | | |

*Figure 22: Business model taxonomy at the end of second iteration (C2E-2)*

*Note. Dimensions and characteristics with changes since previous iteration are stylized in italic.*

### C1.3 Iteration 3: Conceptual-to-Empirical Iteration (C2E-3)

In the third iteration, we review the dimensions from platform and data-driven business model literature against our meta model. Naous et al.’s (2017) work on Analytics-as-a-Service presented three new dimensions and their characteristics about data-driven business models that we should investigate (see Table 29). The ending conditions were not fully met. We transitioned to empirical-to-conceptual phase to further develop the business model taxonomy using platform cases. The operations that were performed during third iteration of our taxonomy development process are shown in Table 30. The taxonomy in Figure 23 reflects the dimensions and characteristics identified at the end of the third iteration.

*Table 29: Overview of new dimensions identified and their sources (C2E-3)*

| Meta-model | | Dimension (in Source) | Definition (in Source) | Source |
|---|---|---|---|---|
| First order construct | Second order construct | | | |
| Value Creation | Key Partnerships | partnerships | Third party actors who contribute value by offering complementary services to enhance its overall platform value proposition | (Naous et al., 2017) |
| | Key Resources | infrastructure | Specialized infrastructure required to main the platform operations | (Naous et al., 2017) |
| Value Delivery | Customer Relationships | customer relationships | company's different approaches to engage with their customers | (Naous et al., 2017; Freischel et al., 2021) |

*Table 30: Inductive operations overview (C2E-3)*

| Operation type | Description |
|---|---|
| Add | Added “partnerships” dimensions and its characteristics<br>Added “infrastructure” dimensions<br>Added “customer relationships” dimensions and its characteristics |

| | Dimension | Characteristics | | | |
|---|---|---|---|---|---|
| **Value Creation** | ML service model | model deployment service | Model development & deployment service | hybrid service with ready to use services | |
| | core VP | complexity abstraction | optimization | insight generation | |
| | | customizability | automation | continuous learning | |
| | complementary services | yes | no | | |
| | model scope | context changing | content changing | | |
| | ML lifecycle activities | data processing | validation | retraining | |
| | | Building/training | monitoring | inference | |
| | complementary resources | yes | no | | |
| | assets | tangible | intangible | | |
| | data strategy | proprietary | crowd-sourced | public | |
| | | customer-supplied (on-demand) | customer-supplied (continuous) | | |
| | data type | numeric/sensor data | textural/ document data | natural language data | |
| | | visual data | mixed data | | |
| | infrastructure | to be refined | | | |
| **Value Delivery** | integrated complementor | selected complementors | crowd | selected complementors and crowd | none |
| | partnerships | technology partners | OEM partners | independent software vendors | system integrators |
| | Customer relationships | online customer portals | communities | support product evaluation | customer support |
| | industry focus | industry specific | industry unspecific | | |
| | customer Segments | B2B | B2C | other | |
| | interface | graphical | code-based | none | |
| **Value Capture** | pricing model | one-off payment | regularly based on variable rates | regularly based on fixed rates | |
| | | freemium | free trial | | |
| | revenue streams | subscription | transaction | advertising | profit-sharing |
| | costs | tangible | intangible | | |

*Figure 23: Business model taxonomy at the end of second iteration (C2E-3)*

*Note. Dimensions and characteristics with changes since previous iteration are stylized in italic.*

### C1.4 Empirical validation

To validate the preliminary business model taxonomy from conceptual-to-empirical phase, we applied the taxonomy on four sampled cases: Superb AI Suite, Hyper Sense, Snowflake and IBM Watsonx Studio. The gaps identified led to additional operations after each application of the taxonomy of a sampled case (see Table 31). After reviewing all the cases, we further aligned the taxonomy iteratively to adapt to all the cases concurrently (see Table 32). The preliminary business model taxonomy at the end of C2E phase is shown in Figure 24.

*Table 31: Overview of operations in each case validation*

| Operation type | Superb AI Suite | HyperSense | Snowflake | IBM Watsonx Studio |
|---|---|---|---|---|
| Add | Added “*model hosting*” dimension & 2 characteristics<br>Added add “*infrastructure*” dimension and 3 characteristics<br>Added 2 characteristics in “*partnerships*” dimension<br>Added “*education*” in “*revenue streams*” dimension | Added 2 characteristics in “*core VP*” dimension<br>Added 1 characteristic in “*model hosting*” dimension | Added 3 characteristics in “*infrastructure*” dimension<br>Added “partnership programs” to “revenue streams” | None |
| Adapt | None | None | Adapted machine learning service model characteristic to “<br>*model development & deployment with ready to use services*” to specify emphasize the supplementary nature of ready-to-use services | |
| Merge | None | None | | |
| Delete | None | None | None | |

*Table 32: Overview of additional operations using insights from case validation*

| Operation type | Additional refinements |
|---|---|
| Add | None |
| Adapt | Split “complementary services”, and “complementary resources” dimensions. |
| Merge | Merged 8 “*core VP*” characteristics into 4 characteristics<br>Merged “profit sharing” into “partnership program” |
| Delete | Removed “*no*” characteristic from both “*complementary services*” and “*complementary resources*” dimensions<br>Removed 2 characteristics from “*data strategy*” dimension<br>Removed “*integrated complementor*” dimension due to overlap with “*partnerships*”<br>Removed “*customer segment*” as all cases are “*B2B*”, with no clear differentiation<br>Removed 5 more characteristics from lack of evidential data |

| | **Dimension** | **Characteristics** | | | |
|---|---|---|---|---|---|
| **Value Creation** | ML service model | model deployment service | Model development & deployment service | Model development & deployment service with ready to use services | |
| | core VP | accessibility | tailored intelligence | efficiency | adaptability |
| | complementary services | specialized expertise | automation | governance & compliance | |
| | model hosting | cloud | on premise | edge devices | |
| | Model scope | context changing | content changing | | |
| | ML lifecycle activities | data processing | validation | retraining | |
| | | building/training | monitoring | inference | |
| | complementary resources | tools | ecosystem | data assets | prebuilt models |
| | assets | tangible | intangible | | |
| | data strategy | third party | customer-supplied (on-demand) | customer-supplied (continuous) | |
| | data type | numeric/sensor data | textural/ document data | natural language data | |
| | | visual data | mixed data | | |
| | infrastructure | compute | private connectivity | disaster recovery | cloud storage |
| **Value Delivery** | partnerships | technology partners | independent software vendors | system integrators | |
| | | OEM partners | investors | research institution | |
| | Customer relationships | online customer portals | communities | customer support | |
| | industry focus | industry specific | industry unspecific | | |
| | interface | graphical | APIs | SDKs | |
| **Value Capture** | pricing model | usage-based pricing | tiered pricing | free trial | |
| | revenue streams | subscription | advertising | transaction | |
| | | partnership program | education | cloud infrastructure services | |
| | costs | tangible | intangible | | |

*Figure 24: Preliminary taxonomy after C2E phase*

## C2 Empirical-to-Conceptual Iterations (E2C)

### *C2.1 First Empirical-to-Conceptual Iteration (E2C-1)*

The first empirical-to-conceptual (E2C) iteration aimed to establish the foundational dimensions of the MLaaS platform business model taxonomy. The following operations were performed during first empirical-to-conceptual iteration of our taxonomy development process (see Table 33). However, to maintain a practical level of generalizability and inclusiveness across the platform landscape, the taxonomy (see Figure 25) was intentionally kept broad with 16 dimensions and 64 characteristics at this stage. Over-specification based on outlier cases was avoided to preserve its applicability across diverse MLaaS platform configurations. The ending conditions were not fully met, so we proceeded with another empirical-to-conceptual iteration.

*Table 33: Overview of operations in E2C-1 iteration*

| Operation type | Refinement after analysing E2C-1 cases |
|---|---|
| Add | Added “integrated AI assistant” and “marketplace” to “complementary resources” dimension<br>Added “edge devices” to “model hosting” dimension<br>Added “specialized hardware support” to “infrastructure” dimension<br>Added “seat-based pricing” to “pricing model” dimension<br>Added “*data broker*” to “*partnerships*” dimension |
| Adapt | Adapted “online customer portal” to “self-service portal” in “customer relationship” dimension<br>Adapted characteristics of “*Model scope*” to “*predictive*” and “*generative*”.<br>Adapted “partnership program” to “partnership” |
| Merge | Merged “assets” into “complementary resources” dimension<br>Merged “natural language data” into “textual data” in “data type” dimension<br>Merged “OEM partners” into “technology partners” in “partnerships” dimension |
| Delete | Removed “*ML service model*” dimension as all 8 cases provided “*model development and deployment services*” with complementary ready-to-use services. |

|  | **Dimension** | **Characteristics** |  |  |  |
|---|---|---|---|---|---|
| **Value Creation** | core VP | accessibility | tailored intelligence | efficiency | adaptability |
|  | complementary services | specialized expertise | automation | governance & compliance |  |
|  | model hosting | edge devices | public cloud | on premise private cloud | hybrid |
|  | Model scope | predictive | generative |  |  |
|  | ML lifecycle activities | data processing | building/training | validation |  |
|  |  | monitoring | retraining | deploying |  |
|  | complementary resources | tools | data assets | prebuilt models |  |
|  |  | ecosystem | integrated AI assistant | marketplace |  |
|  | data strategy | third party | customer-supplied (on-demand) | customer-supplied (continuous) |  |
|  | data type | numeric/sensor data | textual/natural language data | visual data | mixed data |
|  | Infrastructure | compute | disaster recovery | cloud storage |  |
|  |  | private connectivity | specialized hardware support |  |  |
| **Value Delivery** | partnerships | technology partners | independent software vendors | system integrators |  |
|  |  | data brokers | investors | research institution |  |
|  | customer relationships | self-service portal | communities | customer support |  |
|  | industry focus | industry specific | industry unspecific |  |  |
|  | interface | graphical | APIs | SDKs |  |
| **Value Capture** | pricing model | usage-based pricing | tiered pricing | free trial | seat-based pricing |
|  | revenue streams | subscription | transaction | advertising | cloud infrastructure services |
|  |  | partnership | training | certification |  |
|  | costs | tangible | intangible |  |  |

*Figure 25: Business model taxonomy after E2C-1 iteration*

*Note. Box in grey reflected the presence of changes from the previous version of taxonomy.*

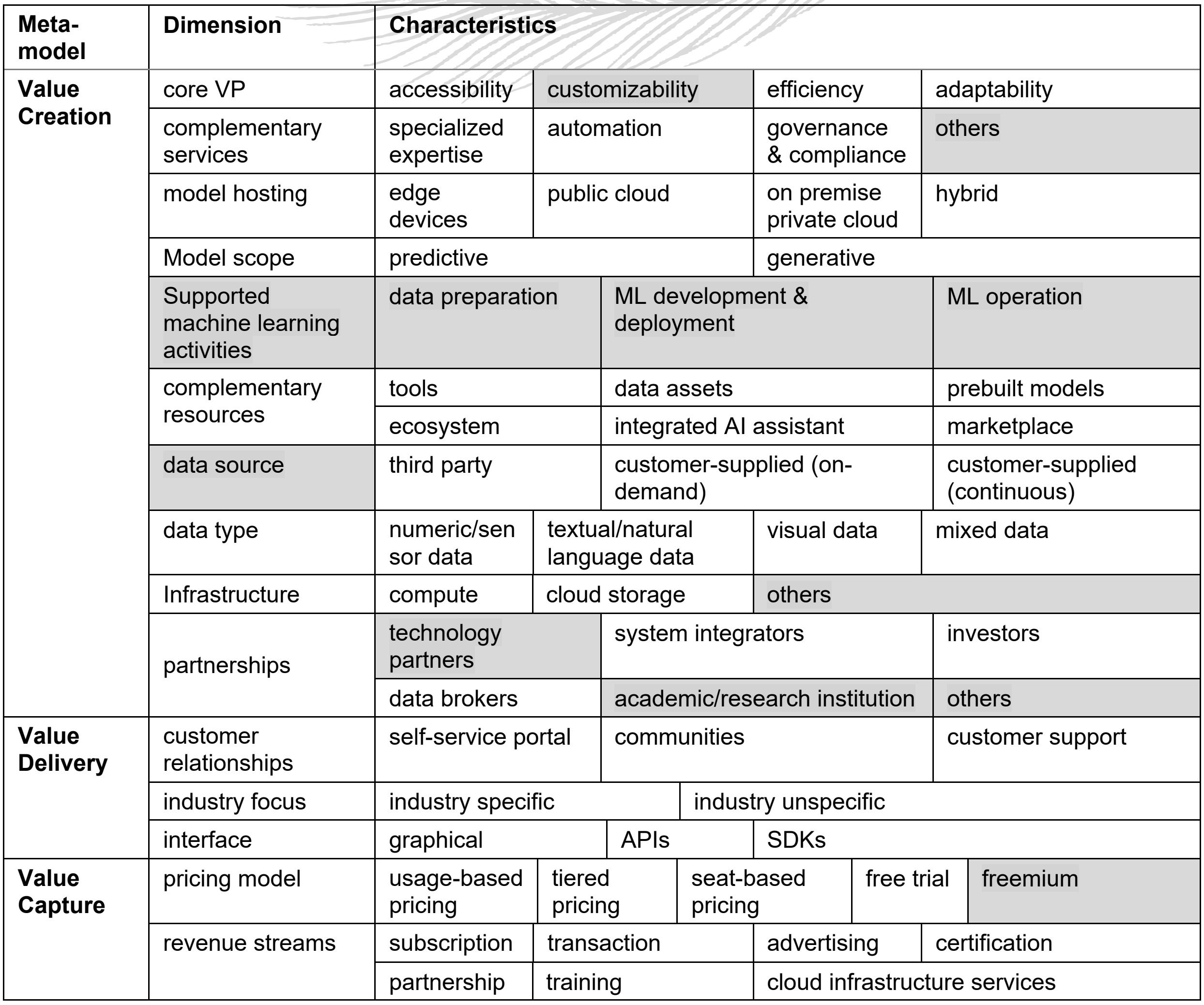
### C2.2 Second Empirical-to-Conceptual Iteration (E2C-2)

The second empirical-to-conceptual (E2C) iteration focused on stress-testing the taxonomy against edge cases and emerging platform archetypes that departed from the dominant models captured in the first round. This iteration analysed platforms such as *Modular* (an AI development platform for LLMs), *CodeOcean* (research reproducibility infrastructure for machine learning models), *Databricks Mosaic AI* (a enterprise-scale AI platform for machine learning and LLM), and *ZAM* (a startup developing agent-like tooling). The following operations were performed during second empirical-to-conceptual iteration of our taxonomy development process (see Table 34).

*Table 34: Overview of operations in E2C-2 iteration*

| Operation type | Refinement after analysing E2C-2 cases |
|---|---|
| Add | Added “*freemium*” to “*pricing model*” dimension<br>Added “others” to “partnerships”, “infrastructure” and “complementary services” dimensions |
| Adapt | Adapted “*ML lifecycle*” as “*Supported machine learning activities*” and grouped post-inference model operational activities together<br>Adapted “tailored intelligence” to “customizability” in “core VP” dimension.<br>Adapted “research institution” to “academic/research institution” in “partnerships” dimension |
| Merge | Merged “independent software vendors” into “technology partners”<br>Merged non-core infrastructure services into “*others*”. Merged ‘*on premise/private cloud*’ into ‘*hybrid*’. |
| Delete | Delete “*costs*” dimension. |

In response, design choices prioritized enhancing conceptual clarity and structural coherence while preserving the taxonomy’s flexibility. For example, the introduction of “*freemium*” in the pricing model dimension reflected new user acquisition strategies evident in developer-focused platforms like *Modular* and *CodeOcean*. “Partnership” dimension was further consolidated to group all technology-related partners together and leaving partners specific for selected few platforms in “others” characteristic. “Infrastructure” dimension large focused on “compute” and “cloud storage” while keeping other add-on cloud infrastructure services together in “others”. Due to lack of empirical evidence support, “costs” dimension was removed. These controlled deletion, additions, and renaming of the dimension and characteristics allowed the taxonomy to absorb increasing diversity and robustness in MLaaS platform strategies without compromising coherence and conciseness. The ending conditions were not fully met. Another iteration was required to further refine the business model taxonomy (see Figure 26).



| Meta-model | Dimension | Characteristics | | | | |
|---|---|---|---|---|---|---|
| **Value Creation** | core VP | accessibility | customizability | efficiency | adaptability | |
| | complementary services | specialized expertise | automation | governance & compliance | others | |
| | model hosting | edge devices | public cloud | on premise private cloud | hybrid | |
| | Model scope | predictive | generative | | | |
| | Supported machine learning activities | data preparation | ML development & deployment | ML operation | | |
| | complementary resources | tools | data assets | prebuilt models | | |
| | | ecosystem | integrated AI assistant | marketplace | | |
| | data source | third party | customer-supplied (on-demand) | customer-supplied (continuous) | | |
| | data type | numeric/sensor data | textual/natural language data | visual data | mixed data | |
| | Infrastructure | compute | cloud storage | others | | |
| | partnerships | technology partners | system integrators | investors | | |
| | | data brokers | academic/research institution | others | | |
| **Value Delivery** | customer relationships | self-service portal | communities | customer support | | |
| | industry focus | industry specific | industry unspecific | | | |
| | interface | graphical | APIs | SDKs | | |
| **Value Capture** | pricing model | usage-based pricing | tiered pricing | seat-based pricing | free trial | freemium |
| | revenue streams | subscription | transaction | advertising | certification | |
| | | partnership | training | cloud infrastructure services | | |

*Figure 26: Business model taxonomy after E2C-2 iteration*

### C2.3 Third Empirical-to-Conceptual Iteration (E2C-3)

The second empirical-to-conceptual (E2C) iteration focused on improving the robustness of the taxonomy and mutual exclusivity of the taxonomy characteristics. This iteration examined four strategically selected platforms with hybrid configurations and industry-specific adaptations: Dataloop (MLaaS platform with data-centric ecosystem), Iguazio (enterprise MLaaS platform), Google Vertex AI (integrated enterprise cloud-originated platform), and PrimeHub (regional predictive AI platform serving multiple industries). We focused deeper on the primary characteristics of value proposition, infrastructure, data source, partnerships from the previous iteration, generating five new dimensions. The following operations were performed during third empirical-to-conceptual iteration of our taxonomy development process (see Table 35). When the ending conditions of the business model taxonomy were fully met, we finalized our taxonomy (see Figure 27) and proceeded with taxonomy evaluation phase.

*Table 35: Overview of operations in E2C-3 iteration*

| Operation type | Refinement after analysing E2C-2 cases |
|---|---|
| Add | Added five new dimensions and 13 characteristics |
| Adapt | Adapted '*partnerships*' dimension into '*key partners*'; '*model hosting*' into '*model deployment*'; '*ML lifecycle activities*' dimension into '*supported activities*' dimension and only two characteristics. Adapted '*complementary services*' characteristics to focus on their purpose. |
| Merge | Merged three characteristics of '*interface*' dimension into two; and seven characteristics from '*revenue streams*' dimension into just three. |
| Delete | Removed '*infrastructure*', '*value proposition*', '*data source*' dimensions. Removed '*hybrid*' characteristic from '*model deployment*' dimension. Removed 2 characteristics from '*Key Partners*' dimension. |

|  | **Dimension** | **Characteristics** |  |  |  |
|---|---|---|---|---|---|
| **Value Creation** | deployment flexibility | self-managed | pre-configured |  |  |
|  | user control | Configurable modular components | Full control on machine learning workflow | Full control on machine learning workflow with AutoML option |  |
|  | complementary services | yes | no |  |  |
|  | model deployment | edge devices | multi-tenant cloud | hybrid |  |
|  | model scope | predictive | generative |  |  |
|  | supported activities | development focus | development and operational focus |  |  |
|  | integrated resources | tools | data assets | prebuilt models | AI assistant |
|  | external resources | marketplace | partner ecosystem | both |  |
|  | data source | third party | customer | both |  |
|  | data flow | on-demand | continuous |  |  |
|  | data type | numeric/sensor data | textual/natural language data | visual data | mixed data |
| **Value Delivery** | key partners | technology partners | system integrators | data brokers |  |
|  |  | academic/research institution | others |  |  |
|  | customer relationships | self-service portal | communities | customer support |  |
|  | industry focus | industry specific | industry unspecific |  |  |
|  | interface | graphical | APIs/SDKs | both |  |
| **Value Capture** | pricing model | consumption-based | seat-based pricing | tiered pricing | free trial |
|  | revenue streams | direct sales | indirect sales | ancillary services |  |

*Figure 27: Business model taxonomy after E2C-3 iteration*

### C3 Taxonomy Evaluation

The last iteration served as a rigorous taxonomy evaluation phase, focusing on assessing the structural and usability properties of the latest taxonomy. This iteration examined four strategically selected platforms with hybrid configurations and industry-specific adaptations: BigML (developer-centric platform), Microsoft Azure AI (enterprise cloud-originated platform), DataRobot (integrated enterprise data platform), and VessL AI (domain-focused predictive AI platform). The within-case analysis confirmed that all existing dimensions and characteristics from the prior iteration sufficiently captured the diversity and complexity of the cases studied. Adjustments were made to "*deployment flexibility*", "*model scope*", and "*data flow*" dimensions by adding "both" to improve the business model taxonomy's ability to fit all sampled cases. Structural validity and practical usability were affirmed in the refined taxonomy. This iteration validated the taxonomy's comprehensive applicability across diverse MLaaS platform configurations and confirmed its alignment with Nickerson et al.'s (2013) objective and subjective ending conditions. Consequently, the refined taxonomy was deemed stable, sufficiently robust, and ready for taxonomy analysis.

## D. Taxonomy Ending Conditions Overview

*Table 36: Overview of ending conditions*

| | **Ending Conditions** | **Design Iteration** | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | C2E-1 | C2E-2 | C2E-3 | E2C-1 | E2C-2 | E2C-3 | E2C-4 |
| **Objective** | All objects, or at least a representative sample of objects have been examined | | | | | | x | x |
| | In the last design iteration, no object was merged with a similar object or split into multiple objects | | | | | | x | x |
| | For every characteristic of every dimension in the taxonomy at least one object is classified | | | | | x | x | x |
| | In the last design iteration, no new dimensions or characteristics were added to the taxonomy | | | | | | x | x |
| | In the last design iteration, no dimensions or characteristics were merged or split | | | | | | x | x |
| | There is no duplication of dimension in the taxonomy; no dimension is repeated, and every dimension is unique | | x | x | x | x | x | x |
| | There is no duplication of characteristics within the dimensions; every characteristic is unique within its dimension | | | x | x | x | x | x |
| | There is no combination of characteristics and there is no cell duplication in the taxonomy; each cell is unique and there is no repetition of cells | | | | | x | x | x |
| **Subjective** | The taxonomy is concise; the number of dimensions make the taxonomy to be meaningful, but not overwhelming | | | | | | | x |
| | The taxonomy is robust; the dimensions and characteristics of the taxonomy enable researchers to sufficiently differentiate the objects of interest | | | | | | | x |
| | The taxonomy is comprehensive; the taxonomy can be used to classify a sample or all objects within the research domain | | | | | | | x |
| | The taxonomy is extendible; new dimensions or characteristics of an existing dimension can easily be added | x | x | x | x | x | x | x |
| | The taxonomy is explanatory; the identified dimensions and characteristics can be utilized to explain about an object | x | x | x | x | x | x | x |

# E. Taxonomy-Case Mapping for Taxonomy Analysis

## E1 Value Creation

*Table 37: Taxonomy case mapping for value creation mechanisms (part 1)*

| **Case (ID)** | **value creation** | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Dimension** | **deployment flexibility** | | | **user control** | | **model scope** | | | **model deployment** | | **complementary services** | | |
| **Characteristic** | **self-managed** | **pre-configured** | **both** | **full developer control** | **full developer control with optional workflow automation** | **predictive** | **generative** | **both** | **public and private cloud** | **edge, public, private cloud** | **value proposition enhancing** | **use case extending** | **both** |
| Barbara (1) | yes | no | no | yes | no | yes | no | no | no | yes | yes | no | no |
| Superb AI (2) | yes | no | no | no | yes | yes | no | no | no | yes | yes | no | no |
| Modular (3) | yes | no | no | yes | no | no | yes | no | no | yes | yes | no | no |
| Dataloop (4) | no | no | yes | no | yes | no | no | yes | yes | no | yes | no | no |
| BigML (5) | no | no | yes | no | yes | yes | no | no | yes | no | no | no | yes |
| Alibaba Machine Learning Platform for AI (6) | no | no | yes | no | yes | no | no | yes | no | yes | yes | no | no |
| PrimeHub (7) | no | no | yes | yes | no | yes | no | no | yes | no | yes | no | no |
| HyperSense (8) | no | yes | no | no | yes | yes | no | no | yes | no | yes | no | no |
| CodeOcean (9) | yes | no | no | yes | no | yes | no | no | yes | no | no | yes | no |
| Amazon SageMaker (10) | no | yes | no | no | yes | no | no | yes | yes | yes | no | no | yes |
| Microsft Azure AI  (11) | no | yes | no | no | yes | no | no | yes | no | yes | no | no | yes |
| Jfrog ML (12) | no | no | yes | no | yes | no | no | yes | no | yes | no | no | yes |
| Snowflake (13) | no | yes | no | no | yes | no | no | yes | yes | no | no | no | yes |
| Domino Data Lab (14) | no | no | yes | yes | no | no | no | yes | no | yes | no | no | yes |
| H2O.ai (15) | no | no | yes | no | yes | no | no | yes | yes | no | yes | no | no |
| Iguazio (16) | no | no | yes | yes | no | no | no | yes | yes | no | yes | no | no |
| Databricks Mosaic AI (17) | no | no | yes | no | yes | no | no | yes | yes | no | no | no | yes |
| Dataiku (18) | no | yes | no | no | yes | no | no | yes | yes | no | yes | no | no |
| Google Vertex AI (19) | no | yes | no | no | yes | no | no | yes | yes | yes | no | no | yes |
| IBM Watsonxx (20) | no | yes | no | no | yes | no | no | yes | no | yes | yes | no | no |
| Cloudera AI (21) | no | no | yes | no | yes | no | no | yes | yes | no | no | yes | no |
| DataRobot (22) | no | no | yes | no | yes | no | no | yes | yes | no | yes | no | no |
| VESSL (23) | no | yes | no | yes | no | no | yes | no | yes | no | no | yes | no |
| ZAMS (24) | yes | no | no | no | yes | no | no | yes | yes | no | yes | no | no |

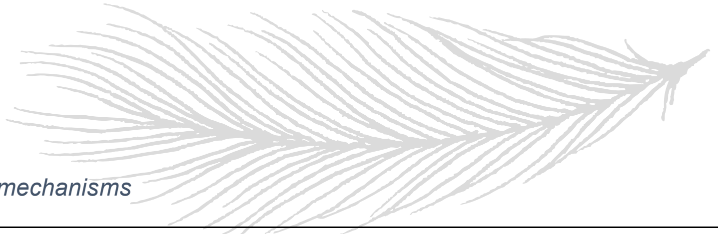
*Table 38: Taxonomy case mapping for value creation mechanisms (part 2)*

| **Case (ID)** | **value creation** | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Dimension** | **supported activities** | | **integrated resources** | | | | | **data source** | | **data flow** | | | **data type** | | |
| **Characteristic** | **development focus** | **development and operational focus** | **tools** | **prebuilt models** | **data assets** | **cloud infrastructure** | **AI assistant** | **customer** | **customer and third party** | **on-demand** | **continuous** | **both** | **telemetric or time-series data** | **visual data** | **multimodal data** |
| Barbara (1) | no | yes | yes | no | no | no | no | yes | no | no | yes | no | yes | no | no |
| Superb AI (2) | no | yes | yes | no | no | no | no | yes | no | no | yes | no | no | yes | no |
| Modular (3) | yes | no | yes | yes | no | no | yes | yes | no | no | no | yes | no | no | yes |
| Dataloop (4) | no | yes | yes | yes | yes | no | no | no | yes | no | yes | no | no | no | yes |
| BigML (5) | yes | no | yes | yes | yes | no | no | no | yes | no | no | yes | no | no | yes |
| Alibaba Machine Learning Platform for AI (6) | no | yes | yes | yes | no | yes | no | no | yes | no | no | yes | no | no | yes |
| PrimeHub (7) | no | yes | yes | yes | no | no | no | yes | no | yes | no | no | no | no | yes |
| HyperSense (8) | no | yes | yes | yes | no | no | no | yes | no | no | no | yes | no | no | yes |
| CodeOcean (9) | no | yes | yes | yes | yes | no | yes | no | yes | yes | no | no | no | no | yes |
| Amazon SageMaker (10) | no | yes | yes | yes | yes | yes | yes | no | yes | no | no | yes | no | no | yes |
| Microsft Azure AI (11) | no | yes | yes | yes | no | yes | yes | no | yes | no | no | yes | no | no | yes |
| Jfrog ML (12) | no | yes | yes | yes | no | no | no | no | yes | no | yes | no | no | no | yes |
| Snowflake (13) | no | yes | yes | yes | yes | no | no | no | yes | no | no | yes | no | no | yes |
| Domino Data Lab (14) | no | yes | yes | yes | yes | no | yes | no | yes | no | no | yes | no | no | yes |
| H2O.ai (15) | no | yes | yes | yes | no | no | yes | yes | no | no | yes | no | no | no | yes |
| Iguazio (16) | no | yes | yes | yes | no | no | no | yes | no | no | no | yes | no | no | yes |
| Databricks Mosaic AI (17) | no | yes | yes | yes | yes | no | no | no | yes | no | no | yes | no | no | yes |
| Dataiku (18) | no | yes | yes | yes | yes | no | yes | no | yes | no | no | yes | no | no | yes |
| Google Vertex AI (19) | no | yes | yes | yes | no | yes | yes | yes | no | no | no | yes | no | no | yes |
| IBM Watsonx (20) | no | yes | yes | yes | no | yes | yes | no | yes | no | no | yes | no | no | yes |
| Cloudera AI (21) | no | yes | yes | yes | no | no | yes | yes | no | no | yes | no | no | no | yes |
| DataRobot (22) | no | yes | yes | yes | no | no | no | no | yes | no | no | yes | no | no | yes |
| VESSL (23) | no | yes | yes | yes | no | no | no | no | yes | yes | no | no | no | no | yes |
| ZAMS (24) | no | yes | yes | yes | no | no | yes | yes | no | no | yes | no | no | no | yes |

## E2 Value Delivery

*Table 39: Taxonomy case mapping for value delivery mechanisms*

| **Case (ID)** | **value delivery** | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Dimension** | **key partners** | | | | **customer relationships** | | | **Industry focus** | | **sales channels** | |
| **Characteristic** | **technology partners** | **system integrators** | **data brokers** | **academic or research institutions** | **self-service portal** | **communities** | **customer support** | **specific** | **broad** | **platform interface and partners** | **platform interface, partners and marketplace** |
| Barbara (1) | yes | yes | no | yes | yes | no | no | yes | no | no | yes |
| Superb AI (2) | yes | no | no | no | no | no | yes | yes | no | yes | no |
| Modular (3) | yes | no | no | no | yes | yes | no | no | yes | yes | no |
| Dataloop (4) | yes | yes | yes | no | yes | yes | no | no | yes | no | yes |
| BigML (5) | yes | no | no | no | yes | no | yes | no | yes | yes | no |
| Alibaba Machine Learning Platform for AI (6) | yes | yes | no | no | yes | yes | yes | no | yes | no | yes |
| PrimeHub (7) | yes | yes | no | no | no | yes | no | yes | no | yes | no |
| HyperSense (8) | yes | no | no | no | yes | no | no | yes | no | yes | no |
| CodeOcean (9) | yes | no | no | yes | yes | yes | yes | yes | no | yes | no |
| Amazon SageMaker (10) | yes | yes | yes | no | yes | yes | yes | no | yes | no | yes |
| Microsft Azure AI (11) | yes | yes | no | no | yes | yes | yes | no | yes | no | yes |
| Jfrog ML (12) | yes | yes | no | no | yes | yes | no | no | yes | yes | no |
| Snowflake (13) | yes | yes | yes | no | yes | yes | yes | no | yes | no | yes |
| Domino Data Lab (14) | yes | yes | yes | no | yes | yes | yes | no | yes | no | yes |
| H2O.ai (15) | yes | yes | no | no | yes | yes | yes | no | yes | yes | no |
| Iguazio (16) | yes | yes | no | no | yes | no | no | no | yes | yes | no |
| Databricks Mosaic AI (17) | yes | yes | yes | yes | yes | yes | yes | no | yes | no | yes |
| Dataiku (18) | yes | yes | no | no | yes | yes | yes | no | yes | yes | no |
| Google Vertex AI (19) | yes | yes | yes | no | yes | yes | no | no | yes | no | yes |
| IBM Watsonx (20) | yes | yes | yes | yes | yes | yes | yes | no | yes | no | yes |
| Cloudera AI (21) | yes | yes | no | no | yes | yes | yes | no | yes | yes | no |
| DataRobot (22) | yes | yes | no | no | yes | no | no | no | yes | no | yes |
| VESSL (23) | yes | no | no | yes | no | yes | no | no | yes | no | yes |
| ZAMS (24) | yes | yes | no | no | yes | no | yes | yes | no | yes | no |



## E3 Value Capture

*Table 40: Taxonomy case mapping for value capture mechanisms*

| **Case (ID)** | **value capture** | | | | | | |
|---|---|---|---|---|---|---|---|
| **Dimension** | **pricing model** | | | | **revenue streams** | | |
| **Characteristic** | **consumption-based** | **tiered** | **seat-based** | **freemium** | **direct sales** | **indirect sales** | **ancillary services** |
| Barbara (1) | no | yes | no | no | yes | yes | yes |
| Superb AI (2) | n/a | n/a | n/a | yes | yes | no | yes |
| Modular (3) | yes | yes | no | yes | yes | no | no |
| Dataloop (4) | no | yes | no | no | yes | yes | yes |
| BigML (5) | yes | yes | no | yes | yes | no | yes |
| Alibaba ML Platform for AI (6) | yes | yes | no | no | yes | no | yes |
| PrimeHub (7) | n/a | n/a | n/a | n/a | yes | yes | yes |
| HyperSense (8) | n/a | n/a | n/a | n/a | yes | no | yes |
| CodeOcean (9) | yes | no | no | yes | yes | no | yes |
| Amazon SageMaker (10) | yes | yes | no | yes | yes | yes | yes |
| Microsft Azure AI  (11) | yes | no | no | yes | yes | yes | yes |
| Jfrog ML (12) | no | yes | no | no | yes | no | yes |
| Snowflake (13) | yes | yes | no | yes | yes | yes | yes |
| Domino Data Lab (14) | no | yes | yes | no | yes | yes | yes |
| H2O.ai (15) | yes | yes | no | yes | yes | no | yes |
| Iguazio (16) | yes | yes | no | yes | yes | yes | yes |
| Databricks Mosaic AI (17) | yes | yes | no | yes | yes | yes | yes |
| Dataiku (18) | no | yes | no | yes | yes | yes | yes |
| Google Vertex AI (19) | yes | no | yes | yes | yes | yes | yes |
| IBM Watsonx (20) | yes | no | no | yes | yes | yes | yes |
| Cloudera AI (21) | yes | no | no | yes | yes | no | yes |
| DataRobot (22) | yes | yes | no | yes | yes | yes | yes |
| VESSL (23) | yes | yes | no | yes | yes | no | no |
| ZAMS (24) | no | yes | no | yes | yes | no | yes |

*Table 41: Discriminatory dimensions of business model archetypes and their characteristics*

| Archetypes | Cloud orchestrator (n=5) | Data orchestrator (n=4) | Aggregator (n=8) | Niche specialist (n=7) |
|---|---|---|---|---|
| *Dimension* | *Characteristics* | | | |
| deployment flexibility | Pre-configured (100%) | Both (100%) | Both (100%) | No dominant characteristic |
| complementary services | Both (100%) | Both (100%) | Value proposition enhancing (63%) | Value proposition enhancing (71%) |
| model scope | Both (100%) | Both (100%) | Both (86%) | Predictive (71%) |
| Model deployment | Edge, public and private cloud (100%) | Public and private cloud (75%) | Edge, public and private cloud (63%) | Edge, public and private cloud (63%) |
| Integrated resources | Tools (100%),<br>prebuilt models (100%),<br>cloud infrastructure (100%),<br>AI assistant (80%) | Tools (100%),<br>prebuilt models (100%),<br>data assets (100%),<br>AI assistant (75%) | Tools (100%),<br>prebuilt models (100%) | Tools (100%),<br>prebuilt models (71%) |
| data sources | Customer and third party (80%) | Customer and third party (100%) | Customer and third party (100%) | Customer (71%) |
| key partners | Technology partners (100%),<br>system integrators (100%) | Technology partners (100%),<br>system integrators (100%),<br>data brokers (75%) | Technology partners (100%),<br>system integrators (100%) | Technology partners (100%) |
| industry focus | Generic (100%) | Generic (100%) | Generic (100%) | Specific (100%) |
| sales channel | Platform interface, partners and marketplace (100%) | Platform interface, partners and marketplace (100%) | Platform interface and partners (100%) | Platform interface and partners (100%) |
| pricing model | Consumption-based (100%)<br>Freemium (100%) | Consumption-based (100%)<br>Tiered (100%) | Tiered (88%)<br>Freemium (100%) | Freemium (71%) |
| revenue stream | Direct sales (100%)<br>Indirect sales (80%)<br>Ancillary services (100%) | Direct sales (100%)<br>Indirect sales (100%)<br>Ancillary services (100%) | Direct sales (100%)<br>Indirect sales (63%)<br>Ancillary services (88%) | Direct sales (100%)<br>Ancillary services (86%) |

*Note. n = the number of cases. % = percentage of cases that demonstrates the characteristic. Characteristics demonstrated by more than 50% of cases are considered dominant.*

## F. Comparison between Business Model Archetypes

*Table 42: Comparison between the business model archetypes (part 1)*

| | Dimension | Characteristic | Cloud orchestrator | Data orchestrator | Aggregator | Niche specialist |
|---|---|---|---|---|---|---|
| value creation | deployment flexibility | self-managed | None | None | None | 1, 24 |
| | | pre-configured | 6, 10, 11, 19, 20 | None | None | 9, 23 |
| | | both | | 4, 13, 14, 17 | 3, 5, 12, 16, 15, 18, 21, 22 | 2, 7 |
| | user control | full developer control | 6, 19 | 13, 17 | 3, 12 | 1, 7, 9, 23 |
| | | full developer control with optional workflow automation | 10, 11, 20 | 4, 14 | 5, 16, 15, 18, 21, 22 | 2,8,24 |
| | model scope | predictive | None | None | 3 | 1, 2, 7, 8, 9 |
| | | generative | None | None | 5 | 23 |
| | | both | 6, 10, 11, 19, 20 | 4, 13, 14, 17 | 12, 16, 15, 18, 21, 22 | 24 |
| | model deployment | public and private cloud | None | 4, 13, 17 | 5, 12, 18 | 1, 2 |
| | | edge, public, private cloud | 6, 10, 11, 19, 20 | 14 | 3, 16, 15, 21, 22 | 7, 8, 9, 23, 24 |
| | complementary services | value proposition enhancing | 6 | None | 3, 16, 15, 18, 22 | 1, 2, 7, 8,  24 |
| | | use case entending | None | None | 21 | 9, 23 |
| | | both | 6, 10, 11, 19, 20 | 4, 13, 14, 17 | 5, 12 | |
| | supported activities | development focus | None | None | None | 2, 8, 24 |
| | | development and operational focus | 6, 10, 11, 19, 20 | 4, 13, 14, 17 | 3, 5, 12, 16, 15, 18, 21, 22 | 1, 7, 9, 23 |
| | integrated resources | tools | 6, 10, 11, 19, 20 | 4, 13, 14, 17 | 3, 5, 12, 16, 15, 18, 21, 22 | 1, 2, 7, 8, 9, 23, 24 |
| | | prebuilt models | 6, 10, 11, 19, 20 | 4, 13, 14, 17 | 3, 5, 12, 16, 15, 18, 21, 22 | 7, 8, 9, 23, 24 |
| | | data assets | 10 | 4, 13, 14, 17 | None | 9 |
| | | cloud infrastructure | 6, 10, 11, 19, 20 | | None | |
| | | AI assistant | 10, 11, 19, 20 | 13, 14, 17 | 16, 18, 22 | 9, 24 |
| | data source | customer | 19 | None | None | 1, 2, 7, 8, 24 |
| | | customer and third party | 6, 10, 11, 20 | 4, 13, 14, 17 | 3, 5, 12, 16, 15, 18, 21, 22 | 9, 23 |
| | data flow | on-demand | None | None | None | 1, 2, 24 |
| | | continuous | None | None | None | 7, 9, 23 |
| | | both | 6, 10, 11, 19, 20 | 4, 13, 14, 17 | 3, 5, 12, 16, 15, 18, 21, 22 | |
| | data type | numeric/sensor data | None | None | None | 1 |
| | | visual data | None | None | None | 2 |
| | | multimodal data | 6, 10, 11, 19, 20 | 4, 13, 14, 17 | 3, 5, 12, 16, 15, 18, 21, 22 | 7, 8, 9, 23, 24 |

*Note. Dimensions and characteristics with high discriminatory power are highlighted in grey. These elements could differentiate archetypes exclusively.*

*Table 43: Comparison between the business model archetypes (part 2)*

| | Dimension | Characteristic | Cloud orchestrator | Data orchestrator | Aggregator | Niche specialist |
|---|---|---|---|---|---|---|
| value delivery | key partners | technology partners | 6, 10, 11, 19, 20 | 4, 13, 14, 17 | 3, 5, 12, 16, 15, 18, 21, 22 | 1, 2, 7, 8, 9, 23, 24 |
| | | system integrators | 6, 10, 11, 19, 20 | 4, 13, 14, 17 | 12, 16, 15, 18, 21, 22 | 1, 7 |
| | | data brokers | 10, 19, 20 | 4, 13, 14, 17 | None | None |
| | | academic/research institutions | 20 | 17 | None | 1, 23, 24 |
| | customer relationships | self-service portal | 6, 10, 11, 19, 20 | 4, 13, 14, 17 | 3, 5, 12, 16, 15, 18, 21, 22 | 1, 2, 8, 9, 24 |
| | | communities | 6, 10, 11, 19, 20 | 4, 13, 14, 17 | 3, 5, 12, 16, 15, 18, 21, 22 | 7, 9, 23 |
| | | customer support | 6, 10, 11, 19, 20 | 4, 13, 14, 17 | 3, 5, 12, 16, 15, 18, 21, 22 | 1, 2, 7, 8, 9, 23, 24 |
| | industry focus | specific | | | | 1, 2, 7, 8, 9, 23, 24 |
| | | generic | 6, 10, 11, 19, 20 | 4, 13, 14, 17 | 3, 5, 12, 16, 15, 18, 21, 22 | |
| | sales channels | platform interface and partners | | | 3, 5, 12, 16, 15, 18, 21, 22 | 2, 7, 8, 9, 23, 24 |
| | | platform interface, partners and marketplace | 6, 10, 11, 19, 20 | 4, 13, 14, 17 | | 1 |
| value capture | pricing model | consumption-based | 6, 10, 11, 19, 20 | 4, 13, 14, 17 | 3, 16, 22 | 1, 23 |
| | | tiered | 6, 10 | 4, 13, 14, 17 | 5, 12, 16, 15, 18, 21, 22 | 1, 2, 23, 24 |
| | | seat-based | 19 | 14 | 5 | 24 |
| | | freemium | 6, 10, 11, 19, 20 | 3, 17 | 3, 5, 12, 16, 15, 18, 21, 22 | 1, 2, 8, 9, 24 |
| | revenue streams | direct sales | 6, 10, 11, 19, 20 | 4, 13, 14, 17 | 3, 5, 12, 16, 15, 18, 21, 22 | 1, 2, 7, 8, 9, 23, 24 |
| | | indirect sales | 10, 11, 19, 20 | 4, 13, 14, 17 | 16, 15, 18, 21, 22 | 1, 7 |
| | | ancillary services | 6, 10, 11, 19, 20 | 4, 13, 14, 17 | 5, 12, 16, 15, 18, 21, 22 | 1, 2, 7, 8, 9, 24 |

*Note. Dimensions and characteristics with high discriminatory power are highlighted in grey. These elements could differentiate archetypes exclusively.*