DEPARTMENT OF
INFORMATION SYSTEMS
AT FREIE UNIVERSITÄT BERLIN

MASTER THESIS

Analyzing the Scaling Intensity of AI-enabled Platform and Service Startups

Anna Schapiro

Supervisor	Univ.-Prof. Dr. Daniel Fürstenau
Second Supervisor	Univ.-Prof. Dr. Natalia Kliewer
Semester	Winter Semester 2024/2025
Author	Anna Schapiro
Student ID	5039650
Email	schapiro.anna@gmail.com
Phone	+49 152 5706 6134
Program	Master in Information Systems
Submission Date	November 19, 2024

Contents

List of Acronyms

AI Artificial Intelligence

GDP Gross Domestic Product

AI startup Artificial Intelligence Startup

IT Information Technology

AIaaS AI as a Service

DDDM Data Driven Decision Making

HR Human Resources

CSV Comma Separated Values

JSON JavaScript Object Notation

API Application Programming Interface

B2B Business to Business

B2C Business to Consumer

VC Venture Capital

ID Identifier

NaN Not a Number

SVR Support Vector Regressor

RBF Radial Bias Function

MLP Regressor Multi-Layer Perceptron Regressor

ReLU Rectified Linear Unit

adam Adaptive Moments

List of Figures

List of Tables

Abstract

This study explores the scaling potential of artificial intelligence (AI) in platform and service startups, focusing on the economic implications of AI-driven growth models. By analyzing approximately 17,000 startups from the PitchBook database and applying regression models inspired by West (2019) and Schulte-Althoff, Fürstenau, and Lee (2021) scaling approach, it was found that AI-powered service startups have significantly higher scaling rates than their non-AI service counterparts, with scaling exponents on average 0.6 to 2.53 times higher. Specifically, the scaling exponents for gross profit and venture capital increased from 0.7 for non-AI service startups to 1.0 for AI-enabled service startups, and revenue scaling exponents rose from 1.1 for non-AI service startups to 2.9 for AI service startups.

The study also examines various locations and industries of startups, revealing that AI plays a crucial role in reshaping traditional growth patterns across different sectors and regions.

The results indicate that artificial intelligence significantly enhances the scaling capabilities of service startups, enabling them to achieve faster and more effective growth. However, this advantage does not extend to platform startups, where AI does not appear to have the same impact. By highlighting the simplicity of scaling formulas, the study illustrates their practical usage in evaluating the innovation potential and growth prospects of AI-based business models. While the findings are more descriptive than predictive, they still offer valuable insights for researchers, investors, and founders seeking to leverage AI as a strategic tool to gain a competitive edge in an increasingly digital economy.

Keywords: AI in business models, AI scalability, Scaling exponents, AI-enabled service startups, Economic implications of AI, West scaling and Schulte-Althoff scaling

1 Introduction

The global market capitalization for Artificial Intelligence (AI) reached a value of 184 billion US dollars in 2024 (Statista 2024). According to the Statista Research Department, the value will grow by over 70% to a remarkable 827 billion US dollars by 2030. AI has the potential to contribute significantly to the Gross Domestic Product (GDP) not only for single countries but for whole geographical regions, where it is predicted to add billions of dollars by 2030 (Ahmed 2019; Statista 2024).

This rapid development requires an analysis of the extent to which well-known concepts from the information economy can be transferred to AI ventures. Since the field of AI is relatively new, and the technology is evolving rapidly, startups being early technological drivers are particularly interesting.

Various studies conclude that AI startups have different scaling requirements and opportunities compared to conventional digital startups (Schulte-Althoff et al. 2021; Weber, Beutter, Weking, Boehm, & Krcmar 2021). For example, AI startups face unique challenges during scaling in areas such as data management, human resources, and infrastructure (Colombelli, D'Amico, & Paolucci 2023; Jia & Stan 2021; Jöhnk, Weißert, & Wyrtki 2021; Schlegel, Schuler, & Westenberger 2023).

Artificial Intelligence is revolutionizing the way businesses operate (Paluch & Wirtz 2020). AI-enabled platforms and AI-enabled service startups are emerging at an unprecedented rate, leveraging AI's capabilities to offer innovative solutions. Analyzing the scaling of these startups is relevant for understanding how they grow in a competitive market. This master thesis can provide quantitative insights into the success of startups related to using AI in their business models.

Existing research on AI business models is still in its infancy, and there is a pressing need to understand how AI differentiates itself from traditional IT in terms of business scaling.

For example, research shows that AI startups need to orchestrate internal processes differently than traditional IT startups in terms of specific resources like data, financing, personnel, strategy, technology, organizational infrastructure, and compliance requirements. The ongoing research is mainly based on successful case studies. Some articles use expert interviews and identify pitfalls or specific scaling-related aspects with sample sizes of 100 or 112 expert interviews (Ermakova et al. 2021; Jia & Stan 2021; Schapiro, Keutner, Friedrich, & Sanwald 2024; Weber et al. 2021)

There is only one large quantitative study that analyses and compares scaling in various AI startups with non-AI startup businesses. The authors have found that there is a difference in scaling among three types of business models: service, platform, and AI startups (Schulte-Althoff et al. 2021).

Based on these findings, this master thesis also aims to investigate quantitatively, whether AI-enabled service and AI-enabled platform startups are scaling differently compared to counterpart startups, which do not use AI as a central part of their business models.

Therefore, I focus on the research questions:

- **Question 1: Does AI in the core of business models influence the scaling efficiency of service startups?**
- **Question 2: Does AI in the core of business models influence the scaling efficiency of platform startups?**

This data-driven work aims to provide answers beyond case studies and questionnaires. Answers to these questions might be relevant for researchers, investors, founders, and other market participants.

This master thesis explores startups, defined as companies younger than ten years. The data from the PitchBook economic database (pitchbook 2022), is analyzed using various quantitative regression methods. The used variables include *net income, gross profit, Venture Capital (VC), revenue*, and the number of *employees*. The scaling models are inspired by West (2019) and Schulte-Althoff et al. (2021). Further, this work is also focusing on revenue-to-employee scaling.

Based on this approach, the following conclusions are drawn:

- **Answer to question 1:** Yes, when comparing companies with more than 80 employees (for those without AI business models) and more than 100 employees (for those with AI business models), service AI startups have a scaling exponent that is *0.60 to 2.53 times higher* than that of non-AI service startups.

  The scaling analyses with revenue, gross profit, and VC raised as variables support this finding: AI in service startups brings higher scalability or exponential growth. However, these findings do not have a robust predictive power but are more descriptive.

- **Answer to question 2:** No. With my methods, no answer could be given below the marks of 15 (for those without AI business models) or 70 employees (for those with AI business models). Above those marks, with the removal of the top 1% revenue achievers, a "No" would be the answer. The scaling analysis with other variables like net income, gross profit, and venture capital also supports this finding by showing that these two categories of startups do not follow the ”universal” scaling pattern. However, these findings do not have a strong predictive power; rather, they are more descriptive.

These findings indicate that AI substantially impacts scalability, particularly within service startups. This seems to be different for platform startups.

This work is structured in the following way: Chapter 2 covers the fundamental concepts essential to understanding the topic, including definitions of AI, AI-enabled startups, and scaling. It further explores different startup types and the specifics of AI-enabled service and platform startups. The chapter concludes with a discussion on scaling mechanisms in business models and their universality. Chapter 3 describes the data collection and preparation processes. It also includes visualizations of the dataset. Chapter 4 consists of an explanation of the functional form used for scaling analysis (power function) and a discussion on the robustness of the models through k-fold cross-validation. Chapter 5 presents the core analysis, beginning with West’s and Schulte-Althoff’s scaling analyses. The chapter includes sub-analyses on geographical and industry-specific data to explore the broader context of scaling in AI-enabled startups. It also answers the research questions and interprets the findings, concluding the implications of AI-enabled scaling in service startups. Chapter 6 summarizes the results, explores their practical implications, and discusses limitations. It concludes with an outlook on future research directions. Chapter 7 wraps up the work with a concise summary of the key findings, implications, and recommendations for future research. Chapter 8 includes supplementary materials, such as a bibliometric analysis, further dataset-related visualizations, tables for the sub-analyses, and MLP Regressor details.

2 Theoretical Background

Since this research topic is still evolving, there are various definitions and perspectives on fundamental terms. To clarify these perspectives, I will first define the term "Artificial Intelligence" in a quantitative manner. This is important because what constitutes "intelligence" is up for debate. In my opinion, this approach best reflects what most studies focus on when they discuss AI within the context of startup scaling.

2.1 Definition of AI

AI is a rapidly evolving field not only in business but also in research. To pinpoint what AI is, I decided to use a quantitative method to deduce the definition of AI in the current research literature.

The keyword analysis in bibliometrics provides an overview of keywords used together in scientific publications on any subject. Figure 2.1 depicts a result of my bibliometric analysis of 111 articles regarding the research question.

Further details for literature, data, search queries, paper preparation, and analyzing options are explained in Appendix 8.1.

It is reasonable to interpret the mind map of keywords in Figure 2.1 as a visual representation of current research topics as seen by the five leading ”Information Systems” journals.

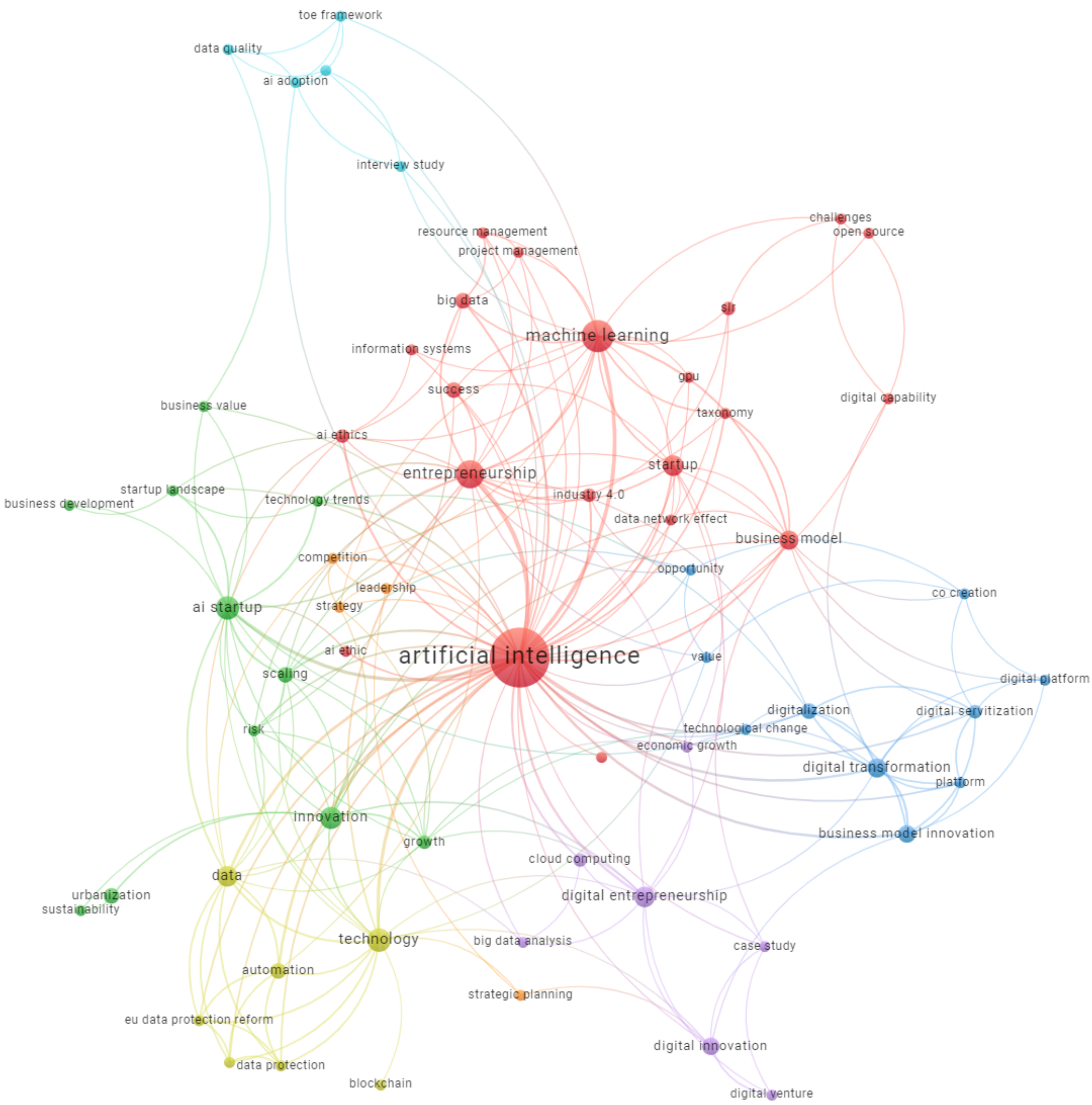


Figure 2.1: Bibliometric analysis: keyword co-occurrence analysis, own figure

The size of a node indicates how frequently a keyword is mentioned. A node that is double the size indicates that the keyword is used twice as often. From the mind map, it is evident that "Artificial Intelligence" is one of the most commonly used terms. It can be defined as a central concept that connects and influences various domains.

Table 2.1: AI-related topics and relevant terms

Category	**Relevant Terms**
Technological Topics	Machine Learning, Big Data, GPU, Industry 4.0, Data Network Effect, Cloud Computing
Digitalization and Business Models	Digital Platform, Digital Transformation, Digital Servitization, Digitalization, Business Model Innovation
Ethics and Regulation	AI Ethics, EU Data Protection Reform
Management and Strategy	Resource Management, Project Management, Leadership, Strategic Planning
Economic Impact	Economic Growth, Value, Opportunity, Business Value
Entrepreneurship and Innovation	Entrepreneurship, Digital Entrepreneurship, Innovation, AI Startup
Challenges and Opportunities	Challenges, Risk, Scaling, Competition

The mind map in Figure 2.1 and Table 2.1 illustrate that "Artificial Intelligence" is a complex and interconnected concept encompassing various technological, economic, ethical, and strategic dimensions. This analysis shows that the literature on the topic relates to multiple scientific fields, including information systems, business administration, computer science, and several other disciplines. Based on the graph in Figure 2.1, I define "Artificial Intelligence" as follows:

> Artificial Intelligence (AI) is a technological concept that relies on machine learning for digital transformation. It is often used in connection with areas such as entrepreneurship, startups, innovation, and transformative business models. AI is influenced by data quality and data analytics processes, and it plays a central role in the development of digital capabilities, for example, in platforms. It promotes automation and innovation, providing companies with opportunities to advance their strategy through data analytics, technologies, and technological changes.

2.2 Definitions of AI Startup

In general, the business model represents the focal business logic of a venture and is essential to the successful commercialization of any technology (Weber et al. 2021). An Artificial Intelligence Startup (AI startup) is a digital startup in which AI is a core element of the business model (Schulte-Althoff et al. 2021). In general, startups are conceptualized as growth-oriented companies with innovative behavior (Carland, Hoy, & Boultom 1984). More specifically, digital startups are described as ventures that market, offer, or support a digital good or service (Zhao & Collier 2016). They rely on aspects of digital media and IT to exploit market opportunities (Davidson & Vaast 2010).

This often involves the use of new digital technologies such as natural language processing, machine learning, big data analytics, 3D printing, deep learning, virtual reality, internet of things devices or services, or cloud computing (Schulte-Althoff et al. 2021). AI is an ever-evolving frontier of emerging computing capabilities. The machine learning technologies at the core of today's AI are more autonomous, adaptive, and opaque than any previous "intelligent" IT artifacts (Berente, Gu, Recker, & Santhanam 2021).

AI technologies, which include robotics and autonomous driving, pattern recognition, generative AI, text-to-speech, speech-to-text, and virtual agents, are being utilized across different domains (Berente et al. 2021). AI startups can be found in domains such as natural language processing, machine learning, predictive analytics, and intelligent systems (Schulte-Althoff et al. 2021).

2.3 Definitions of Scaling

Scaling refers to how the components of a system respond as its size changes (Schulte-Althoff et al. 2021; West 2019). The ability to scale is crucial for startups, particularly in securing funding, enhancing productivity, and accelerating the adoption of new products and technological innovations (Schulte-Althoff et al. 2021). Scaling in this thesis follows the definition of Schulte-Althoff et al. (2021) that says scaling is the measurable change in growth.

AI startups can extend traditional scaling methods by adapting powerful machine learning models to various business applications (Aggarwal 2018; Schulte-Althoff et al. 2021). Or by offering new types of services that surpass humans in terms of perception and cognition (Brynjolfsson & Rock 2010; Schulte-Althoff et al. 2021).

2.4 Overview Different Startup Types

Understanding the fundamental types of startups is crucial to analyzing their scaling intensity effectively. Broadly, startups can be categorized into four types: service, platform, AI-enabled service, and AI-enabled platform (Weber et al. 2021). Each type has distinct characteristics and operational models that influence its scaling potential and business dynamics. This section aims to provide a scientific definition and differentiation of these four types, laying the foundation for further analysis.

Service startups primarily offer tangible or intangible services directly to consumers or businesses. These services could range from consulting, healthcare, and education to more niche markets like pet grooming or specialized repair services. The primary value proposition of service startups is the direct delivery of expertise or labor-intensive tasks that meet specific customer needs. The scalability of service startups is often limited by the need for human intervention and customization, making rapid expansion challenging without a proportionate increase in resources (Weber et al. 2021).

Platform startups, on the other hand, facilitate interactions between two or more interdependent groups (Øverby & Audestad 2021), typically consumers and producers. Examples include e-commerce websites, social media platforms, and ride-sharing apps. Platform startups thrive on network effects, where the platform's value increases as more users join (Øverby & Audestad 2021). The scalability of platform startups is inherently higher than service startups because they can grow without a corresponding rise in operational costs. The primary challenge lies in balancing the interests of different user groups and maintaining an efficient and scalable technological infrastructure (Weber et al. 2021).

AI-enabled service startups integrate AI technologies into their service offerings to enhance efficiency, accuracy, and scalability. These startups leverage AI for tasks such as data analysis, customer support, and predictive maintenance. For instance, AI-enabled customer support systems can handle large volumes of inquiries with minimal human intervention, significantly boosting scalability (Lins et al. 2021; Ollig 2022). The integration of AI not only automates routine tasks but also provides advanced capabilities like real-time decision-making and anomaly detection, setting these startups apart from traditional service models (Weber et al. 2021).

AI-enabled platform startups combine the network-driven model of platforms with the advanced capabilities of AI. These startups use AI to optimize matchmaking between users, personalize user experiences, and automate various aspects of platform management. For example, AI can be employed in ride-sharing platforms to predict demand, optimize routes, and dynamically adjust pricing, thereby enhancing the platform's efficiency and user satisfaction (Lins et al. 2021; Weber et al. 2021). The dual advantages of network effects and AI-driven automation give these startups a unique edge in operational efficiency, improving customer satisfaction.

Differentiating between these types of startups has its problems; for example, a platform that sells AI services to customers is difficult to categorize (Ejsmont et al. 2024). In this case, I decided to categorize this type of business model as a platform because the business is primarily a marketplace for producers and consumers of services. Also, a case study by Jia and Stan (2021) underlines that for their startup, the company had implemented an internal data platform, a platform especially designated for AI optimization, with the intention to use the data network effects. Because the customer could not access the platform directly, and this offering resembles a platform from the customer's perspective, I categorized this business model as a platform. If the company resembles a service from the customer's perspective, the classification would lead to an AI-enabled service startup.

After separating these four types, it is possible to understand their unique scaling challenges and opportunities. Service and platform startups represent more traditional models with distinct scalability constraints and enablers. In contrast, AI-enabled service startups and AI-enabled platform startups introduce a layer of technological sophistication that can significantly alter their growth. For instance, while a traditional service startup might struggle with scaling due to human resource limitations, an AI-enabled service startup business model might overcome these limitations.

Service and platform startups offer a more conventional approach, while AI-enabled service and AI-enabled platform startups leverage cutting-edge technologies to overcome traditional scaling barriers (Weber et al. 2021). This foundational understanding is crucial for analyzing the scaling intensity of AI-enabled platforms and services.

2.5 AI-enabled Service Startup Specifics

To understand AI's critical role in service startup business models, it is essential to investigate whether these businesses can function without AI technology. If the startup can not function without AI, it means that AI is at the core of the service-business model.

This analysis begins by examining how AI enables or significantly enhances the services offered by startups. For instance, services like fraud detection or disease diagnosis are areas where AI outperforms human capabilities, making these offerings not just enhanced but fundamentally enabled by AI (Weber et al. 2021). Without AI, such services would either be impossible or require extensive human resources, thus not being scalable or economically viable.

The underlying novelty of AI startup business models also hinges on their ability to leverage AI capabilities in previously unattainable ways with traditional Information Technology (IT) solutions. While IT-business models have existed for some time, the unique capabilities of AI, such as continuous learning and real-time decision-making, offer new value propositions that are integral to their service (Weber et al. 2021).

Finally, to clearly distinguish how AI constitutes the core of these business models, it is crucial to consider the business logic and value creation mechanisms of these AI-enabled services. The taxonomy of AI startups reveals that AI technology impacts the overall business logic by introducing new roles for data and altering the traditional value creation process (Weber et al. 2021). This fundamental shift indicates that the service offerings, in their current form, are deeply intertwined with AI technology, making them significantly enabled and fundamentally dependent on AI for their existence and scalability (Weber et al. 2021).

2.6 AI-enabled Platform Startup Specifics

When approaching the research question, it is essential to clarify AI's role in platform startups. To understand the critical role that AI plays in the business models of platform startups, it is essential to investigate whether these businesses can function without AI. If the startup can not function without AI, it means that AI is at the core of the platform business model. This question is relevant as it distinguishes between startups that merely integrate AI as an auxiliary tool and those for which AI is essential.

For example, AI is the backbone for some platforms, enabling core functionalities such as real-time data analytics and personalized user recommendations, which would be unfeasible without it (Weber et al. 2021). In contrast, other platforms might use AI to optimize operations but could still function less efficiently without it.

One of the main aspects of platform startups in this work is the question of whether the platform is accessible to other parties or not; the "matchmaking" character of the platform needs to be dominant. For example, Netflix matches film producers and film consumers on its platform. In this case, AI enables precise and fast matchmaking. If a company, on the other hand, uses internal and closed AI platforms, for example, to analyze market conditions and its customers' behavior, but no external party can assess this platform, this "internal" platform will not be viewed as relevant for the AI-enabled platform startup classification, even though an AI-platform is used in the business model. From the point of view of a customer or potential matchmaking-interested party, this platform is not accessible, and, most likely, only an AI-optimized service from this company can be bought.

2.7 Scaling Mechanisms in Business Models

Scaling can be understood through various definitions and perspectives. In AI ventures, scaling refers to the ability to grow by leveraging digital and human resources effectively (Vartak 2022). AI ventures create and capture value through repetition within a domain, between domains, and by acquiring resources for repetitive value generation and capture (Zebhauser, Rothe, & Sundermeier 2023). Digital ventures, including AI startups, benefit from digital infrastructure that enables flexible and scalable design, allowing rapid adaptation to changing environments (Zebhauser et al. 2023). This scalability is crucial in winner-take-all markets, where being faster than competitors can determine success (Schilling 2002; Zebhauser et al. 2023). Here, I simply use the word scaling in the sense of change in the growth of startups, corresponding to the work of Schulte-Althoff et al. (2021).

The results of my systematic literature research are presented in the following paragraphs. Additionally, I will elaborate on the mechanisms behind scaling service and platform business models.

In AI ventures, data network effects are a significant scaling mechanism. The value increases for both AI-enabled services and platforms as more data is collected and analyzed, improving the AI models and attracting more users. This continuous learning and data accumulation create a feedback loop that enhances the service or platform over time, exemplified by startups that build competitive advantages through superior algorithm performance (Ollig 2022; Weber et al. 2021).

Automation is another critical mechanism for scaling. AI ventures can automate repetitive tasks, reducing the need for human intervention and increasing efficiency. In AI-enabled services and platforms, automation can enhance scalability by processing large amounts of data quickly and accurately, as seen in AI as a Service (AIaaS) offerings that optimize hardware resources and handle infrastructure failures automatically (Lins et al. 2021). Automation allows AI ventures to serve more clients with fewer resources, facilitating rapid growth and scaling.

The abstraction of AI refers to simplifying AI technologies so that users can interact with them without needing in-depth technical knowledge. This abstraction can drive scaling for AI-enabled services and platforms by making AI more accessible to a broader audience. AIaaS offerings exemplify this by guiding users through AI development and deployment processes without requiring them to understand complex algorithms (Lins et al. 2021). By lowering the barriers to entry, AI ventures can attract more users and scale their operations more effectively. As exemplified by AIaaS, abstraction and automation could supplement each other (Lins et al. 2021).

Data Driven Decision Making (DDDM) is essential for scaling in AI ventures. By leveraging data analytics, AI startups can make informed decisions that improve performance and scalability. DDDM helps organizations align their processes with AI strategies, ensuring compatibility and facilitating AI adoption (Jöhnk et al. 2021).This method allows AI companies to continuously optimize operations, improve customer satisfaction, and discover new market opportunities, making it essential for successful scaling (Mittapally 2024).

Additionally, network effects are crucial for scaling platforms and AI-enabled platforms. As more users join a platform, the value of the platform increases for all users, creating a self-reinforcing growth loop (Øverby & Audestad 2021). This is evident in digital platforms that transform industries by creating lock-in effects and fostering growth through network-driven dynamics (Ollig 2022). For AI-enabled platforms, network effects are amplified by data network effects, where the increasing user base generates more data, further enhancing the AI capabilities and attracting even more users (Jia & Stan 2021).

In conclusion, scaling intensity in AI ventures is driven by mechanisms such as data network effects, automation, data-driven decision-making, AI abstraction, and network effects. These mechanisms enable AI startups to grow rapidly by leveraging digital and human resources effectively, creating a competitive edge in dynamic markets. Understanding and harnessing these scaling mechanisms is essential for AI ventures aiming to achieve sustainable growth and long-term success.

2.8 Universal Scaling Mechanisms

Universal scaling refers to the consistent patterns that can be observed in various systems, such as cities, organisms, and companies, across different scales. Understanding these patterns can provide insights into their growth and sustainability in the context of AI-enabled platforms and service startups. Companies, like organisms, typically demonstrate low power growth and eventually plateau due to constraints such as resource limitations and market saturation (West 2019). This suggests that startups, initially experiencing rapid growth, will eventually face similar constraints (West 2019). By exploring the scaling laws applicable to companies, we can better understand the underlying dynamics of AI startups and predict their potential and long-term success.

Net income, gross profit, and the number of employees are crucial metrics for analyzing the scaling intensity of startups. These metrics show a company's financial health and operational efficiency. Net income represents the profitability of a company after all expenses have been deducted, serving as a measure of its ability to generate profits sustainably (West 2019). Gross profit, on the other hand, reflects the efficiency of core business activities by showing the difference between sales and the cost of goods sold. The number of employees indicates the size and capacity of the company to undertake operations and manage increasing demands. Analyzing these metrics allows us to assess how effectively a startup is scaling its operations in response to market demands and resource constraints.

The underlying mathematical form of scaling is assumed to be a power function (West 2019). West (2019) specifically analyzes the scaling exponent by considering that a company's income and expenses both rise power-oriented as it employs more workers. He divides scaling into three cases: sublinear, linear, and superlinear. The concept of linearity in this context refers to the axis scaling, which is logarithmic, making power functions appear "linear".

He discovered that expenses initially grow in a sublinear fashion and later transition to a linear form as more employees are added, resulting in an exponent of less than or equal to 1. In contrast, he found that income can either grow in a superlinear manner (exponent greater than 1) or in a sublinear manner (exponent less than 1), but not in between. If the sublinear growth of income persists for too long, the company risks "dying".

He mentioned a scaling mechanism for small companies, particularly in his model. Since the expenses of young companies are sublinear while their income is typically linear or even superlinear, these companies can increase the value of their shares. This, in turn, attracts more shareholders and fuels further growth. However, this scaling effect diminishes when the expenses become linear, meaning the exponent equals 1. Nevertheless, scaling for young companies or startups can be achieved this way.

In Chapter 9 of the book by West (2019), the researcher examines the scaling of companies by plotting net income and gross profit for each company in the database against the number of employees. As illustrated in Figure 2.2, West (2019) also calculates the mean values for net income and gross profit in relation to the number of employees. These mean values are represented as white dots and are connected to form a line. Since the plot is displayed on a log-log scale and the line connecting the mean values appears straight, this is interpreted as an indication of power-function growth scaling behavior. The slope of this line represents the exponent of the power function describing the scaling behavior.

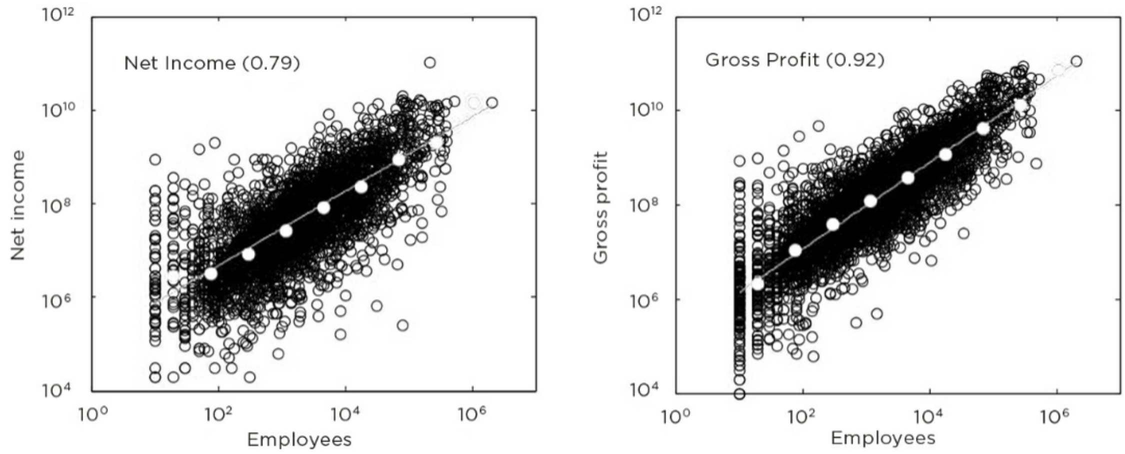


Figure 2.2: Net income and gross profit in USD publicly traded companies in the United States from 1950 to 2009 in dependence on the number of employees in the company, West (2019)

The plots shown in Figure 2.2 use the Compustat data set, which contains around 30,000 companies traded on U.S. markets between 1950 and 2009 (West 2019).

The R^2, which in this case is 0.79 (net income) and 0.92 (gross profit), quantifies how well the power-function growth model fits companies' actual data compared to simply calculating the average. A high R^2 value for AI-enabled startups indicates that the power growth model accurately represents their growth patterns during the early stages. However, as startups scale and encounter various constraints, the fit may become less precise, reflecting the complex and dynamic nature of business growth (West 2019).

This master's thesis adopts the quantitative methods and analytical concepts from West (2019) to better understand the scalability and potential limitations of AI-enabled platforms and service startups.

3 Data

This chapter outlines the methodology used to collect, prepare, and analyze the dataset for this study. The primary data source is the PitchBook database (pitchbook 2022), which contains comprehensive information on various companies, including startups in different sectors. Most users know this website or front-end from PitchBook:

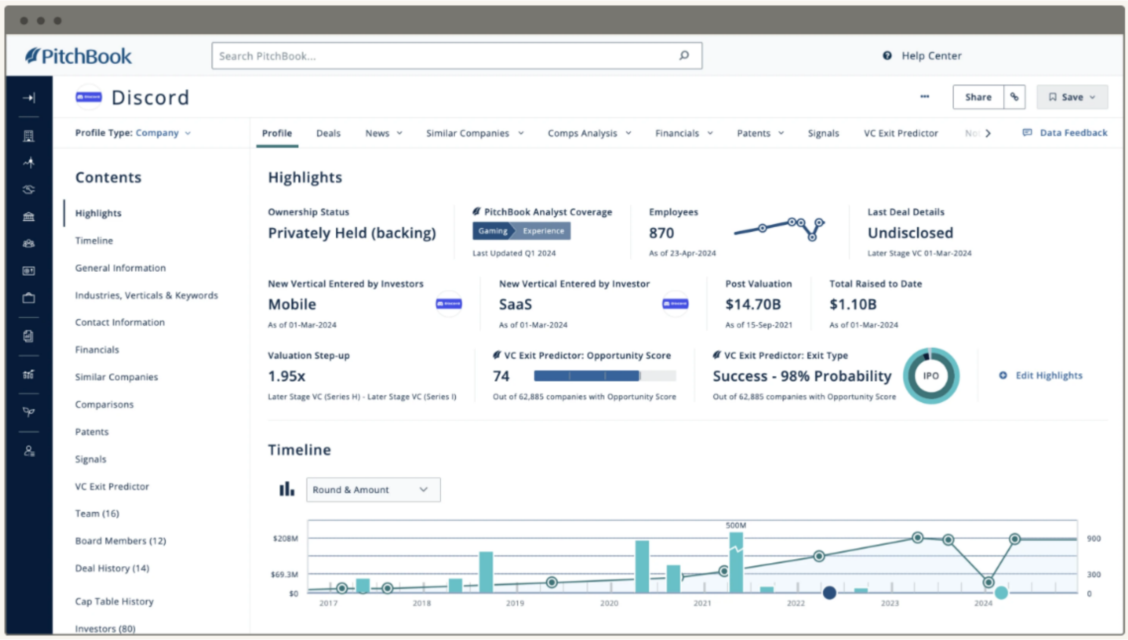


Figure 3.1: PitchBook front-end (pitchbook 2024)

The analysis aims to compare four types of startups: non-AI service startups, AI-enabled service startups, non-AI platform startups, and AI-enabled platform startups. It focuses on identifying and evaluating the differences between these categories. Relevant data will need to be gathered for this analysis.

3.1 Data Collection

3.1.1 PitchBook Dataset Overview

The data for this study was obtained from PitchBook (pitchbook 2022), a relatively big database widely used by investors, researchers, and analysts to gather insights on companies, transactions, investors, and related professionals. According to PitchBook, the platform provides the most comprehensive data on European public and private markets, covering over 4.6 million companies, including approximately 47,000 companies labeled as startups.

For this research, data was accessed through Comma Separated Values (CSV) files, which were derived from JavaScript Object Notation (JSON) files provided by my supervisor via the PitchBook Application Programming Interface (API). Each CSV file corresponds to a specific startup category, defined by the presence or absence of AI capabilities and the business model type (service vs. platform). Separate API queries were constructed for each of the four categories, as outlined below.

3.1.2 Data Extraction and Classification Criteria

The following criteria and query logic defined this study's AI-related and business model categories. The intention is to address the vertical cells and the description cells with these strings.

Table 3.1 first defines what ”AI” means in the database; it means the following strings are part of any cell, like the string ”Artificial Intelligence” or ”Data Analytics”. In the next step, the criteria for, for example, an AI-enabled platform are written. It uses the ”AI” strings and evaluates if other strings like ”Platform” or ”Broker” are present in this special database entry. If both string criteria are present, the database entry is written in the AI-enabled platform JSON folder.

Table 3.1: Data procurement queries for AI-enabled service, AI-enabled platform, and their non-AI counterparts

Category	**Conditions**
AI	*AI = "Artificial Intelligence" OR "Data Analytics" OR "Artificial Intelligence & Machine Learning"*
AI-enabled Platform	*AI AND ("Platform" OR "Broker" OR "Intermediary Service" OR "Marketplace" OR "Digital Ecosystem" OR "Social/Platform Software")*
non-AI Platform	*NOT AI AND (Platform OR Broker OR Intermediary Service OR Marketplace OR Digital Ecosystem OR Social/Platform Software)*
AI-enabled Service	*AI AND (Digital Services OR Commercial Services OR Services (Non-Financial) OR Business Products and Services Business to Business (B2B) OR Consumer Products and Services (B2C) OR Financial Services) AND Information Technology AND NOT (platform OR Broker OR Intermediary Service OR Marketplace OR Digital Ecosystem OR Social/Platform Software)*
non-AI Service	*NOT AI AND (Digital Services OR Commercial Services OR Services (Non-Financial) OR Business Products and Services (B2B) OR Consumer Products and Services Business to Consumer (B2C) OR Financial Services) AND Information Technology AND NOT (platform OR Broker OR Intermediary Service OR Marketplace OR Digital Ecosystem OR Social/Platform Software)*

The resulting JSON files contained data with 364 attributes per entry, providing comprehensive information on each company. Some expressions also include vertical column-specific filters, allowing segmentation based on the primary market or sector. For instance, a startup vertical under ”Artificial Intelligence & Machine Learning” is tagged as AI-enabled, and further logical expressions with strings determine whether it is a platform or service-based startup. The full query with vertical details is in

Appendix 8.3.

Following this categorization, we obtain four separate CSV files: AI-enabled platform (platform AI), non-AI platform (platformNonAI), AI-enabled service (serviceAI), and non-AI service (serviceNonAI) startups.

3.1.3 Data Classification Validation

After the procurement, four CSV files were obtained, corresponding to each startup category. Given the prior theoretical definitions of "service" and "platform" as well as AI-enabled vs. non-AI criteria, it was essential to verify classification accuracy. To assess the consistency of classifications, I reviewed 100 random entries per category. Table 3.2 summarizes the classification results, with details on correct, incorrect, and unclear classifications:

Table 3.2: Startup classification-check, more details in Appendix 8.11

startup class	checked sample size	correct classification	incorrect classification	unclear class
platform AI	100	84	15	2
platform non-AI	100	70	27	3
service AI	100	94	5	1
service non-AI	100	75	13	12

We can analyze the results for "Platform AI/Non-AI" and "Service AI/Non-AI" in greater detail. The classification outcomes for each of the four categories show variations in the number of correctly and incorrectly identified startups. Below is a brief overview and an evaluation of how these findings relate to the foundational concepts established in Chapter 2.

”Platform AI” sample: The sample contains a lot of Human Resources (HR) matching platforms, a lot of marketing-related platforms for influencers and brand matchings and supplier/material in supply chain optimization platforms, and some relatively large Chinese startups for self-driving cars. The entries in the sample did not show any suspicious industries or inconsistent information.

The most incorrect classification is seen within the ”Platform AI” category for startups that offer internal, non-public platforms (e.g., internal AI-enabled platforms). This deviates from this study’s requirement that platforms facilitate external user interactions with a matchmaking function between distinct user groups. These are 15 entries from 100 that were validated.

”Service AI” sample: In this sample, we find some military startups, like autonomous shooting and flying robots, also robots for logistics, startups that sell some better sensor algorithms, startups that focus on cyber security know-how and some text-generating AI products, like chatbots and email bots and of course startups for internal AI platforms installation as an extension of the digitization process.

In the ”service AI” class, four out of the five misclassified startups were AI-enabled platforms, and one changed its business to selling Apple watch bands. So, overall, the service AI class is a very good classified category.

”Service non-AI” sample: To summarize the sample findings, many startups sell gadgets for households, and digitization services for medium and small companies are also present. Services to improve consumers’ lives, like house cleaning, travel services, dancing schools, and education, are also present. In this sample of 100 entries, 13 startups were classified incorrectly. Platforms, like online shops for shoes and electronics, are the most misclassified as non-AI services.

”Platform non-AI” sample: This category is a relatively dispersed sample of startups. Here are, as I would call them, ”Indian online casinos”, regular gaming and gambling platforms, and a lot of payment platforms, which are less known competitors to companies like PayPal; further, there are conventional online shops for groceries and industrial material, there are also some startups for online adult content as well as voucher platforms for shopping centers are also present. In this sample of 100 entries, as many as 27 startups were classified incorrectly.

This misclassification often has the following reason: if a platform is internal and not accessible to matchmaking-interested parties, I view it as an IT service and not as a platform in this study. In addition, three mismatched companies in this sample have, in reality, AI-enabled platforms and, therefore, are not non-AI platforms.

Unclear class: The classification of certain startups in these PitchBook data samples could not be verified through available online sources. This was sometimes due to language barriers, while in other cases, the startups had no online presence, such as a website or LinkedIn profile. Among the 400 randomly tested entries, 18 startups lacked any online presence and may no longer exist as of the time of writing.

3.2 Data Preparation

In this study, every startup from the database is classified into one of four distinct business models. Each startup-founded year falls between 2014 and 2024. This selection criterion establishes the overall context of the research.

3.2.1 Selection of Key Variables

The extensive dataset in PitchBook includes numerous variables that are suitable for analysis. Drawing inspiration from the scaling analysis conducted by Schulte-Althoff et al. (2021), I have identified key variables that align with the criteria for scaling analysis. Additionally, the scaling analysis from West further informed the selection of these scaling variables.

This study examines the selected variables presented in Table 3.3.

Table 3.3: Selected variables in the dataset for this master thesis

Selected variables:	**Variable chosen based on:**
Employees	Basis for every scaling analysis
Revenue	For scaling analysis like Schulte-Althoff et al. (2021)
VC Raised	For scaling analysis like Schulte-Althoff et al. (2021)
Net Income	For scaling analysis like West (2019)
Gross Profit	For scaling analysis like West (2019)
Location Data: HqCity, HqCountry, HqRegion, HqStateProvince	To study regional dependence;
Industry Data: primaryIndustryCode, PrimaryIndustryGroup, PrimaryIndustrySector	To study dependence on industrial branches;

The variable *Employees* is the total number of employees within a company. The variable *Revenue* is defined as the gross income generated over 12 months or one year before any expenses are subtracted (PichBook 2024a). The variable *Gross Profit* is calculated by subtracting the year's costs of goods sold from the total revenue (PichBook 2024a).

The variable *VC Raised* indicates the total amount of venture capital raised up to a specific date (PichBook 2024b). The abbreviation for venture capital is VC. Additional variables, such as “Deal Type” and the specific "Date" of fundraising, also exist (PichBook 2024b). However, this analysis will focus solely on the amount raised in thousands of Euros over the last 10 years.

The variable *Net Income* represents a company's profit after all expenses, taxes, and costs, including operating costs, interest, and depreciation, have been subtracted from its total revenue. This key figure is typically calculated annually as part of financial reporting and is a crucial measure of a company's profitability and economic health. Investors often consider this key metric when assessing a company's performance

(Haan 2022). The relevant chapters will provide further details about other variables if used.

The database's revenue, gross profit, costs, or net income are the last published information. These company data entries are either for 2023 or the previously published data for earlier years.

Before analyzing the data, the underlying information for the selected variables in Table 3.3 must be prepared. I will conduct the data-cleaning process described in the following section.

3.2.2 Data Cleaning in Variables

The following steps were undertaken for each of the four sets to obtain high-quality data:

- **Identification**: Each company has a unique Identifier (ID). For each tuple, which represents a company, only relevant variables, based on Table 3.3, are kept for further cleaning and investigation. IDs are not relevant for analysis but are kept for matching tuples.

- **Screening**: The data was screened for duplicates and missing values. Duplicates were excluded, and any missing data in cells was marked with "Not a Number (NaN)".

- **Filtering**: Some analyses required specific filters, such as examining the number of employees above or below a certain threshold. Whenever such a filter is crucial, it will be mentioned in the relevant chapters.

- **Validation**: Invalid entries, such as "NaN" or placeholder values like 99, 999, 9999, or alike, as well as negative values, were excluded. Additionally, a manual verification of the correct business model classification was conducted, as described in Chapter 3.1.3.

- **Final Dataset**: As a result, four sets have been created: Service AI, Platform AI, Service non-AI, and Platform non-AI, each containing valid data for employees and a selected variable (e.g., revenue).

To provide an overview of the data volumes (i.e., numbers of tuples) at every stage mentioned above, Figure 3.2 shows a visualization of the procedure for the employee-revenue-scaling analysis.

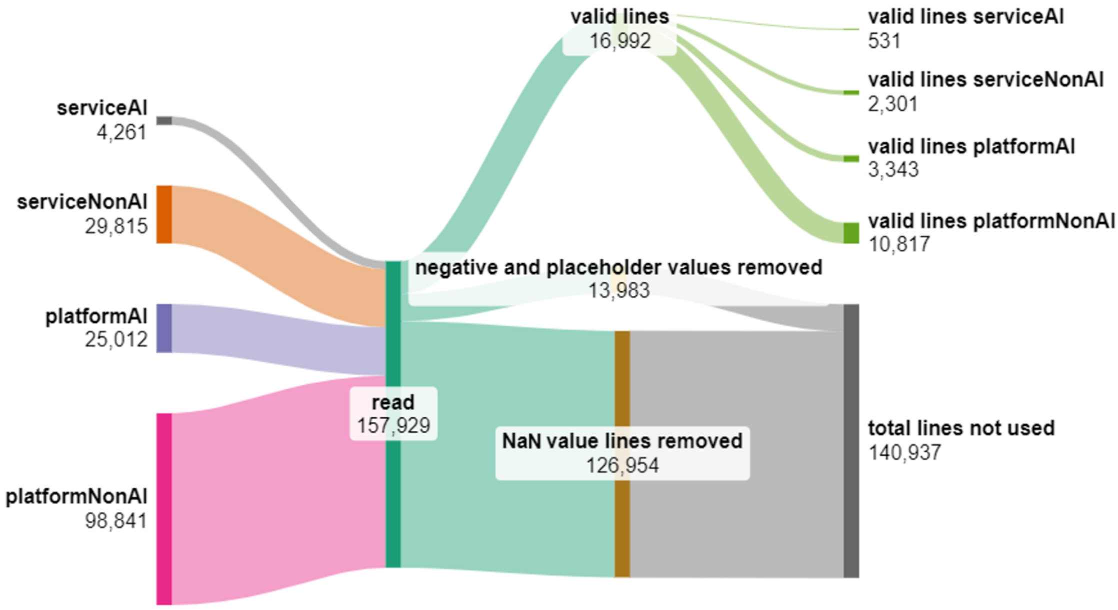


Figure 3.2: Data preparation flow for employee-revenue scaling analysis, own figure

First, I gather all four CSV files containing data for the four business models. Together, these files contain approximately 158,000 data points across all categories. The next step is data cleaning, removing NaN values, placeholder entries, and negative values. As a result, only a few valid and useful lines remain for each dataset. Most of the processed entries do not meet the basic criteria of having both the number of employees and the second variable with a value greater than zero (for example, revenue greater than zero). After this step, the valid data tuples are archived for each dataset, allowing the study of employee-revenue scaling.

In the particular case of employee-revenue-scaling analysis, this handling means that all startups with zero revenue are excluded.

3.2.3 Logarithmic Representation of Data

In the following chapters, we will examine various data presented in log-log graphs. These graphs illustrate the relationship between a specific variable and the number of employees in each company. Each plot displays a cluster of data points along with the best-fitted power function to explore the scaling patterns. Some of the power-function fits in these graphs may appear surprising, and since their interpretation is not straightforward, we need to discuss them in detail, particularly in relation to the specific data we will analyze after the data-cleaning process.

In the following example, I illustrate the point more clearly. Both pictures in Figure 3.3 display the same scatter plot of revenue against employees, using linear and log-log axes. The red line represents the best fitted power function for the data points.

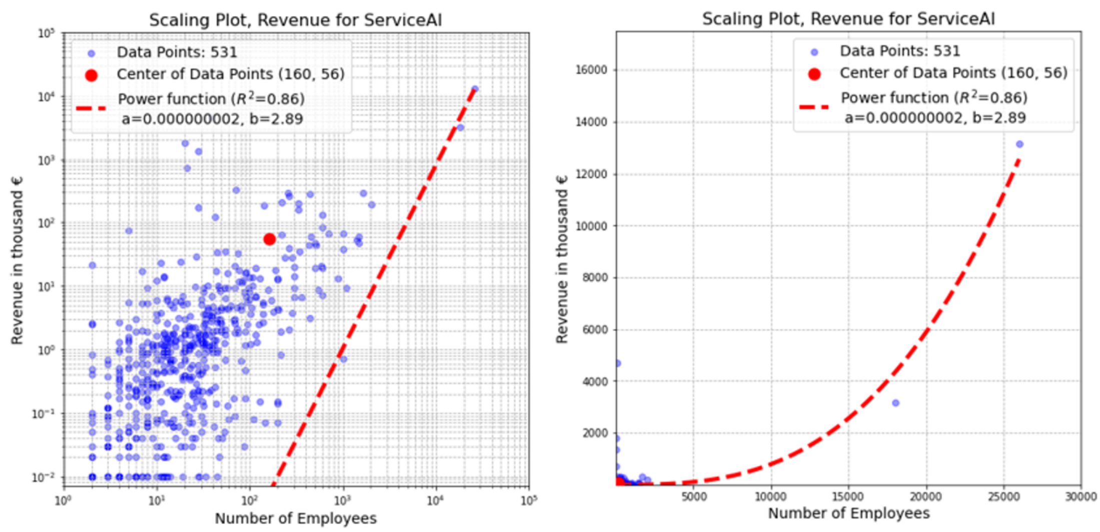


Figure 3.3: Comparison of the same power function fit in two graphical representations, own figure

The optimal power function, represented by the red line in both graphs, appears as a straight line in the log-log representation, with a slope determined by the exponent.

However, the red line in the log-log plot seems to be misaligned with most data points. The right plot clarifies this "misalignment" by highlighting the presence of two outliers, which have both a high number of employees and high revenues. In comparison, all other data points have relatively low employee numbers and revenues, typically below 100,000€. Since the primary goal of any startup is to generate revenue, the two outliers, representing successful startups, are crucial to the analysis. These data points cannot be excluded when fitting the power function.

These two outliers significantly influence the fit of the power function, as shown in the graph on the right. The optimization process minimizes the sum of squares, meaning that the larger values of the two outliers heavily influence the fit in the log-log representation. In comparison, the numerous smaller values have minimal impact. Further details are in Appendix 8.9.1.

The outliers in this dataset are valid real-world measurements and will not be excluded from the analysis. These two outliers represent what are known as "unicorn" startups.

This study further employs the log-log representation because it effectively illustrates and facilitates comparison of the slopes across different types of startups, particularly in Chapter 5.

3.3 Dataset Visualization

3.3.1 Summary of Descriptive Statistics

The raw data table, which is used in this analysis, and also histograms of revenue, employee, and VC raised, are in Appendix 8.2.

Employee Numbers

The average number of employees in the service AI category is 57, which is relatively low compared to non-AI Services, which has an average of 81. The Platform AI category has the lowest average employee count, at 44, while for non-AI platforms, this is close to 49. This suggests that companies in the service sector tend to employ more staff than those in the platform sector.

Revenue

The average revenue also shows significant differences. Service AI generates an average of 56 thousand €, while Service non-AI exhibits a remarkable increase with an average of 79 thousand €. In comparison, the average values for Platform AI (47 thousand €) and Platform non-AI (78 thousand €) are also notable, with Platform AI recording the lowest revenue. These results may indicate that companies in the service sector, particularly in the non-AI domain, are more efficient in revenue generation. It is important to notice that these revenue values are very low and, in the context of European or North American regions, approximately zero.

VC Raised

Regarding the raised VC (VC Raised), the values show that Service AI receives the highest average sum at 20 thousand €, followed by Service non-AI at 19 thousand €. Platform AI and Platform non-AI have lower average values of 17 and 16 thousand €, respectively, which may suggest a lower risk willingness or investment attractiveness in the platform sector.

Net Income and Gross Profit

Regarding net income, Service non-AI shows an average of 11 thousand €, representing the highest value among the categories. Gross profit is also relevant, with Service non-AI (32 thousand €) having the highest figure, followed by Service AI (23 thousand €). The values for Platform AI and Platform non-AI are comparatively lower, indicating that the service segments demonstrate better profitability.

3.3.2 Selected Status Variables

Business Status

One of the startup's characteristics is its business status in PitchBook. The "Business Status" is an attribute provided by PitchBook that indicates whether a company is generating revenue, has gone out of business, or has been acquired by other companies.

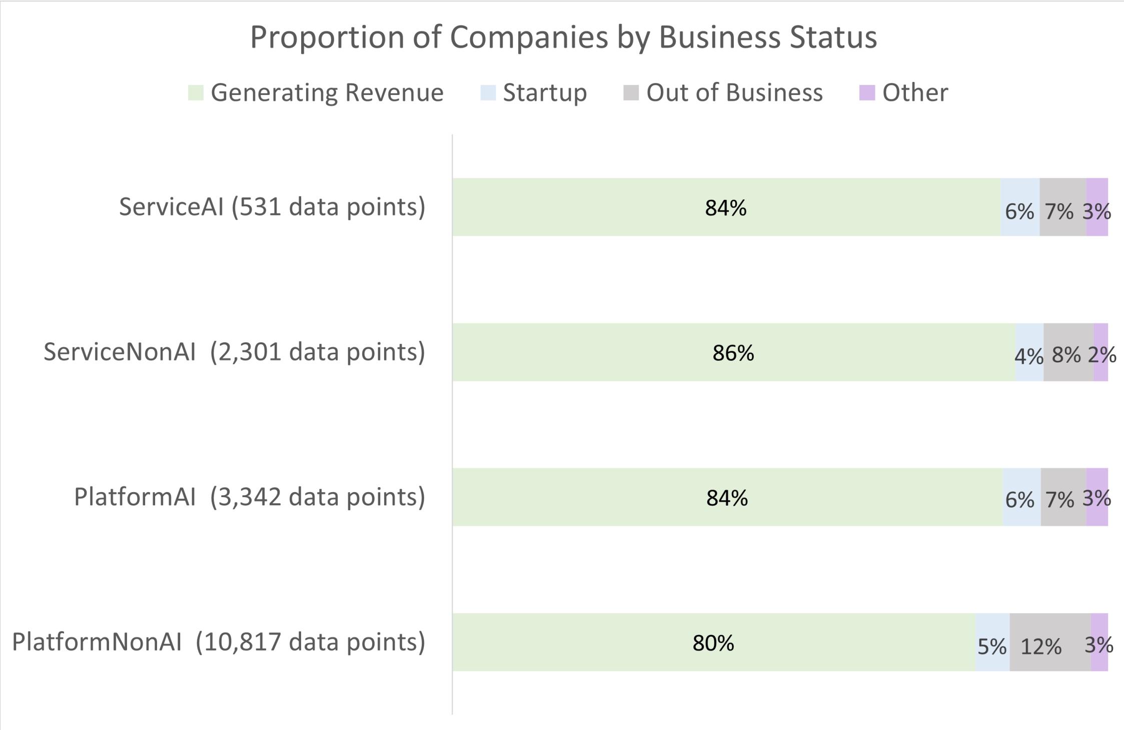


Figure 3.4: Business status of startups in the PitchBook database, own figure

As shown in Figure 3.4, most companies are in the ”Generating Revenue” status, qualifying them for the scaling analysis. Another notable status is ”Startup”.

Financing Status

Since PitchBook is also a database for investors, it is interesting to examine the startup's ”Financing Status”.

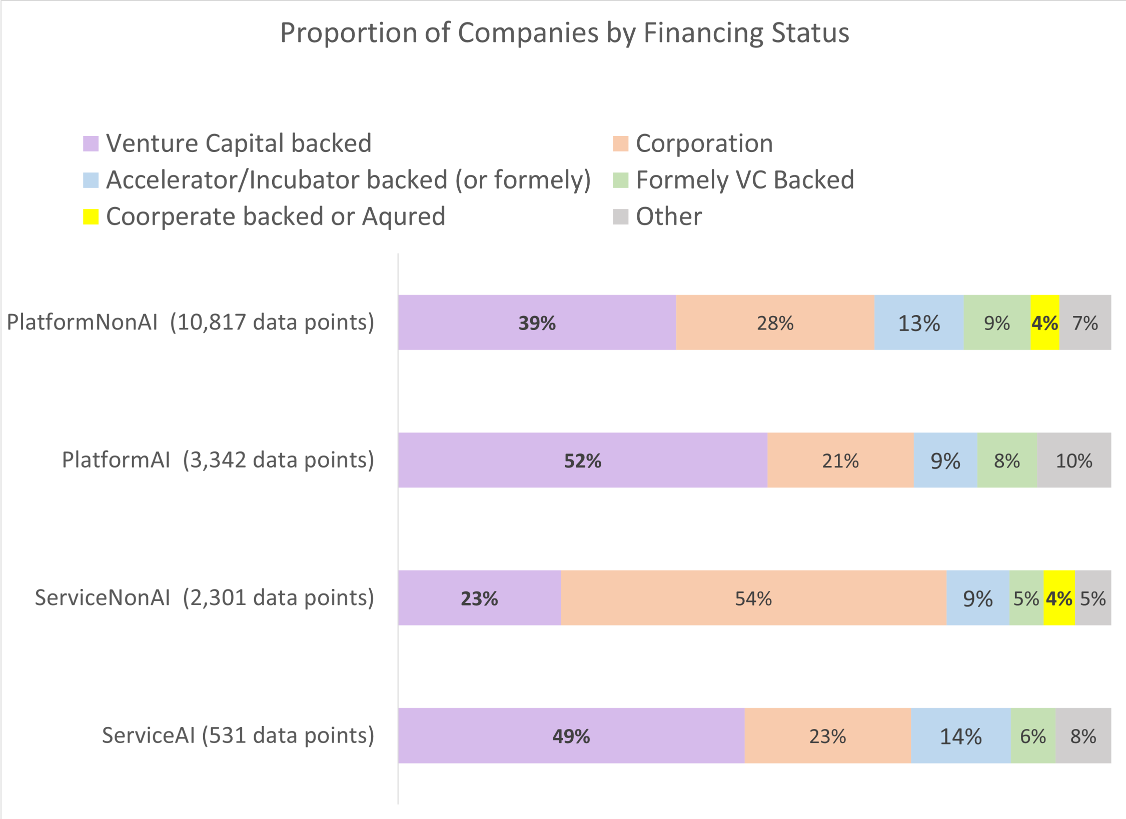


Figure 3.5: Financing status of startups in the PitchBook database alongside business model, own figure

You can see in Figure 3.5 that, especially for service AI startups compared to service non-AI startups, there are more than double as many startups with venture capital. Overall, it seems that AI startups are more often venture capital-backed than non-AI startups.

Geographical Status

After a small analysis of the previous variables, I also present selected geographical information of the data set.

North America and Western Europe are the predominant continents in terms of startup locations in the dataset. Interestingly, Southeast Asia is primarily represented in platform startups. For more information on startups by continent, please refer to Appendix 8.3.

The dataset also includes information about the countries where the headquarters of these startups are registered. This information is illustrated on the map shown in Figure 3.6.

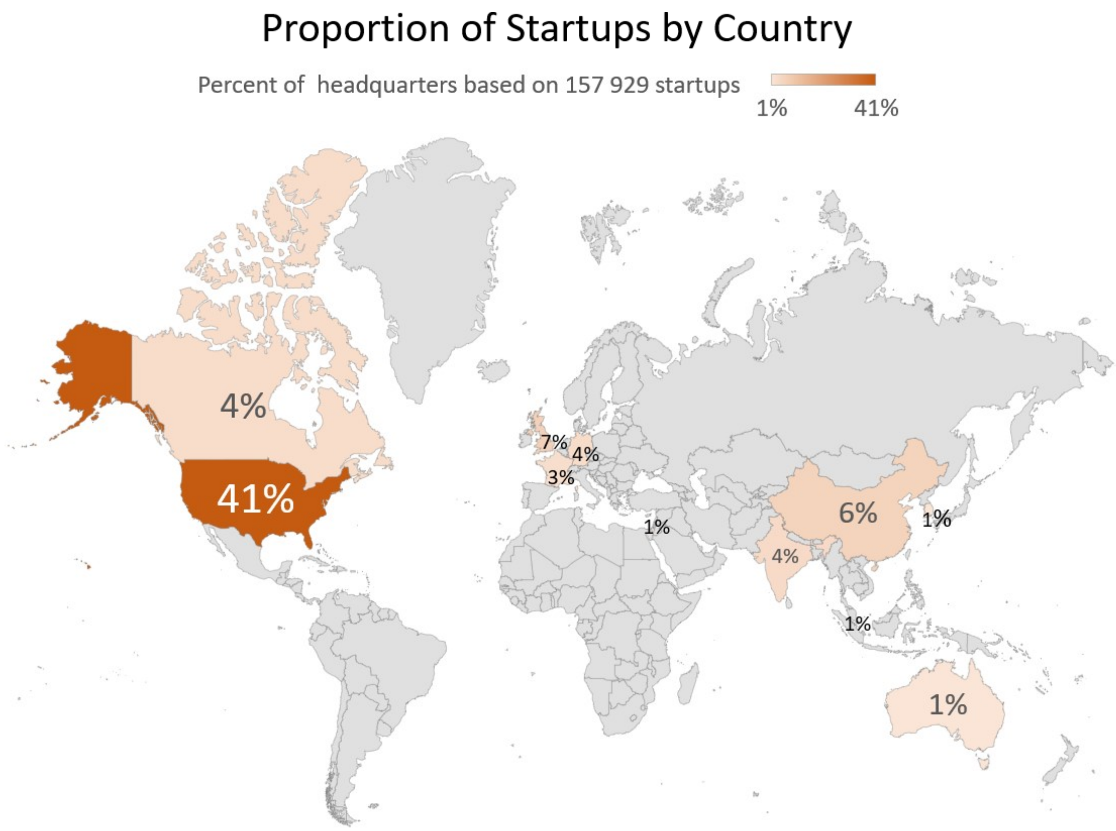


Figure 3.6: Headquarters country of startups in PitchBook, own map

In Map 3.6, countries with less than 1% of the data points are grouped into the "Other" category, shown in grey, this category accounts for 26.7% of all values in the dataset. The countries that contribute around 1% of the data points include Israel, South Korea, and Singapore. Overall, the dataset is predominantly focused on the United States of America.

Year Founded Status

This information is interesting because it demonstrates that different types of startups do not experience the same peak in founding years.

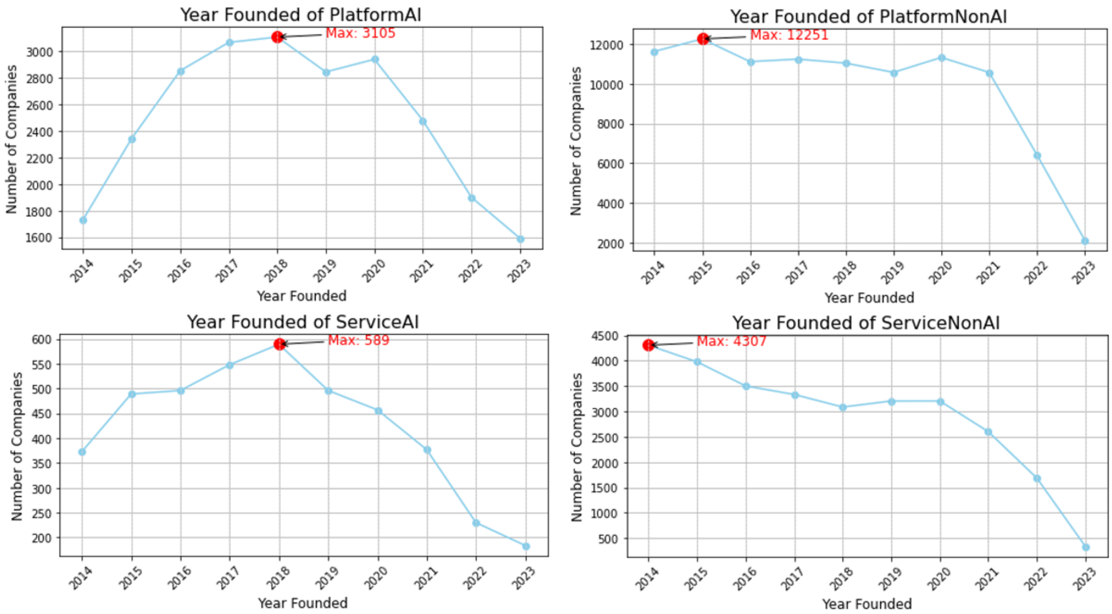


Figure 3.7: Founding year of startups by business model type, own figure

It is important to keep in mind that new startups will continue to be added later to the database, which may correct the current downward trend in the data over time.

Further visualizations and additional data-related details can be found in Appendix 3.3.

4 Methodology

This is an empirical work using quantitative and exploratory methods. The goal is to measure and then understand how scaling works in startups that are using AI in their business models. Since the topic is relatively new, there is a lack of data-oriented studies. It is important to examine the "hard facts" of startups alongside interviews or questionnaires, which provide valuable insights but can be influenced by subjective impressions. However, a measurable difference in scaling with and without AI would provide a more reliable understanding of the impact of AI on business models.

In the previous Chapter 3, we covered the dataset and the selected variables, which are listed in Table 3.3. This chapter will present the mathematical regression models associated with these variables.

4.1 Functional Form: Power Function

Geoffrey West performed a scaling analysis, which is as follows: He demonstrated the power scaling law using the Compustat dataset, which includes information on all 28,853 companies that traded on U.S. markets over a period of sixty years, from 1950 to 2009. West calculated the mean net income and mean gross profits for each value of employee numbers. When plotting the logarithmic mean values against the logarithm of employee numbers, the results displayed a linear relationship.

In West's study, the mean net income dependent on the number of employees exhibits a power law relationship. A similar power law relationship is observed for mean gross profit. These two power laws are derived from the data analysis.

This study aims to investigate these findings and assess whether they are applicable to the Pitchbook dataset, which focuses on startups. Assuming that the power law is valid for this dataset as well, I will fit the data accordingly and calculate the R^2 value to evaluate the quality of the fit for the Pitchbook data. The formula used for the scaling analysis and for testing the power law fit is specified below, along with a description of the relevant variables from the dataset.

$$y = a \cdot x^b$$

Here:

- y: dependent variable (e.g., net income, gross profit, VC raised, or revenue)

- a: scaling factor
- x: independent variable (e.g., number of employees)
- b: exponent

This method utilizes a general power function, where b represents the exponent and a is a scaling factor.

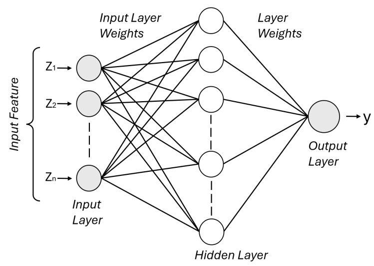
In West's model, the exponent is a crucial aspect of the power formula. This exponent is categorized into three types of growth: sublinear, linear, and superlinear. The term "linear" specifically refers to the logarithmic scale, where power functions appear linear.

4.2 Other Functional Forms

My primary mathematical model for analyzing the relationship between the studied variables, such as revenue and the number of employees, will be the power function. This approach assumes that the scaling of revenue per employee follows a power law.

Furthermore, it may be interesting to evaluate various mathematical models, as one of them could also demonstrate a high R^2 value and provide additional insights into the data.

Table 4.1: Overview of various regression methods

Method	**Formula**
Power Scaling / Power law Function	$y = a \cdot x^b$
Support Vector Regression	$y = \sum_{i=1}^{N} \alpha_i K(x, x_i) + b$
Polynomial Regression	$y = a + b_1 \cdot x + b_2 \cdot x^2 + b_3 \cdot x^3 + \dots$
Linear Interpolation	$y = y_0 + \frac{(x-x_0)(y_1-y_0)}{x_1-x_0}$
Neural Network - MLPRegressor	

Every model listed in Table 4.1 will be explained in more detail below.

The Power Scaling or Power Law function is presented in Table 4.1. This model is often utilized in scenarios where a nonlinear relationship between two variables is anticipated. For example, revenue may grow at a rate that is proportional to the number of employees.

The Support Vector Regression (SVR) with a Radial Basis Function (RBF) kernel is capable of modeling nonlinear relationships. The general formula for the RBF kernel version of SVR is quite complex and lacks a straightforward explicit form, unlike linear Support Vector Machines (SVM). Conceptually, this method works by mapping the input data into a higher-dimensional space to find a linear separation, and then projecting it back into the original space, resulting in a nonlinear prediction.

$$y = \sum_{i=1}^{N} \alpha_i K(x, x_i) + b$$

Here:

- y: predicted variable (e.g., net income, gross profit, VC raised, or revenue)
- x is the number of employees.
- $K(x, x_i)$ is the kernel function (in this case, the Radial Bias Function (RBF) kernel), which measures similarity between x and the training points x_i.
- α_i are the learned coefficients.
- b is the bias term.

Using the RBF kernel, Support Vector Regressor (SVR) can fit complex, non-linear curves to the data, which allows it to capture the non-linear relationship between employees and revenue.

Polynomial Regression The Polynomial Regression model expands on linear regression by including higher-degree terms:

$$y = a + b_1 \cdot x + b_2 \cdot x^2 + b_3 \cdot x^3 + \dots$$

In this equation:

- y: predicted variable (e.g., net income, gross profit, VC raised, or revenue)
- x is the number of employees.
- a is the intercept (the expected revenue when no employees exist).
- $b_1, b_2, b_3, \dots$ are coefficients for the linear, quadratic, cubic, and higher-order terms, respectively.

This model is useful when the relationship between revenue and employees is nonlinear, and the data suggests the need for curvature or more complex behavior than a simple line can provide.

Linear Interpolation is a simple way to estimate revenue between two known points.

$$y = y_0 + \frac{(x - x_0)(y_1 - y_0)}{x_1 - x_0}$$

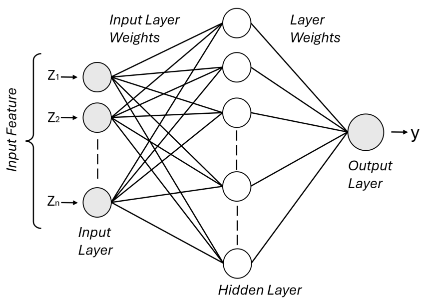
Here:

- y_0 and y_1 are the predicted values (e.g., net income, gross profit, VC raised, or revenue) corresponding to employee numbers x_0 and x_1
- x is the number of employees for which you want to estimate the revenue

Linear interpolation can struggle with data that is not well-behaved, especially when there are multiple data points with the same x or employee value. This method provides a smoothed estimate by using the average revenue for companies with the same number of employees.

It is also possible to use the median or infinitesimal-small shift for multiple occupancy values instead of the average. The experimental usage shows that the simple average has the best R^2 overall.

The Multi-Layer Perceptron Regressor (MLP Regressor) is an artificial neural network used for regression tasks. The general form of the neural network is complex to represent in a simple equation, as it consists of multiple layers of neurons, each connected by weights and biases. Figure 4.1 shows a simplified version of such a neural network.



$$y = f(W_L \cdot f(W_{L-1} \cdot \ldots f(W_1 \cdot \mathbf{x} + b_1) + b_{L-1}) + b_L)$$

where

$$x = [z_1, z_2] = [\text{employees}, \text{revenue}]$$

Figure 4.1: Neural network for MLP regressor with formula and structure, own figure

In this formulation:

- y: predicted output (e.g., net income, gross profit, VC raised, or revenue number)

- f is the activation function applied at each layer (here Rectified Linear Unit (ReLU), (other activation functions had lower R^2).
- W_i represents the weight matrix for layer i, where W_L is the weight matrix of the last layer and W_1 is for the first layer.
- b_i is the bias vector for layer i, where b_L is the bias for the last layer and b_1 is for the first layer.
- $\mathbf{x}$ is the vector of input features, which consists of:
  - z_1: number of employees (training-set)
  - z_2: corresponding revenue (training-set)

The MLPRegressor was trained with all data points for a descriptive functional form, and afterward, the calculated curve was used to determine R^2 with all data points. Fine-tuning options for the MPL-Regressor are:

- Changing the activation functions of the perceptions from "ReLU" to "tanh" or "logistic". The last two functions had worse R^2 than the simple "ReLU".
- Changing the solver from the default Adaptive Moments (adam). This does change the trajectory, but the highest R^2 is faster achievable with an "adam"-solver.

One of the more complex aspects of fine-tuning hyperparameters is determining the optimal number of hidden layers. I decided to test this option, even though it requires substantial computational resources. The optimal number of hidden layers varied for each category of startups. Ultimately, I was unable to identify a globally optimal number of layers within my setup. I calculated the optimum within the range of 1 to 1000, but this limitation could be broadened if additional computational power were available. For further implementation details, I recommend referring to Appendix 8.8.

The MLPRegressor can capture very complex, nonlinear relationships, making it a powerful tool for tasks where the revenue function is difficult to approximate using traditional regression models.

4.3 Robustness and Optimization

Some of the mathematical models provided may tend to overfit, making robustness essential. When a new data point (for example, an employee-revenue tuple) is added to the dataset, the mathematical model should continue functioning effectively.

To achieve this, one can use cross-validation. This method typically divides the dataset into five equally sized buckets of data points. Four buckets are used for training the model, while the fifth bucket is used for validation. The advantage of this method is that one can permute the five buckets, creating five different test folds. I calculated the average R^2 value from these five test results. The procedure is illustrated in Figure 4.2.

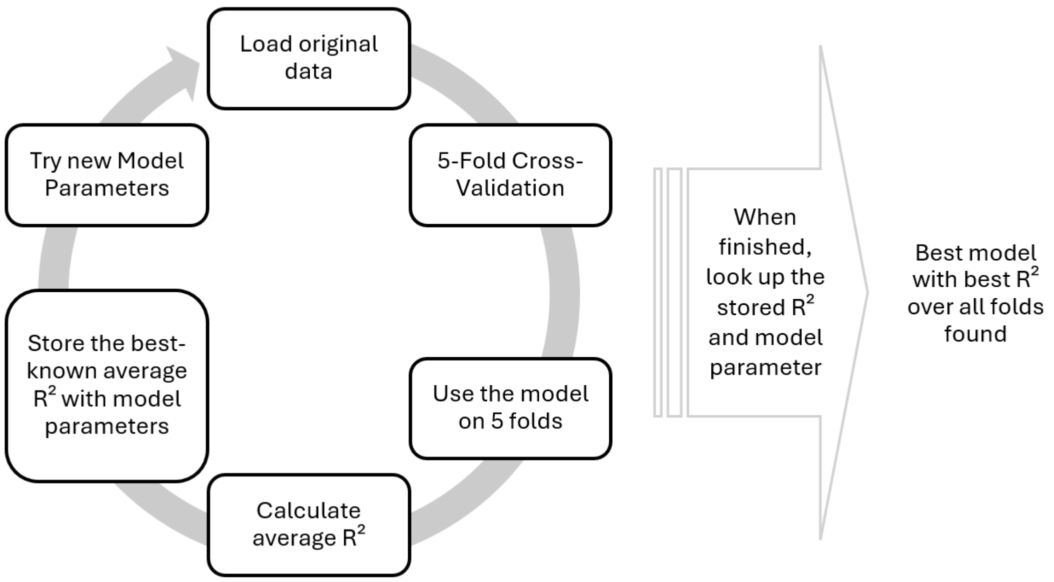


Figure 4.2: Process for achieving robust and optimal results, own figure

The next step is an automated iterative process, where cross-validation is used to optimize the tuning parameters for achieving maximum R^2, shown in Figure 4.2. This includes adjusting the hyperparameters for the MLP regressor, selecting the degree for polynomial regressions, and so on. That way, optimal parameters regarding the cross-validated R^2 are found.

5 Main Analysis and Results

This chapter contains the results of the scaling analysis based on the prepared and validated data using the methodology described in Chapter 4. The results are visualized and described. The performance of models is evaluated for each of the four startup categories for further interpretation and discussion in Chapter 6.

5.1 Drawing the Baseline: West's Scaling Analysis

The next step is to conduct the same analysis using the same variables as West (2019), but this time based on PitchBook data with a focus on the four types of startups.

Before proceeding, I will provide a brief summary of the variables and findings. I am particularly interested in the net income per number of employees and gross profit per number of employees. In the book, these results are presented as follows:

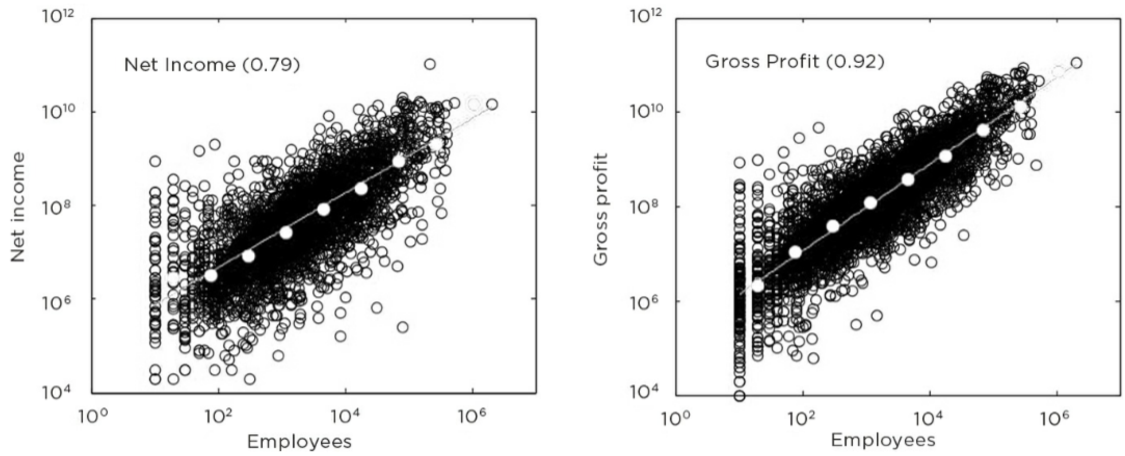


Figure 5.1: Net income and gross profit scaling analysis in USD (West 2019, p. 396)

Figure 5.1 illustrates the income and gross profits of nearly 30,000 companies, plotted logarithmically against their number of employees. These metrics are key indicators of a company's financial health and performance (West 2019, p. 395).

As the plots in Figure 5.1 show, companies tend to scale following a simple power law. In this statistical context, companies can be seen as approximately self-similar, meaning that larger companies are scaled-up versions of smaller ones. For example, Walmart can be considered a larger counterpart to a much smaller, more modest company (West 2019, p. 395).

With these findings in mind, I will explore if the results are comparable across the four startup categories.

5.1.1 Net Income

The first analysis that West performed was to analyze a power law for the dependence of the mean net income of companies on the number of employees. The results of my power law scaling analysis performed on the PitchBook data are shown in Figure 5.2. The net income is in thousand €. Further details are in Appendix 8.4.1.

It is important to note that each plot in Figure 5.2 contains numerous data points that lie far outside the main cluster where most points are concentrated. This issue was previously addressed in Chapter 3. At first glance, the fitted lines (shown in red) may appear unexpected and counterintuitive; however, the dispersion of the data significantly influences the coefficients of the resulting power function.

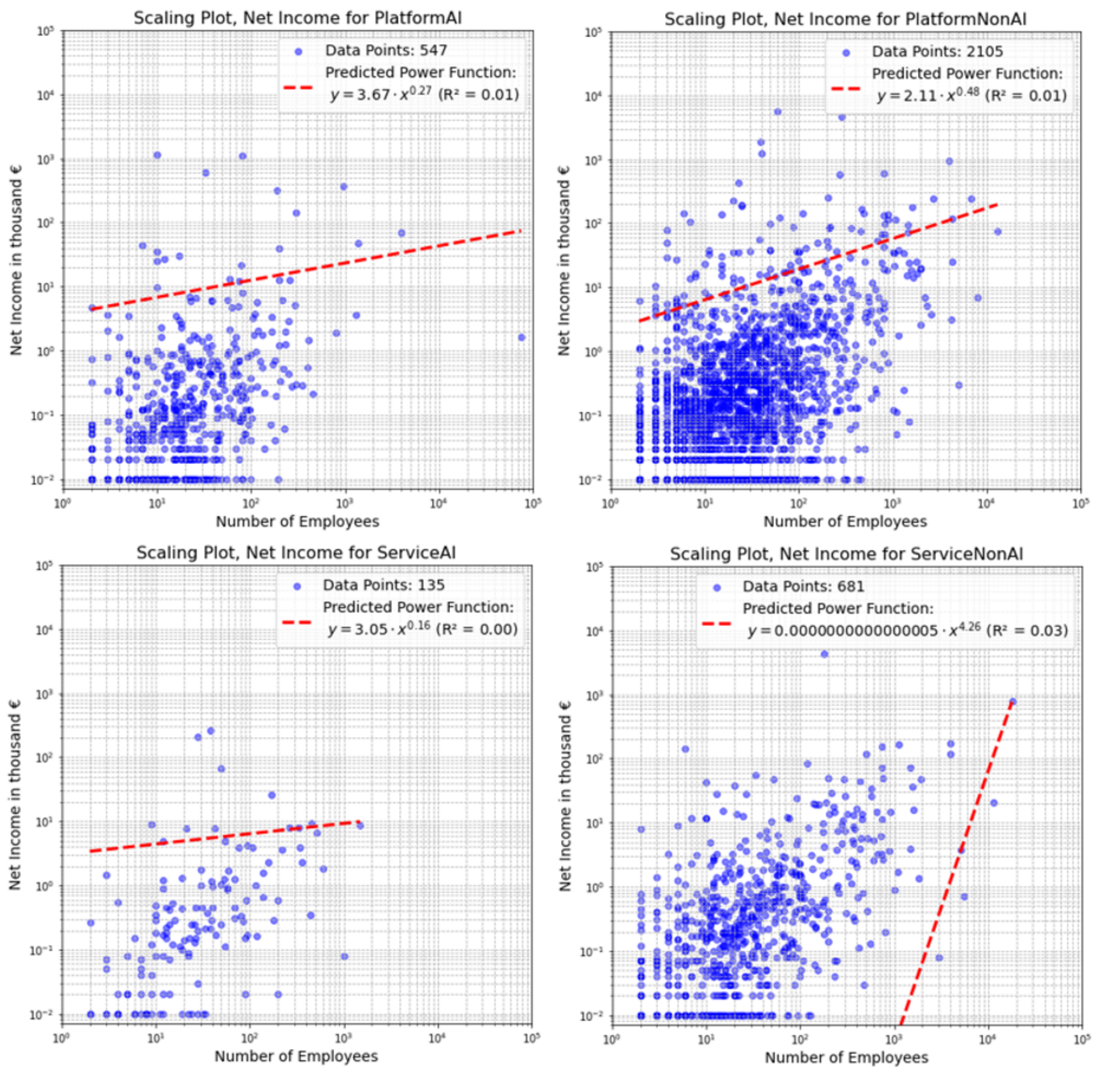


Figure 5.2: Net income based on employees, scaling analysis, own figure

You can see that the R^2 is near zero, which means the model is not a good descriptor for this kind of relationship. This could be because of the startup nature that they are not generating income yet, or the scaling-power-law function is not applicable because many startups are on the way to failing but still counted here. Or the universal law of scaling is only applicable to big companies. Further details are in Appendix 8.4.

5.1.2 Gross Profit

West's second analysis is the power law for the dependence of the mean gross profit on the number of employees. The power law scaling analysis results inspired by West (2019) but performed on the PitchBook data are shown in Figure 5.3.

These plots show a tendency for the gross profit to rise with the number of employees in this log-log plot. It is again important to notice that in every plot in Figure 5.3, there are many data points far outside the main cluster where most data points group at low-profit levels, for details see Chapter 3.2.3.

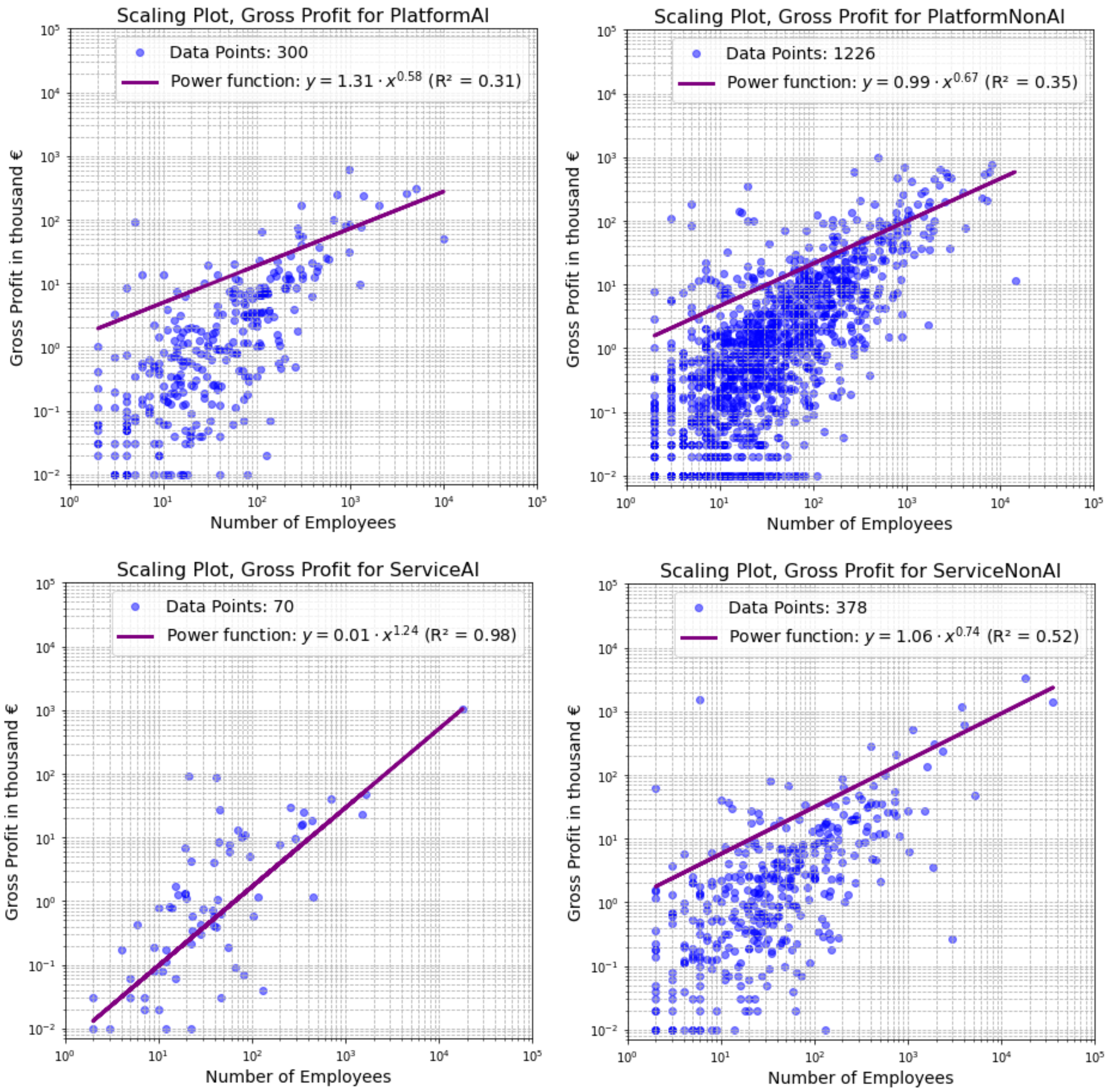


Figure 5.3: Gross profit based on employees - scaling analysis, own figure

In Figure 5.3, we can observe that the R^2 values for service AI are quite strong. However, the other three values do not approach the $R^2 = 0.92$ seen in the graphs from West (2019). One limitation for service AI is the small number of data points, which affects the significance of the analysis. Nevertheless, the exponent for service AI, which is 1.24, is noteworthy because an exponent greater than 1 indicates that the companies are exceptionally innovative, as interpreted by West (2019). Further details are in Appendix 8.4.

5.2 Drawing the Baseline: Schulte-Althoff's Scaling Analysis

Similar to the study by Schulte-Althoff et al. (2021), I will analyze the revenue and venture capital raised variables. These two attributes in PitchBook have much greater data completeness, resulting in more data points for analysis.

5.2.1 Venture Capital Raised - Scaling Anaysis

The power law from the Chapter 4.1 above has been used to explore the dependence of the raised VC on the number of employees in the startups, see Figure 5.4.

The data points in Figure 5.4 are transparent, and when they overlay, the color becomes darker, this is useful as the plots contain many more companies or data points than the previous plots. The VC is in units of thousand €.

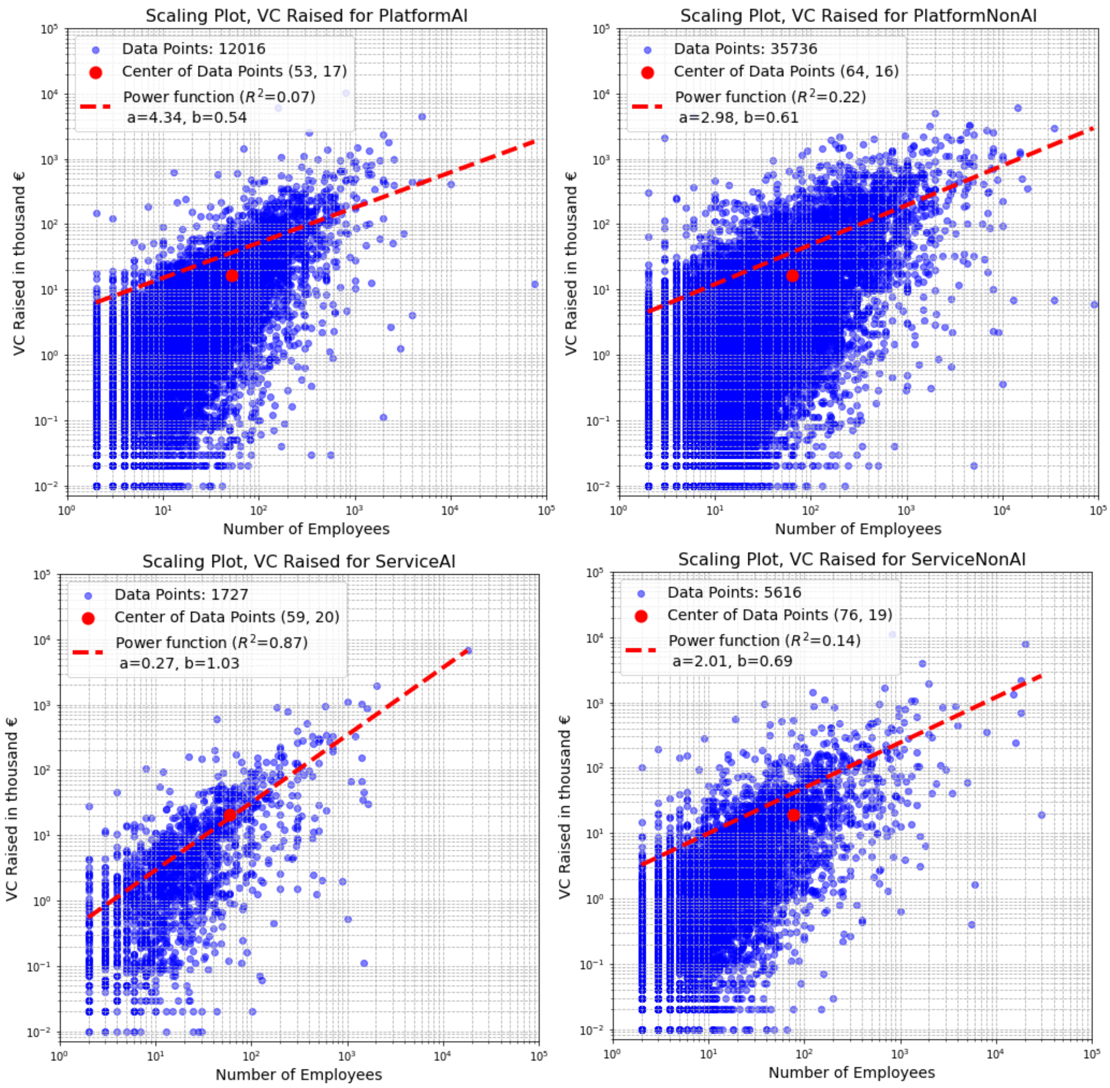


Figure 5.4: VC raised based on employees, scaling analysis, own figure

The graphs indicate that many startups across various categories receive very low funding. However, service AI startups might follow a power law growth pattern, as indicated by a relatively high R^2 value and a significant number of companies in this category. This suggests a notable relationship between the number of employees and the volume of VC funding for service AI startups. In contrast, the analysis shows that the other three categories do not exhibit such a strong relationship.

In this note, I again mention that the fitted power law, represented by the red line in the plots, does not align with the main cluster of the data points. For further details, see Chapter 3.2.3.

5.2.2 Revenue - Scaling Analysis

Figure 5.5 shows the dependence of revenue in units of thousand € on the number of employees in the startups of our data set as scatter plots.

In these plots, we see some tendency towards rising revenue with an increasing number of employees. We fit a power law function again, see the red line in the plots.

The plots in Figure 5.5 contain far fewer companies and thus data points than the plots for VC-raised in Figure 5.4. And again, service AI startups stand out; even though the number of companies analyzed is not as high as for VC raised, the R^2 is still high, and the exponent in the power function is exceptionally high. Also, it should be mentioned that the exponent in the service Non-AI startups is relatively high, even though the R^2 value is small.

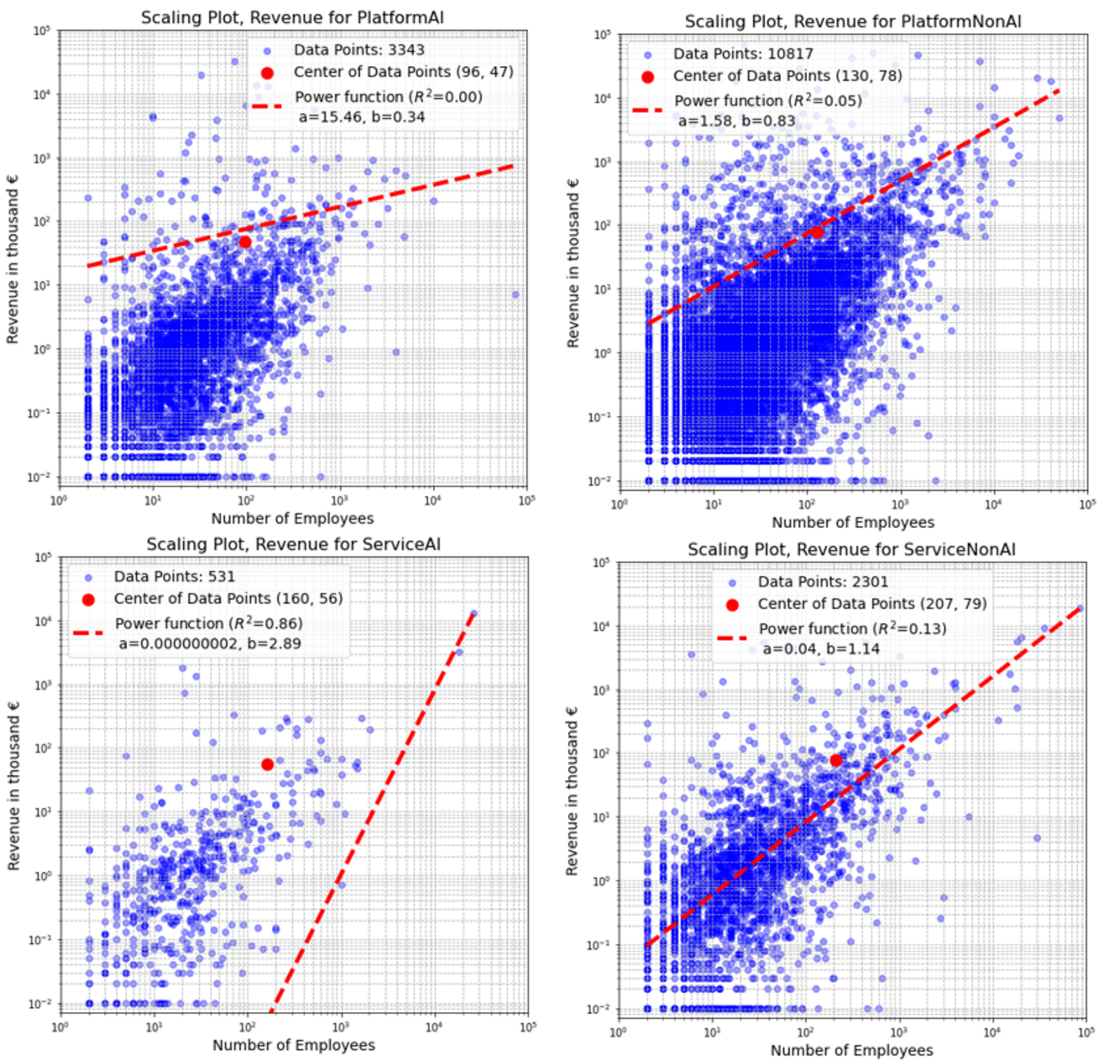


Figure 5.5: Revenue based on scaling analysis, own figure

The service AI function in Figure 5.5 may appear questionable, but fitting includes extremely high revenue achievers; see Chapter 3.2.3.

After analyzing all startup categories, we can conclude that service AI startups follow a power law scaling while other startups do not.

5.3 Investigation of Geographical Data

West (2019) explores in his book "Scale" how scaling laws govern both biological and social systems. He describes scaling as the principle through which certain properties, such as metabolic rates or city infrastructure, change predictably but nonlinear with size. West argues that these scaling laws are universal and can be applied to understand the growth patterns and efficiency of cities. He illustrates that as cities grow, they tend to become more efficient, displaying economies of scale in infrastructure, which can also be applied to organizations and businesses (Giuliano, Kang, & Yuan 2019; West 2019).

Luis Bettencourt takes a slightly different but complementary approach to understanding scaling relationships, particularly in urban environments (Bettencourt, Lobo, Helbing, Kuehnert, & West 2007). Bettencourt's research focuses on the quantitative aspects of urban scaling, demonstrating that cities follow mathematical relationships that can predict various outcomes such as economic productivity, innovation rates, and crime (Bettencourt et al. 2007). According to Bettencourt, these scaling laws reveal that as cities grow, they do not just become larger versions of themselves but also undergo qualitative changes. This transformation allows for increased social interactions and economic opportunities, which behave exponentially (Bettencourt et al. 2007; Giuliano et al. 2019).

The relevance of these theories to startups could be approached with Giuliano's research on agglomeration economies and urban form, which provides insights into how specialization within employment centers can drive economic activity (Giuliano et al. 2019). Startups often thrive in specialized clusters where they can share resources, pool labor, and engage in learning exchanges. These agglomeration benefits are crucial for platform businesses that rely on network effects and service models that require specialized skills. The efficiencies gained from this agglomeration can lower costs and enhance innovation, making scaling more feasible and sustainable.

The application of scaling theories to AI-enabled startups underscores the importance of specialization and innovation. For platform businesses, which depend on user interactions and service models that require targeted expertise, being situated in dense, specialized environments can provide a competitive edge. Giuliano's findings suggest that such environments offer better matching of labor and resources, which is essential for the growth and scalability of startups. Therefore, understanding and leveraging the principles of scaling and specialization in an environment can significantly impact the success and expansion of businesses (Giuliano et al. 2019). Also, the findings of Díaz-Santamaría and Bulchand-Gidumal (2021) underline that effect, mainly because EU AI startups seem to be concentrated in capital cities in countries like France, the UK, and Spain. But one Paper, where authors worked with the census of manufactures data from the USA, suggests that "new industries prosper in large, diverse metropolitan areas, but with maturity, production decentralizes to smaller, more specialized cities" (Henderson, Kuncoro, & Turner 1995, abstract). So, with that in mind, we can say that AI startups should be found more often in metropolitan areas because of their technological novelty.

I analyze startup locations based on different business models. The PitchBook database provides four geographic variables: Region (or Continent), Country, State/Province (or Federal State), and City. For each geographic variable, I conduct a scaling analysis of revenue in relation to the number of employees. This analysis aims to identify the best locations for various startup businesses and their respective business models.

5.3.1 Geographical Data for Platform AI Startups

The innovative use of AI in various platforms is typically associated with traditional industrialized countries like the United States, Japan, Germany, and France. Instead, this super-scaling business model is more commonly found in unexpected countries such as Estonia, Ireland, China, and Finland. Interestingly, 9 out of the 10 cities presented seem to follow a power scaling formula. Notably, well-known innovation hubs like Silicon Valley (which includes San Francisco and San Jose) and Shenzhen in China do not appear in the top 10 list. Instead, two Indian cities, Pune and Gurgaon, both with populations of fewer than 1 million residents, have made it onto the list. Further details are in the Appendix, e.g. Figure 8.25.

5.3.2 Geographical Data for Platform non-AI Startups

Platforms that do not use AI are located differently from those that do. Malaysia and Hungary are utilizing the power law scaling the best for this business model.

Many cities appear to be promising locations for platforms without AI to scale, as they all demonstrate meaningful R^2 values and have an exponent greater than 1. If you are interested in identifying areas where opportunities for this business model are growing, notable locations include Hyderabad, Prague, Telangana, and Minnesota. Smaller regions also show significant potential. While many of these opportunities are found in the USA, two are in India, and one is in Malaysia.

The only region where platform startups, both with and without AI, are experiencing superlinear or power growth is Telangana, an Indian province with a population of 35 million, particularly in its capital city, Hyderabad. It is important to note that platform startups without AI tend to have a larger exponent for employee-to-revenue scaling. Further details are in the Appendix, Figure 8.26.

5.3.3 Geographical Data for Service AI Startups

For Service AI startups, the scaling lists are short. Only two countries are at the forefront of this field: India and the United States. Further investigation becomes complicated, particularly at the city level. It is evident that AI startups in Seoul exhibit a wide variation in the ratio of employees to revenue, similar to those in California. This observation is supported by the R^2 value. More details can be found in the Appendix, specifically in Figure 8.27.

5.3.4 Geographical Data for Service non-AI Startups

In the last category of business models, labeled "Service without AI," certain regions, such as South Korea, demonstrate a power law scaling.

However, some areas stand out in this scaling analysis, particularly San Francisco, which exhibits superlinear scaling. This city is well-known for its robust IT industry and expertise. The dataset indicates that San Francisco has a notably large exponent and a highly descriptive R^2 value. Additionally, Colorado and Texas are other regions where service scaling without AI is occurring. More details can be found in the Appendix, specifically in Figure 8.28.

5.3.5 Summary for Geographical Data

It's interesting to observe that some cities or regions have startups that scale exceptionally well. However, not all locations offer the same level of superlinear scaling. Furthermore, areas that utilize AI tend to differ from those that do not.

Traditionally recognized hubs of the technology industry, such as Silicon Valley, tend to host more non-AI startups. In contrast, AI-enabled startups are primarily found in countries like Estonia, Ireland, China, and Finland, as well as in Indian cities like Pune and Gurgaon. Overall, the USA and India appear to be the strongest countries in terms of employee-to-revenue power scaling in the AI sector.

This analysis underlines the theory that AI startups prefer agglomerations, but different ones than non-AI startups.

5.4 Investigation of Industries

Startups operate across various industries that experience fluctuating growth, making it important to examine these industries. In the dataset, there are four variables related to industries: verticals, primaryIndustryCode, primaryIndustryGroup, and primaryIndustrySector. Since verticals are used to classify different startup business models, it is more beneficial to investigate the other three variables. Table 5.1 provides a brief example to illustrate how industry classification works in PitchBook.

Table 5.1: Industry classifications in PitchBook: Example

Short Description	Verticals	Primary Industry Code	Primary Industry Group	Primary Industry Sector
An AI-enabled platform for HR and talent matching	Human Resources, Technology	541512	Professional, Scientific, and Technical Services	Information Technology

Table 5.1 illustrates a platform for talent matching in the field of Human Resources (HR). The vertical entry represents the most specific categorizations, followed by the "PrimaryIndustryCode," which is a standardized Number that allows global comparability among industries. The next level of detail is the Primary Industry Group, which is followed by the broader Industry Sector, both enable comparability worldwide.

In the following chapters, I will compare AI startups with non-AI startups. This means that AI platforms and AI service startups will be examined together, while non-AI platforms and non-AI service startups will be analyzed together as well. By comparing revenue scaling with venture capital (VC) funding scaling, we can evaluate the scaling efficiency of different industry groups and investigate the relationship between funding growth and revenue growth within those industries.

5.4.1 Investigation of primaryindustrySector

Non-AI startups in the healthcare sector are scaling quite effectively, with an exponent of around 2 ($R^2 = 0.6$). In contrast, an analysis of venture capital scaling in this sector shows an exponent of 0.5 ($R^2 = 0.7$). This difference suggests that these startups are three times more efficient at converting funding growth into revenue growth.

AI startups in the healthcare sector are experiencing growth with an exponent of 0.7 ($R^2 = 0.8$), which is similar to the growth observed in the financial services sector. However, in the healthcare AI startup sector, the exponent for VC funding is 0.5. This indicates that revenue scaling and VC funding scaling are occurring at nearly the same rate. In my opinion, this suggests that the healthcare industry is in a "healthy" state, characterized by balanced growth. You can inspect the tables in Appendix 8.7, where there are more details on this topic.

5.4.2 Investigation of PrimaryindustryGroup

First, we will analyze non-AI startups, followed by an analysis of AI startups.

PrimaryindustryGroup non-AI Startups Investors are directing funding toward the "Communications and Networking" sector; however, there is a significant variance in revenue outcomes, as indicated by the R^2 value. This suggests that a considerable portion of the funding is going to startups that are generating poor revenue.

In contrast, the "Computer Hardware" group has experienced lower revenue growth, with funding scaling at an exponent of 0.8 compared to revenue scaling at 1.2. This indicates that this group has received less funding relative to its potential for revenue growth. Nevertheless, the group has utilized the funding it did receive effectively.

In the "Healthcare"-related non-AI startup group, the funding and revenue scaling are aligned, meaning that growth in both funding and revenue, in relation to the number of employees, is proportional. This reflects a balanced growth strategy.

PrimaryindustryGroup AI Startups For AI startups, the leading groups for revenue growth are "Healthcare Technology Systems," "Media," and "Other Financial Services," all of which have an R^2 value above 0.7, though their exponents remain below 1. When comparing this growth to the increase in venture capital (VC) funding, distinct patterns emerge: "Media" (exponent = 2.0), "Commercial Product" (exponent = 2.3), and "Healthcare" (exponent = 0.5) demonstrate varying trends in VC funding growth. Notably, "Healthcare" is the only sector where revenue growth aligns with VC funding. In contrast, the other sectors show that revenue growth is underwhelming, suggesting that there may be some hype surrounding these investments.

5.4.3 Investigation of primaryindustryCode

First, we will analyze non-AI startups, followed by an analysis of AI startups.

PrimaryindustryCode non-AI Startups: For non-AI startups in the "IT Consulting and Outsourcing" code, the revenue growth exponent is 1.0, while the VC funding exponent is 1.6. This indicates that the industry experiences more growth in funding than in revenue, which could suggest that there is some hype surrounding this industry or code.

Conversely, in the "Electronic Equipment and Instruments" sector, the revenue growth exponent is 1.3, but the funding growth exponent is lower at 0.8. This suggests that revenue growth outpaces funding growth, indicating an extremely effective utilization of the available funding.

A notable industry code is "Database Software," which has a revenue growth exponent of 1.4. However, it lacks a strong R^2 value for VC funding growth, meaning while companies in this field demonstrate a revenue-to-employee power function, VC funding does not. So investors do not seem to align the amount of funding with the number of employees.

In contrast, other industry codes with high R^2 values show a clear alignment between funding and revenue scaling. Further details are in the Appendix 8.7.

PrimaryindustryCode AI Startups: Overall, for AI startups, the "Healthcare Technology Systems" category appears to scale at a rate that closely matches the growth rate of VC funding. In contrast, the "Media" sector shows a funding exponent of 2.0 and a revenue exponent of 0.8, highlighting a significant discrepancy. This could suggest that there is some hype, as the scaling on the VC side is occurring at a higher exponent than revenue.

Additionally, the "Commercial Products" category features a high VC funding exponent of 2.3, but demonstrates considerable variability in revenue, as indicated by the low R^2 value. This suggests that while investors base their funding decisions on the number of employees, the resulting revenue growth remains uncertain.

5.4.4 Summary: Industries Results

In summary, after comparing the exponents of power functions for VC funding and revenue growth, you can categorize the industries into three groups:

- The "hyped" industries, so industries where the funding is scaling higher than the revenue is scaling: *"Communications and Networking" group for Non-AI startups, "Media" sector for AI startups, "Commercial Product" sector for AI startups, "IT Consulting and Outsourcing" code for non-AI startups, "Media" code for AI startups, "Commercial Products" Code for AI startups.*

- The "hidden gem" industries that use the funding extremely well, which means the revenue is scaling higher than the funding is scaling: *"Healthcare"sector Non-AI startups, "Computer Hardware" group Non-AI startups, "Electronic Equipment and Instruments" codes for non-AI startups.*

- And the "healthy" industries that scale the funding and revenue at nearly the same exponent or have balanced growth: *"Healthcare" sector and group for AI , "Healthcare"group for non-AI startups, "Electronic Equipment and Instruments" codes for non-AI startups, "Internet Retail " codes for non-AI startups, "Publishing " Codes for non-AI startups, "Education and Training Services B2B " codes for non-AI startups, "Other Healthcare Technology*

Systems " codes for non-AI startups, "Healthcare Technology Systems" codes for AI startups.

5.5 Universal Scaling Findings

The sub-analysis reveals interesting patterns in how location and industry affect scaling behaviors. However, for a comprehensive analysis, it is essential to understand the general trend in business model scaling.

The scaling analysis, guided by the work of West (2019) and Schulte-Althoff et al. (2021), examined variables such as gross profit, revenue, and venture capital. The results indicate that service AI startups follow a "universal" scaling pattern. However, the model that assessed net income relative to the number of employees yielded a low R^2 value, suggesting that this particular analysis did not produce meaningful insights.

In contrast, no "universal" scaling pattern was found for platform AI startups.

5.5.1 Overview Not Robust Models for Revenue

The main focus of this master thesis is on the revenue variable because the revenue is the basis for the net income and profit of a company. That is the reason why in the provided Table 5.3, I have collected all the R^2 values based on startup type and mathematical model to find the best descriptive mathematical model for each startup category.

Table 5.3: Comparison of revenue based on employees, various models with R^2 scores, not robust

Not Robust Model/ R^2	**Platform AI**	**Platform Non-AI**	**Service AI**	**Service Non-AI**
Power Function	0.00	0.05	0.86	0.13
Support Vector Regressor	0.00	0.02	0.28	0.04
Polynomial Regression	0.00	0.06	0.87	0.14
Linear Interpolation	0.12	0.32	0.89	0.96
MLP Regressor	0.00	0.05	0.76	0.14

You can see in Table 5.3 that for the platform AI business type, it is relatively hard to find a good fitting model. But for service AI startups-dataset there are a couple of good fitting models. So basically for the platform AI dataset, it is recommended to use the linear interpolation with an average for duplicate values, for Platform Non-AI, the linear interpolation is the model of choice. You can choose from three relatively strong models for the Service AI dataset: the power function, polynomial

Regression (degree 4), or linear interpolation. Only the linear interpolation has a good R^2 for service Non-AI.

It is essential to say that this model did not have a robustness check, and especially for linear interpolation, you can see in the graphs that the function looks fidgety or overfitting (see Appendix 8.9.4). Overfitting means the function is too near the data points and does not have a generalization or predictive character. For example, the R^2 value is highly sensitive to extreme outliers, so if, hypothetically, a new company is added with 1 employee generating millions of revenue to one business model category, the power function would be optimized extremely for that value. And for that reason, a robustness-check should be done.

5.6 Scaling Factor with Robust Power Function

In this chapter, I am doing a robust revenue scaling analysis. This means I am investigating the power function with predictive intentions for each startup category.

5.6.1 Robust Revenue-to-Employee Power Scaling

Since the power-scaling analysis is the main focus of this work, this chapter presents the plots for various models to support the findings visually. The robust power-scaling analysis for startups from PitchBook, based on employee count and revenue, looks as follows:

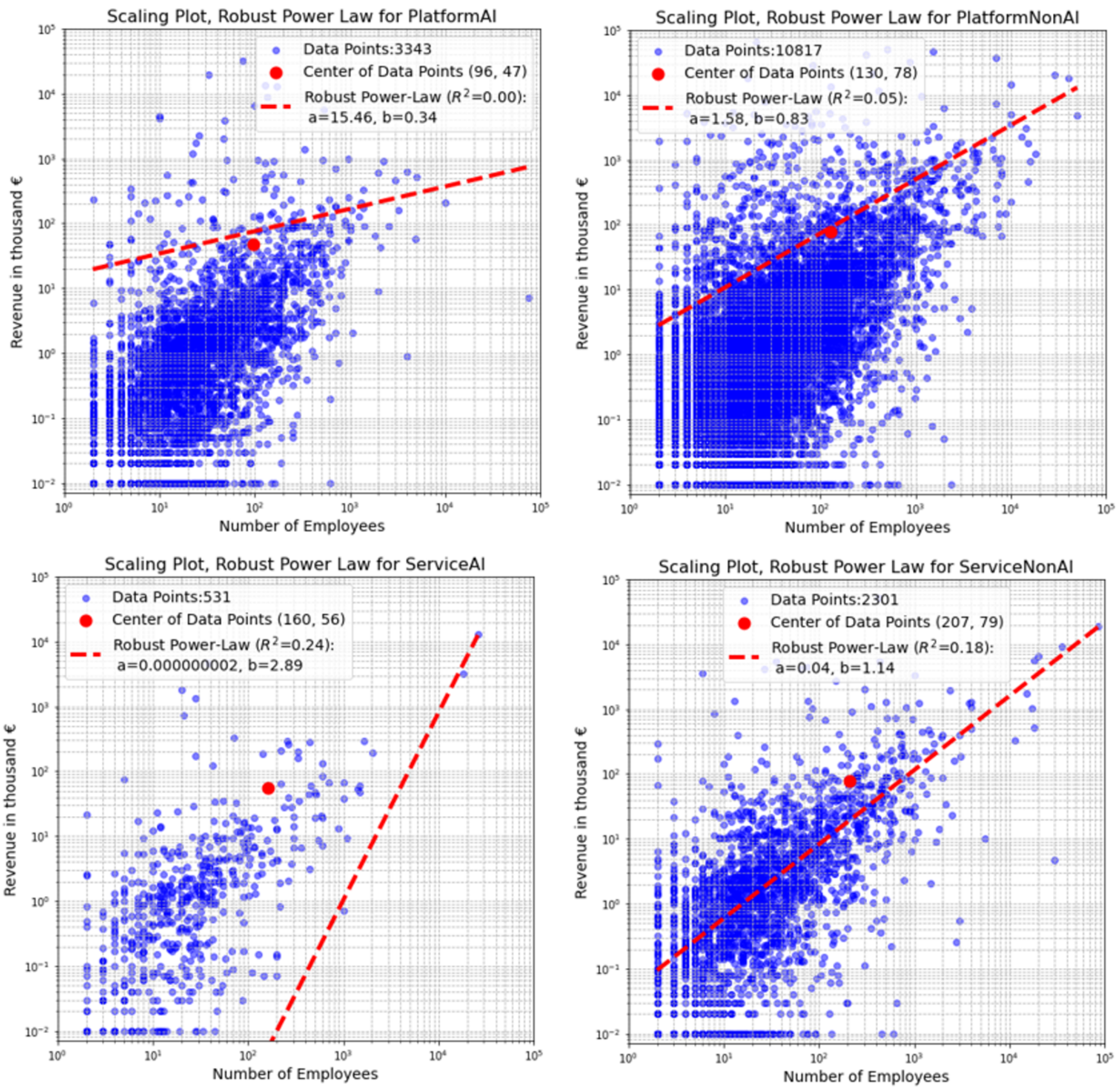


Figure 5.6: Robust power scaling for revenue based on employees. Here extreme values are not plotted. For further details see Appendix 8.9, own figure

The service AI function in Figure 5.6 may look suspicious. However, this is, as discussed already, because the function was optimized on the data, which contains very high revenue achievers. Further details are in Appendix Figure 8.9. The curves presented in the plot are the average revenue based on the power function of five test folds.

You can see the robust power function has generally turned out to have low R^2 for each category.

The robust exponent for service AI startups is 2.89, while for service non-AI startups, it is 1.14. This indicates that the two differ by more than a factor of two. This trend is also evident in the insufficient power function for revenue, as illustrated in Figure 5.5. Additionally, the robust scaling analysis for gross profit shows an exponent of 1.24 (R^2 = 0.18) for service AI startups and 0.74 (R^2 = 0.33) for service non-AI startups, resulting in a difference of 0.6, as detailed in Appendix Figure 8.43.

5.6.2 Using the Knowledge of Other Models

In this chapter, I explain that specific models yield contradictory results when the number of employees falls below a specific threshold. The area of contradictory curves is marked red.

For instance, the MLP Regressor shows a vertical line for employee numbers below 70. At the same time, the polynomial regression yields a nearly horizontal line for a similar range of employee counts; see Figure 5.7.

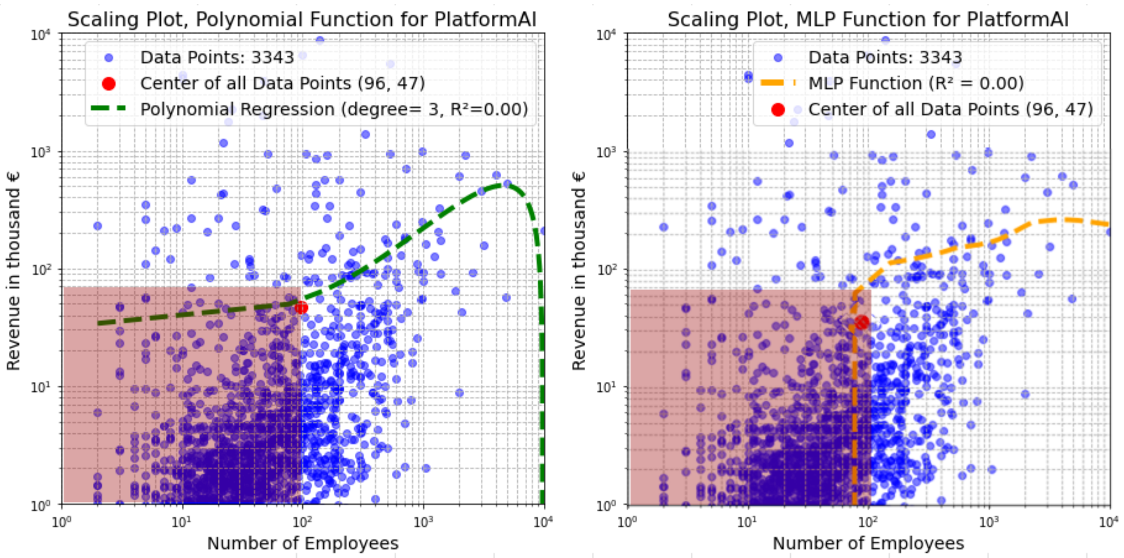


Figure 5.7: Platform AI: Polynomial regression and MLP-regression, own figure

The same situation occurs with platform non-AI startups, which have a limit of around 15 employees, see Figure 5.8

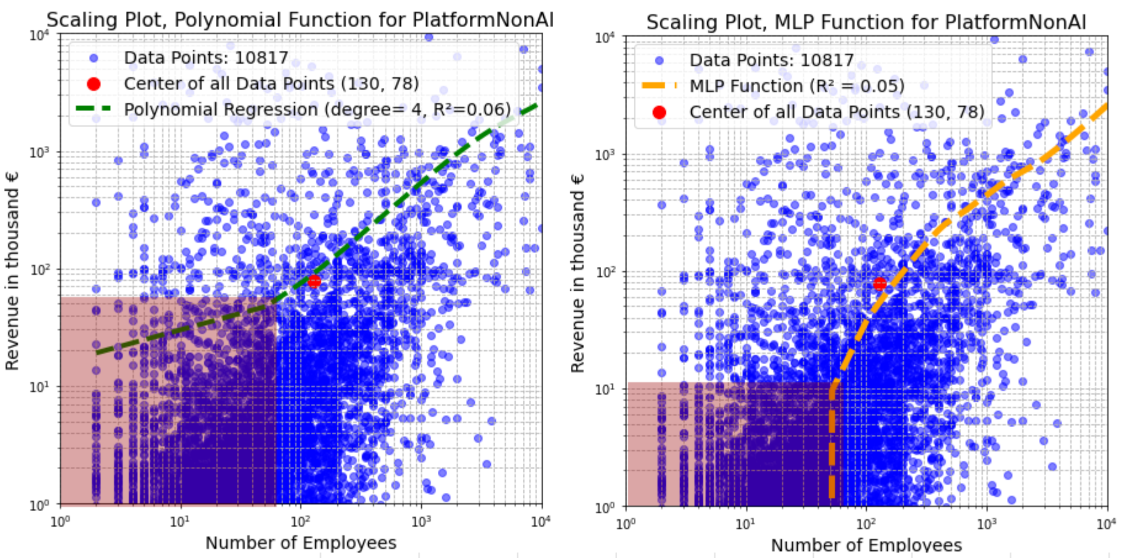


Figure 5.8: Platform non-AI: Polynomial regression and MLP-regression, own figure

Also, for service non-AI startups with nearly 80 employees, contradictory curves are calculated; see Figure 5.9.

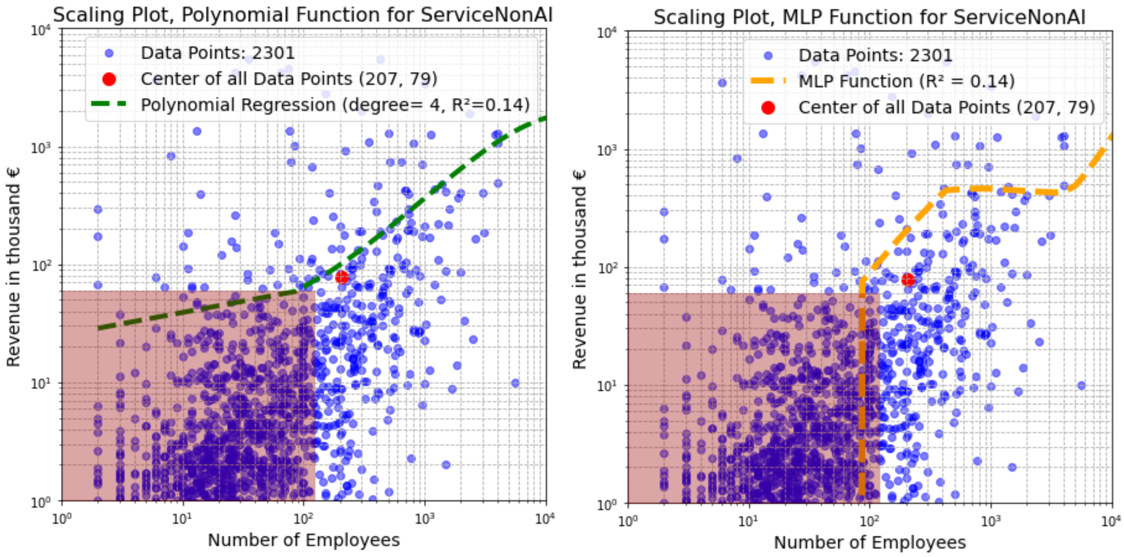


Figure 5.9: Service non-AI: Polynomial regression and MLP-regression, own figure

The service AI curves are attractive because they show a noticeable bend in two models with around 100 employees; see Figure 5.10.

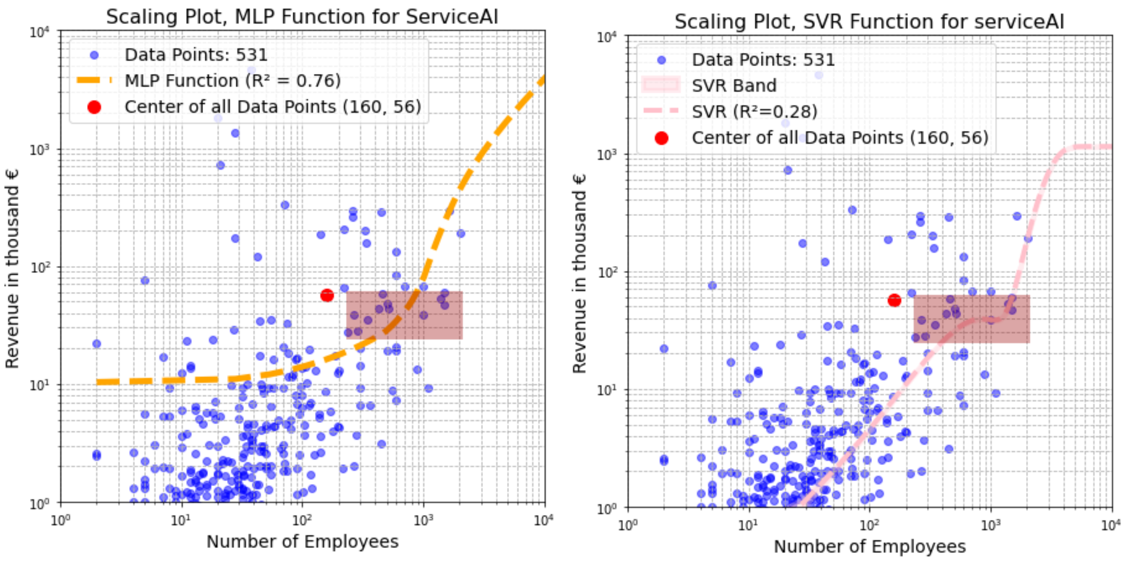


Figure 5.10: Service AI: SV-regression vs MLP-regression, own figure

I decided to separate the employee axis for each startup type according to the upper limit of the anomaly, as shown in Table 5.4.

Table 5.4: Data split based on employee, defined thresholds

Category	Below Threshold	Above Threshold
Platform AI	Count below 70	Count above 70
Platform Non-AI	Count below 15	Count above 15
Service AI	Count below 100	Count above 100
Service Non-AI	Count below 80	Count above 80

I am applying the robust revenue power-scaling analysis again based on this threshold to analyze if the power-scaling model also has problems in the marked areas from the previously presented plots. See Table 5.5.

Table 5.5: Data split based in employee, threshold based power-scaling-analysis

Robust Power Scaling Model	Below Threshold R^2	Above Threshold R^2
Platform AI (Threshold = 70)	0.00	(0.00 old) 0.00
Platform Non-AI (Threshold = 15)	0.00	(0.05 old) 0.05
Service AI (Threshold = 100)	0.00	(0.24 old) 0.24
Service Non-AI (Threshold = 80)	0.00	(0.18 old) 0.18

In the power scaling analysis, the R^2 value below the specified threshold is 0.00. Table 5.5 includes the previous R^2 value, which was recorded before the axis split, for comparison.

This allowed me to determine the limits of the models for each category: Platform AI with 70 employees, platform non-AI with 15 employees, service AI with 100 employees, and service non-AI with 80 employees. Below each of those numbers of employees, the power scaling model has limited meaning. Further plots with different functional forms are in the Appendix 8.9.

5.6.3 Removing the Top 1 % Performer

After experimenting with the filter, I decided to exclude the top 1% of revenue-generating startups from each category, as they had the most significant impact on the robust R^2.

Platform AI: By removing the top 1% of revenue achievers from the platform's AI data points, the robust R^2 value has increased from 0 to 0.13, and the exponent has also risen. This is the first model that has an R^2 greater than zero for this category.

Platform Non-AI : In contrast, when examining the data for platform startups without AI and without the 1%, the R^2 value also increases; however, the exponent remains below one.

For **Service AI** startups, the top 1% in terms of revenue exhibit a noteworthy trend. When including these top earners, the exponent exceeds the highly innovative mark of 1. However, once we exclude the top 1%, the remaining 99% of startups do not demonstrate this trend, and the exponent falls below 1.

Service Non-AI : In the service startups that do not use AI, the removal had a different effect; the mean R^2 value worsened.

5.7 Evaluation of Results

I performed a straightforward power function analysis on various variables, identified specific limits for a dedicated model in each category, and conducted optimizations, which excluded certain data points.

This chapter provides a comprehensive overview of the analysis conducted and meaningfully interprets the scaling model's results.

5.7.1 Overview Descriptive Power Scaling Analysis

First, the power function for each startup category and selected variables is calculated to answer whether any startup business model category can be described using the power scaling analysis.

Table 5.6: Financial variable metrics across different categories summarizing if power scaling was identified

Financial Variable	**Service AI**	**Service Non-AI**	**Platform AI**	**Platform Non-AI**
Net Income	no (R^2=0.00)	no (R^2=0.01)	no (R^2=0.01)	no (R^2=0.03)
Gross Profit	yes (R^2=0.98)	yes (R^2=0.52)	somehow (R^2=0.31)	somehow (R^3=0.31)
Revenue	yes (R^2=0.86)	no (R^2=0.13)	no (R^2=0.00)	no (R^2=0.05)
VC Raised	yes (R^2=0.87)	no (R^2=0.14)	no (R^2=0.07)	no (R^2=0.21)
Summary	**Yes**	**No**	**No**	**No**

In summary, only service-oriented AI startups are experiencing power scaling. The platform startups, in both AI and non-AI categories, are not achieving this kind of scaling. When it comes to non-AI service startups, the results regarding revenue scaling are mixed.

5.7.2 Overview Predictive or Robust Revenue Scaling Models

This section presents an overview of robust employee-to-revenue models and associated R^2. To make the mathematical functions more reliable and robust, I used the 5-fold cross-validation described in the methodology Chapter 4 and also presented a robust version of the same function, which excluded the 1% top revenue achievers.

Table 5.7: Overview R^2 values of different models across various categories

Robust Revenue Models/R^2	**Platform AI**	**Platform Non-AI**	**Service AI**	**Service Non-AI**
Power Function	0.00	0.05	0.24	0.18
Support Vector Regressor	0.00	0.04	0.13	0.25
Polynomial Regression	0.00	0.10	0.14	0.30
Linear Interpolation	0.00	0.04	0.15	0.01
MLP Regressor	0.00	0.06	0.13	0.33
P. Func. (without top 1%)	0.13	0.19	0.15	0.17
SVR (without top 1%)	0.29	0.15	0.10	0.18
Pol. Reg (without top 1%)	0.18	0.19	0.14	0.12
Lin. Interp. (without top 1%)	0.03	0.00	0.24	0.02
MLP Reg. (without top 1%)	0.12	0.11	0.18	0.12

The R^2 values presented are derived from cross-validation, where various versions of each algorithm are tested to identify the highest achieved R^2, displayed in Table 5.7. For a more detailed overview, please refer to Appendix Table 8.10, which includes information on the degrees of polynomial regression or the hidden layer sizes for MLP regression.

The best predictive models for each startup category, based on all data points, are highlighted in red. The other rows represent the same cross-validated functions, excluding the top 1% of revenue achievers, clearly indicating that this 1% significantly impacts the results. The best model for the remaining 99% of companies is highlighted in blue.

Table 5.7 serves as the main result of the revenue variable, summarizing the optimized and robust versions of various mathematical models for each startup category alongside their respective data points. Depending on the specific business model of interest, you can easily identify the strongest revenue model.

Compared to the study by Schulte-Althoff et al. (2021), which reported R^2 values of 0.05 for Non-AI platforms (this study shows 0.19, excluding the top 1%), 0.23 for Non-AI services (this study shows 0.24, excluding the top 1%), and 0.12 for AI startups (this study shows 0.33 and 0.24 in both cases), the robust R^2 values obtained in this analysis, indicate a small improvement in the prediction of scaling for each startup business model.

Further, if we look into the specific models, the polynomial regression for service non-AI startups has a relatively strong R^2, and this model is a polynomial of degree 1, so it is a straight line. Based on that, you can vaguely assume that service non-AI startups scale with a linear function compared to service AI startups, for which scaling is more in line with a power function form.

5.7.3 Power-Scaling Model Implications

Overall, the models used in the analysis demonstrate limited predictive performance, but they provide strong descriptive insights. While the relationship between the variables of revenue and number of employees can indicate trends within the data, these two variables alone cannot pinpoint the specific challenges and causes that affect a particular revenue metric in a startup. Revenue is influenced by various factors that are not explicitly considered in the models. As a result, the models' ability to make predictable statements remains constrained.

Nevertheless, the not robust power model, which has a relatively high R^2 value, demonstrates a good fit for describing the data. It effectively captures the overall trend within the data points and summarizes it well, even though it does not explain the underlying causes.

5.8 Answering the Research Question

The research question, *"Does the integration of AI in core business models influence the scaling efficiency of (1) service startups and (2) platform startups?"* can now be addressed. I am employing scaling analysis, particularly the power scaling function, as my primary method to explore this question.

Service Startups: To answer this question, I consider this topic complex, and you need to look at all analyzed variables. In Chapter 5.5, it is shown that service AI startups follow the "universal" scaling based on gross profit, revenue, and VC raised.

I'm using the robust revenue-scaling power-function exponent to quantify the scaling intensity or difference in exponential growth between service startups with and without AI. Service AI startups have an exponent of 2.89, and service non-AI startups have an exponent of 1.14. Even if the non-robust revenue-scaling exponent is compared to each other, the service AI exponent is the same height. Both analyses show a difference in exponents by a factor of 2.53. The robust gross profit scaling analysis supports, with far fewer data points, though, the finding that there is a difference in scaling up to 60% times higher with AI.

In Geoffrey West's book, the non-robust scaling analysis for gross profit in dependence on the number of employees focuses on the exponent. In this case, both startup types have a bigger exponent, which means this company scales superlinear, this indicates a highly innovative venture.

However, in one sub-analysis, I found a mark below where many mathematical models struggle to find a reasonable functional form. Below that mark, unfortunately, the power scaling function also had problems. For both service startup types, the employee mark below that model could not present a good fit, which is 80 employees for service non-AI and 100 employees for service AI. That means below that mark, the meaningfulness of the scaling efficiency should be seen as skeptical. So, to summarize the findings:

Research question 1: Does AI in the core of business models influence the scaling efficiency of service startups?
Yes, when comparing companies with more than 80 employees (for those without AI business models) and more than 100 employees (for those with AI business models), service AI startups have a scaling exponent that is *0.60 to 2.53 times higher* than that of non-AI service startups.
The scaling analyses with revenue, gross profit, and VC raised as variables support this finding: AI in service startups brings higher scalability or exponential growth. However, these findings do not have a robust predictive power but are more descriptive.

One analysis revealed that when the top 1% of revenue achievers are excluded, the phenomenon of super scaling is no longer applicable. This indicates that for 99% of service AI startups, super linear scaling does not hold significance. However, it's important to note that service AI startups are typically founded later than non-AI startups. In my opinion, this difference may be overlooked, as AI companies are generally younger compared to their non-AI counterparts. This finding should not be used for predictive purposes, as cross-validation indicated that this revenue-based approach is not a strong model.

Platform Startups: If all data points of all companies are used for this scaling revenue analysis, the R^2 value would be zero, so no good answer could be given with that method. Therefore, it is beneficial to remove the best 1% revenue-performer startups in this category to gain a higher R^2 than zero in the power scaling analysis.

To answer the research question, does AI in platform startups make a difference in revenue? Based on the power law exponents for platform AI of 66 and platform non-AI of 64, I would tend to answer this question with no. For the non-robust scaling analysis, the exponents are 0.35 and 0.83. However, this business model growth still does not seem to be well described by the power function.

The book by Geoffrey West focuses on the exponent. In this case, both startup types have a lower exponent, which means those companies scale sublinear. This indicates no innovative ventures.

Based on one sub-analysis where a threshold of employees was found, other mathematical functions had struggled to lay a function over the data points. Unfortunately, the power scaling function had problems below that mark as well. I would say that this scaling exponent and its implications are especially not meaningful for non-AI startups with less than 15 employees and for AI platform startups with less than 70 employees. To summarize the findings:

Research question 2: Does AI in core business models influence platform startups' scaling efficiency?
No. With my methods, no answer could be given below the marks of 15 (for those without AI business models) or 70 employees (for those with AI business models). Above those marks, with the removal of the top 1% revenue achievers, a "No" would be the answer. The scaling analysis with other variables like net income, gross profit, and VC raised also supports this finding by showing that these two categories of startups do not follow the "universal" scaling pattern. However, these findings do not have a strong predictive power; rather, they are more descriptive.

5.9 Interpretation of Results

I am following the interpretation by West (2019). He analyses the scaling exponent this way: a company is driven by income and expenses rising exponentially as more employees work for the company. He found that the expenses start sublinear and later behave with more employees in linear form, so exponent <= 1. But the income, which depends on revenue, is either superlinear (exponent > 1) or sublinear (exponent < 1). If the sublinear income growth continues for too long, the company "dies".

For the gross profit to employees scaling analysis, which should indicate whether companies survive in the future, the service AI with 1.24 ($R^2 = 0.98$) seems to have good perspectives, but platform startups (both R^2 near 0.33) with exponents near 0.67 or smaller have a surviving problem, just like service non-AI startups with an exponent of 0.53 ($R^2 = 0.52$).

The promising revenue of service AI startups, an indicator for income, with an exponent of 3.46 (R^2=0.86), could also underline the finding that service AI startups thrive. A bit contradictory is the revenue of non-AI services, with an exponent of 1.14 ($R^2 = 0.13$), indicating superlinear scaling. However, the net income and VC funding raised appear to show sublinear growth.

Unfortunately, platform startups, either with or without AI, do not seem to scale like the other two categories and even seem threatened by economic "death." However, this result has its weaknesses, which are explained in the limitations Chapter 6.3.

5.9.1 Interpretation of Results for Geographical Data

Alongside the prior interpretation of the results, looking at what the results mean to the locations is interesting. Since the overall tendency for platform non-AI and platform AI startups worldwide and industry-wide is to fail, some cities and provinces are scaling-wise outstanding given this background:

Table 5.8: Sub-analyses: Geo-locations for startups based on business model (BM)

BM type	Good Cities	Good Provinces or Countries	Worldwide trend for this BM
Platform AI	Pune, Sao Paolo, Los Angeles, Tallinn, Dublin, Hyderabad, Tokyo, Gurgaon, Helsinki	Telangana, Illinois, Estonia, Ireland, Japan	Threatened by economic ”death”
Platform non-AI	Hyderabad, Petaling Jaya, Turin, Prague, Mexico City, Jakarta, Noida, Palo Alto, Austin, Pune	Michigan, Quebec, Telangana, Selnagor, Nevada, Uttar Pradesh, Minnesota, Malaysia, Hungary, Austria, Switzerland, Thailand, Mexico	Threatened by economic ”death”
Service AI	-	India, USA	Thriving
Service non-AI	San Francisco	Colorado, Texas, South Korea	Thriving

Table 5.8 shows that besides the trend for startups to fail with some business models, agglomerations on this planet still cultivate scaling, successful startups. That analysis supports the theory that a bigger agglomeration of people, like cities, supports innovative and value creation progress of young companies (Giuliano et al. 2019), in this case, presented from the perspective of new technology like AI.

On the other hand, not all analyzed cities and startups have nearly the same R^2 in the scaling functions, despite the height of the exponent, which means that the ”universal” law of scaling is not applicable everywhere. There are also cities where the R^2 is so low that this power-law function has no meaning. This, in my opinion, undermines the ”universal” character of the scaling law from West (2019). It demonstrates that there are cities with startups following this scaling law and cities with startups that do not. Based on the data, exponential growth change is impossible for startups in every town.

5.9.2 Interpretation of Results for Industries

Even though the overall revenue-to-employee scaling analysis had a relatively low R^2, this sub-analysis was done with R^2 above 0.7 because this mathematical model could more precisely describe some industries. If the exponent of the revenue scaling analysis and the exponent of the VC raised scaling analysis are compared, you can have three sets: one with higher growth in VC raised than revenue, I call them "hyped" industries, one with higher growth in revenue then VC raised, I call them "hidden gem" industries and "healthy" industries, where both growths are the same:

Table 5.9: Summary of industry scaling types, based on startups with and without AI

"Hyped" Industries	**"Hidden Gem" Industries**	**"Healthy" Industries**
Communications and Networking (Non-AI)	Healthcare (Non-AI)	Healthcare (AI and Non-AI)
IT Consulting and Outsourcing (Non-AI)	Computer Hardware (Non-AI)	Electronic Equipment and Instruments (Non-AI)
Commercial Products (AI)	Electronic Equipment and Instruments (Non-AI)	Other Healthcare Technology Systems (Non-AI)
Media (AI)		Internet Retail (Non-AI)
		Education and Training Services B2B (Non-AI)
		Publishing (Non-AI)

You can see that no AI-enabled startups are present in the "hidden gem" industry set. But "Media" and "Commercial Products" are hyped within the AI business models. In the "healthy" industry set, only AI startups are present in "Healthcare".

So for AI investors, it is interesting to look into this set: "Healthcare" sector and group and " Healthcare Technology Systems" Codes for AI startups; they have an exponent of 0.6-0.7 for the Sector and Group (both R^2= 0.8), and especially the exponent of 0.8 in the primary industry Code (R^2=0.98). Further details can be explored in the Appendix 8.7.

6 Discussion

6.1 Summary of Results

After applying the findings made by West (2019) and part of the findings of Schulte-Althoff et al. (2021) to the Pitchbook dataset, the two research questions were answered as follows:

- **Research question 1: Does AI in the core of business models influence the scaling efficiency of service startups?** Yes, when comparing companies with more than 80 employees (for those without AI business models) and more than 100 employees (for those with AI business models), service AI startups have a scaling exponent that is *0.60 to 2.53 times higher* than that of non-AI service startups. The scaling analyses with revenue, gross profit, and VC raised as variables support this finding: AI in service startups brings higher scalability or exponential growth. However, these findings do not have a robust predictive power but are more descriptive.
- **Research question 2: Does AI in the core of business models influence the scaling efficiency of platform startups?** No. With my methods, no answer could be given below the marks of 15 (for those without AI business models) or 70 employees (for those with AI business models). Above those marks, with the removal of the top 1% revenue achievers, a "No" would be the answer. The scaling analysis with other variables like net income, gross profit, and VC raised also supports this finding by showing that these two categories of startups do not follow the "universal" scaling pattern. However, these findings do not have a strong predictive power; rather, they are more descriptive.

6.1.1 Comparison with Existing Literature

The findings of this study expand existing knowledge on scaling in startups, especially in the domain of AI-enabled service businesses. However, it is important to note that the literature provides very few quantitative studies on AI-enabled scaling. For example, studies that directly and quantitatively compare service startups to AI-enabled service startups do not exist yet. Additionally, there are no existing quantitative studies specifically addressing the scaling differences between platform startups and AI-enabled platform startups.

One of the few studies approaching this matter was written by Schulte-Althoff et al. (2021), where the authors group service AI and platform AI startups under a single category: AI startups. My thesis builds on this existing work by expanding this approach and making further distinctions. Specifically, my analysis reveals that platform AI startups do not exhibit superlinear scaling behavior or follow a power-law growth as expected. This finding may explain why the power law function of AI startups in Schulte-Althoff et al. (2021) has relatively small R^2 values.

As analyzed in Chapter 2.7 for service AI startups, the scaling mechanisms appear to be largely driven by specific factors such as human-labor intensive process automation, AI automation and abstraction, data-driven decision-making (DDDM), and data network effects. These scaling mechanisms combined represent the first time these effects are quantitatively measured beyond different companies. Interview-based studies are more common in this research area.

More precisely, the scaling barrier in traditional service startups, as identified by Weber et al. (2021), is primarily human labor. The introduction of AI can significantly raise this barrier, offering a clear shift in effectiveness for service business models. This shift in overcoming the scaling barrier may be demonstrated in my findings, providing a new view on the role of AI in service business model scalability.

To anticipate the criticism that my analysis is partially based on such low R^2, in some disciplines, these values are seen as high enough, as they represent a 1% increase in explained variance per 0.01 increase in R^2. Especially compared with the paper from Schulte-Althoff et al. (2021) where the analyzed startups had an R^2 between 0.05 and 0.23.

However, as discussed in Chapter 6.3, the findings related to AI-enabled platform startups should be viewed cautiously. The use of AI in platform models is not mature yet, and the effects of AI on platform scalability are still emerging.

6.2 Practical Implications

The findings of this study suggest several practical recommendations for investors, startup founders, researchers, and consulting professionals aiming to maximize scalability and growth potential. However, the predictive character of these recommendations should be viewed more as suggestions than a scientifically proven predictive model instruction.

Recommendations for Investors:
Investors seeking high-growth opportunities should prioritize service AI startups and some promising service non-AI startups. Service AI startups, in particular, exhibit

enhanced scalability, especially as they grow beyond specific employee thresholds. Additionally, choosing startup locations strategically can be advantageous. For example, cities highlighted in Table 5.8 will likely foster startup success due to their established entrepreneurial ecosystems and resources. Furthermore, industry selection is critical, as can be seen in Table 5.9, focusing on startups in industries marked as "hidden gem" or "healthy" to avoid industries labeled as "hyped".

Recommendations for Founders:
For founders, the location and sector of their startup are important factors in growth potential. Establishing a startup in a major urban area or cluster, as identified in Table 5.8, may provide access to essential resources, networks, and funding opportunities. Additionally, a focus on AI-enabled service models is recommended for those aiming to scale rapidly. Industries such as healthcare or other sectors listed as "healthy" in Table 5.9 offer promising revenue-scaling opportunities. By incorporating specialized AI-driven services, founders can enhance their scalability potential, positioning their startups to meet increasing demand effectively and achieve dramatic growth.

Implications for Research:
These findings highlight the importance of further investigating the role of AI in different business models and industries. Future research could explore the long-term effects of AI adoption on scalability in diverse sectors, examining why AI offers higher scalability advantages in service-focused startups but not in platform-based ones. Additionally, analyzing other influential factors, such as company structure, market conditions, or technological adaptability, may provide deeper insights into the mechanisms behind AI-driven scalability. Such research could enhance the understanding of AI's impact on business model efficacy and growth dynamics.

Practical Applications and Consulting:
For consulting professionals, these insights offer a basis for developing targeted scaling strategies for startups to scale up. Consultants can leverage these findings to advise AI-enabled service startups on maximizing scalability while providing platform startups with alternative growth strategies. Additionally, consulting professionals can use this study's location and industry-specific data to offer tailored advice on market entry, funding opportunities, and industry selection. Policymakers may also find this data useful for designing targeted support programs to stimulate growth in sectors and regions with strong AI potential, thus promoting economic innovation.

These recommendations provide actionable insights that align with the observed scaling patterns in AI-enabled and non-AI service sectors, supporting strategic investment, operational decisions, and further research in the startup ecosystem.

6.3 Limitations

The biggest limitation of this work, in my opinion, lies in the method by which the business model is extracted from the self-description of startups. As startups often change their economic focus and industries, it is hard to say if the self-description that is once put into PitchBook or on their websites really describes the business model currently or in the long term. Also, it is hard to tell if startups are really using AI and not just pretending to do so for the benefit of attracting investors or business opportunities. Furthermore, the question arises of whether startups really use their own AI models or just use publicly available AIs, like ChatGPT, via an API.

A critical weakness of the analysis is the technology life cycle. For example, AI technologies, as pointed out in a study by Mishra and Tripathi (2024) need to add business value in the ”creation”, ”survival” and ”growth” phase of technology to have success. For conventional service solutions, these three phases take 3-9 years, 9-15 years for AI service solutions, and machine and deep learning platforms 15-30 years. To have the final success measured, the ”equilibrium” phase afterward needs to be reached, which could also take 1-5 extra years. For conventual platforms, there is no timeline given (Mishra & Tripathi 2024, Fig. 4).

As the value creation of AI technologies is the main innovative part of AI startups, these time spans should be considered. It would be beneficial to make this analysis additionally take into account these timelines and have an unrestricted time span for companies to be analyzed. According to this timeline, while service AI startups in this database have just begun to show success, the platform AI startup functions have not yet experienced this scaling change. This would especially explain why the platform AI startups do not show this AI-enabled boost in revenue.

In the referred book from West (2019) is a graph that illustrates the survivorship of companies based on their lifespan, Figure 6.1. I added a line where the definition of startup ends in this analysis.

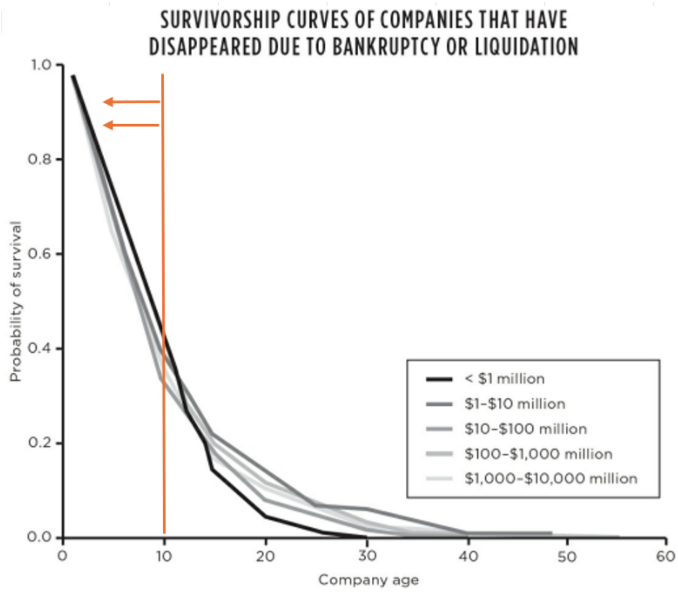


Figure 6.1: Probability of survival for ventures (West 2019, chapter 9, p. 21). The graph was modified to visualize the startup time span.

You can see in Figure 6.1 that the probability of companies surviving past the 10-year mark is around 40 %. So, even if the analyzed startups in this master thesis scale enormously, it does not mean they will survive beyond 10 years. The scaling of a startup business type should not be confused with a company's economic sustainability.

I focused on companies in this data set that are unlikely to survive the next 10 years by some estimates. It is counterintuitive to analyze the growth of companies that are likely to disappear in just a few years. Therefore, conducting a sustainable scaling analysis would also be beneficial.

6.4 Outlook

In this chapter, I will present further ideas related to the master thesis topic. One key suggestion to enhance this work is to manually review all data sets and filter out incorrectly classified companies, ensuring they are placed in the correct category. This process would significantly strengthen the analysis at the cost of very labor-intensive data labeling.

Additionally, it would be valuable to apply exponential and logarithmic functions, particularly to platform startups. These business models leverage network effects, and utilizing both functions may better capture this scalability compared to the power function.

We should also consider examining the application of service AI across various industries, as illustrated in the AI Index Report Perrault and Clark (2023). Notably, the cybersecurity and drone industries have demonstrated resilience during the COVID-19 crisis (Perrault & Clark 2023, p. 196). Based on an analysis of 100 companies (as shown in Table 3.2, "AI Service" sample), it appears that there are a greater number of startups in these two industries. This raises the question of how scaling occurs specifically regarding them.

Lastly, it would be interesting to categorize startups according to the AI technologies they utilize. Investigating whether scaling varies based on the type of technology employed could yield important insights.

7 Conclusion

This study investigated the scaling potential of artificial intelligence (AI) in platform and service startups, focusing on the economic impact of AI-driven growth models. By analyzing approximately 17,000 startups from the PitchBook database and applying regression models within West (2019) and Schulte-Althoff et al. (2021) scaling approach, it was found that AI-powered service startups have significantly higher scaling rates than their non-AI service counterparts, with scaling difference in exponents from 0.6 to 3.41 times higher. Specifically, the scaling exponents for gross profit and venture capital increased from 0.7 for non-AI service startups to 1.0 for AI-enabled service startups, and revenue scaling exponents rose from 1.1 for non-AI service startups to 2.9 for AI service startups.

Furthermore, this study examined differences in scaling intensity across locations and industries and showed that AI plays a critical role in reshaping growth patterns in service-based startups across sectors. In contrast, this effect was not observed in platform startups, suggesting that the scaling benefits of AI may be business model dependent.

The results indicate that AI enhances the scaling potential of service startups and facilitates faster growth. By emphasizing the simplicity and effectiveness of the scaling formulas utilized in this study, it underscores their practical usefulness in assessing the innovation capabilities and growth prospects of AI-driven business models. However, it is important to note that these findings are more descriptive than predictive.

In summary, these findings might have valuable implications for researchers, investors, and founders seeking to leverage AI as a strategic tool for competitive advantage in an increasingly digital economy. This work contributes to the growing understanding of AI's potential to drive transformative growth in startups and suggests promising avenues for future research on AI-enabled scalability.

8 Appendix

8.1 Bibliometric Analysis

Bibliometric analysis focuses on the data evaluation of scientific literature (Öztürk, Kocaman, & Kanbach 2024), with particular attention to the specific format of articles and proceedings. Every Article has an abstract, authors line, keywords line, body of actual text, and reference section.

VosViewer is a program I used for my bibliometric analyses. Here, I have opted for the Keyword Analysis and the Co-Word Analysis, simply because the volume of literature I have is so extensive, 111 articles, proceedings, and book chapters, and with these two Analysis options, you can have a relatively calculation-wise lightweight overview.

VosViewer, the program I used to make the bibliometric analyses, has essentially three approaches:

- Authorship Analysis: This involves evaluating the authorship of articles to determine who has collaborated with whom and the number of articles they have written together. Author clusters could be identified.

- Citation Analysis: This method assesses who has cited whom, allowing for the creation of graphs that illustrate the relationships between different works. This approach uses the reference section of each paper.

- Keyword Analysis: This approach examines the keywords of the articles to identify which thematic areas are currently being explored frequently.

- Co-Word Analysis: This method takes the title and abstract of all papers and, calculates the co-occurrence of important words or word-sequences of texts.

I have opted for the Keyword Analysis and the Co-Word Analysis simply because the volume of literature I have is so extensive, 111 articles, proceedings, and book chapters, and with these two Analysis options, you can have a relatively calculation-wise lightweight overview.

The reference analysis option requires significant data preparation, while the authorship analysis does not provide an overview. Therefore, these two analysis approaches are not used here.

The articles that were used were from a list of relevant databases from my faculty. A big part of the literature research is systematic literature research, which I have used from a previous project (Schapiro et al. 2024), where a group of 5 students, including me, have worked for half a year on a similar topic. The additional research in this thesis is not systematic. I have used the databases listed in Table 8.1.

Table 8.1: Overview of databases used for literature review

Database	**Description**
Primo	Internal search platform at FU that searches multiple databases and library collections.
Web of Science	Comprehensive bibliographic and citation database for scientific publications and citations across various disciplines.
EBSCOhost (Academic Search Ultimate and Business Source Premier)	Platform offering a variety of databases, providing scholarly articles from various disciplines, including economics and social sciences.
AISEL	The Association for Information Systems Electronic Library, a database for scholarly articles in the field of information systems.
Google Scholar	A search engine designed for scientific content.
ScienceDirect	A scientific database providing access to over 2,500 peer-reviewed journals across various disciplines.

My current literature search (as of October 1, 2024) consists of 111 publications. Figure 8.1 shows the steps that needed to be taken to provide those graphs:

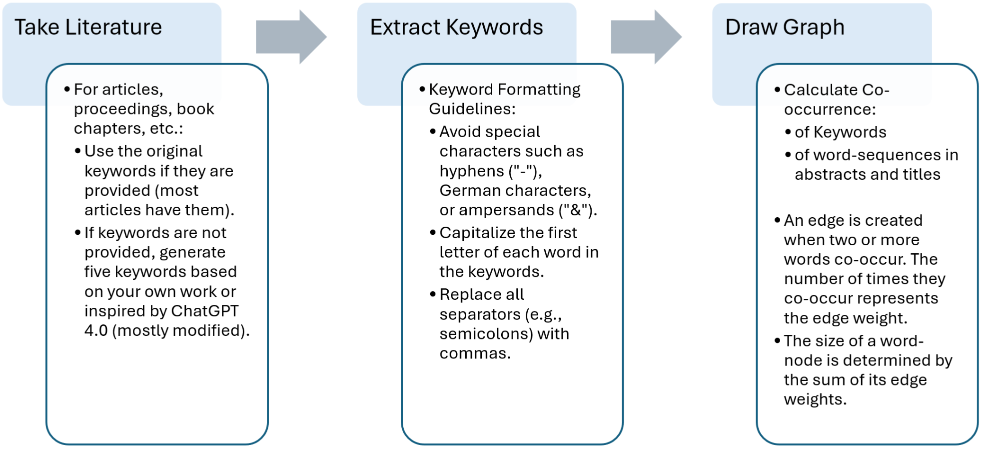


Figure 8.1: Process for bibliometric analysis, own figure

The first step is to find the literature, read it, and put it in the database of important literature, it should be done either way for a master thesis, and since I work with latex as a writing tool, I can format the literature in the .bib format. This format looks for example:

```
@proceedings{schulte-althoff_scaling_2021,
    title = {A Scaling Perspective on AI Startups},
    url = {https://scholarspace.manoa.hawAIi.edu/server/api/core/bitstreams...},
    doi = {10.24251/HICSS.2021.784},
    abstract = {Digital startups use of AI technologies has significantly ...},
    author = {Dr. Schulte-Althoff, Matthias and Prof. Fürstenau, Daniel ...},
    urldate = {2023-11-21},
    year = {2021},
    langid = {english},
    keywords = {Artificial Intelligence, AI Startup, Scaling, Digital Startup,...}
}
```

You can see that the fields of title, abstract, and keywords are already collected.

Keywords Co-Occurrence Represented as a Heatmap

First, you can display the most relevant keywords to the topic, especially if you want to know which termini are right. know is mostly related to the topic in the literature. The more often a word occurs, the darker red it is shown.

Here are the boundary conditions used: (min size cluster = 4, lin/log modularity, min occurrence = 2, select from 65 termini all).

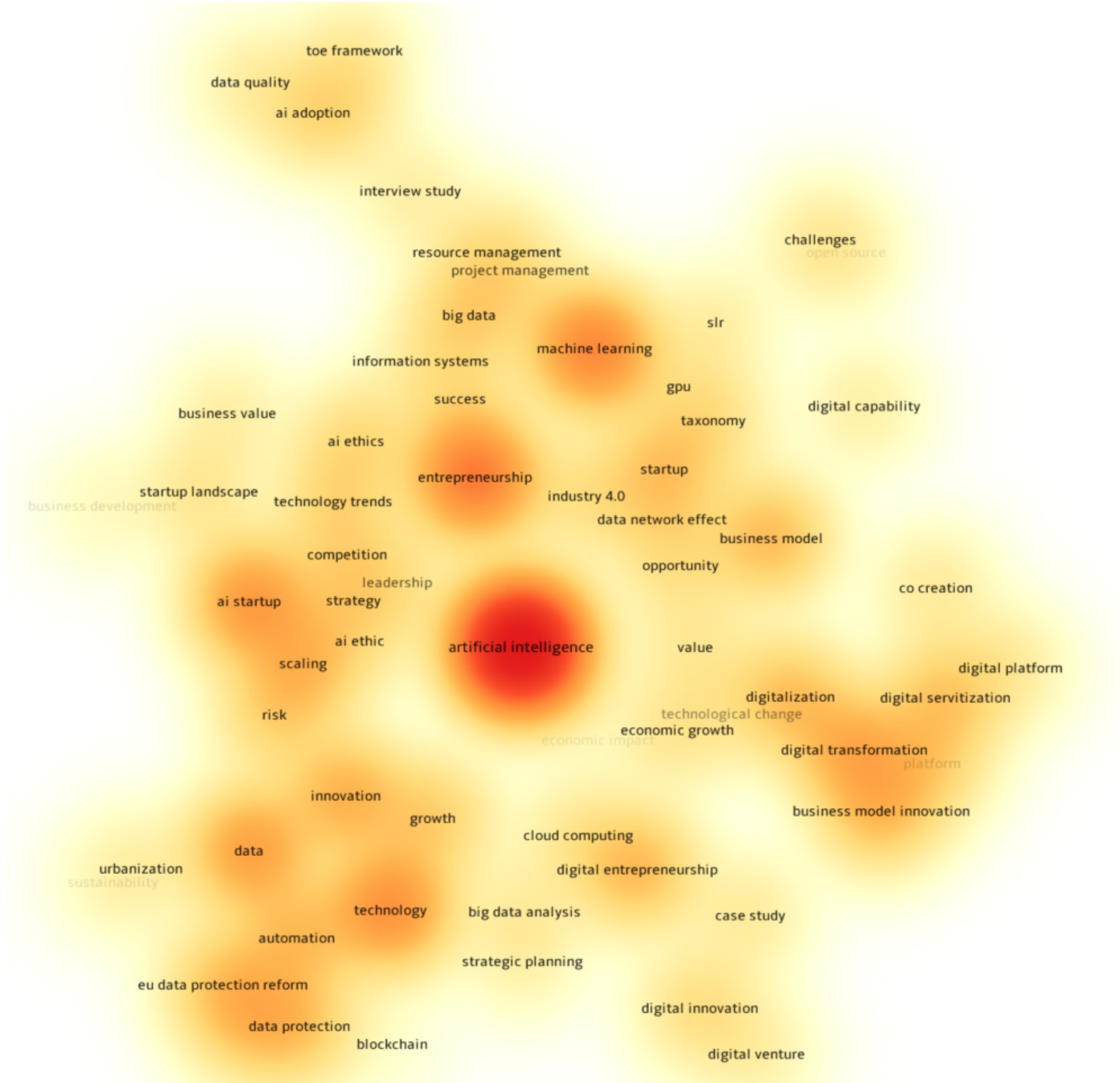


Figure 8.2: Heatmap of bibliometric analysis, own figure

For a more detailed understanding of the context of the words, which are represented in a darker tone in Figure 8.2, the next subsection provides a more complex representation.

Bibliometric Data Keywords Co-Occurrence of All Papers as Clusters

The graph in Figure 8.3 shows the co-occurrences of keywords that are used in the papers. If you want to zoom in, here is the full graph on a webpage:

`https://app.vosviewer.com/?json=https%3A%2F%2Fdrive.google.com%2Fuc%3Fid%3D1j5E6w6fwPZI1CB2T_74N244fXAyEwipd`.

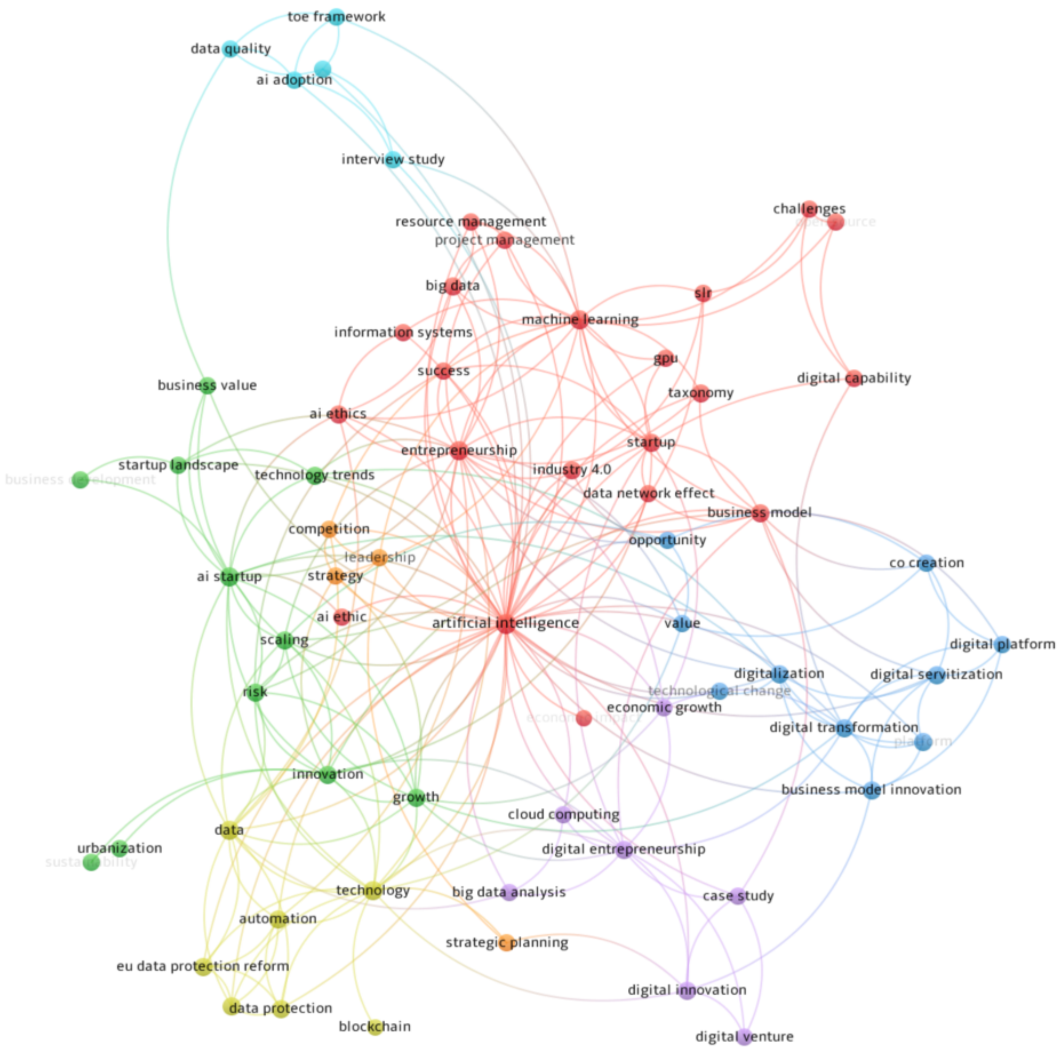


Figure 8.3: Co-occurrence of keywords as clusters, own figure

Boundary conditions for the graph are as follows: (min size cluster = 1 , lin/log modularity, min occurrence = 2, select from 65 termini all). The clustering is calculated based on the number of clusters to which an item belongs. Here it is set by me on 1. More details are in the user manual of VosViewer (van Eck & Waltman 2018).

The 6 clusters have the natural tendency to fall align with different scientific disciplines as elaborated in Table 8.2.

Table 8.2: Content description of clusters: keyword co-occurrence

Cluster	Description
Red	Main cluster related to AI, machine learning, entrepreneurship, big data, and ethics. Generally, it has an information systems view on the topic.
Blue	Literature related to digital transformation, business model innovation, and platforms. A business administration perspective.
Purple	Focuses on digital entrepreneurship. Literature often consists of case studies explaining successful examples. Business administration perspective.
Orange	A small cluster about strategic planning and leadership includes books that try to deduce new types of desirable leadership figures based on AI opportunities.
Dark Yellow	A technical cluster related to computer science, covering topics like technical data handling and laws related to data processing.
Green	The AI startup cluster, where factors like innovation trends, assessment bases for venture capital, urbanization, and other mostly external factors for AI scaling are grouped. Includes my supervisor's paper.
Light Blue	Focuses on assessments of organizational readiness, often based on interview studies. Features a business administration perspective for leaders.

Abstract and Title Co-Occurrence of Word-Sequences

Keywords are often given from journals, and they can be oriented on the field in which the journal is publishing. That is why it is beneficial to look into the titles and abstract texts that authors write themselves.

8.1.1 How Authors Perceive the Research Topic

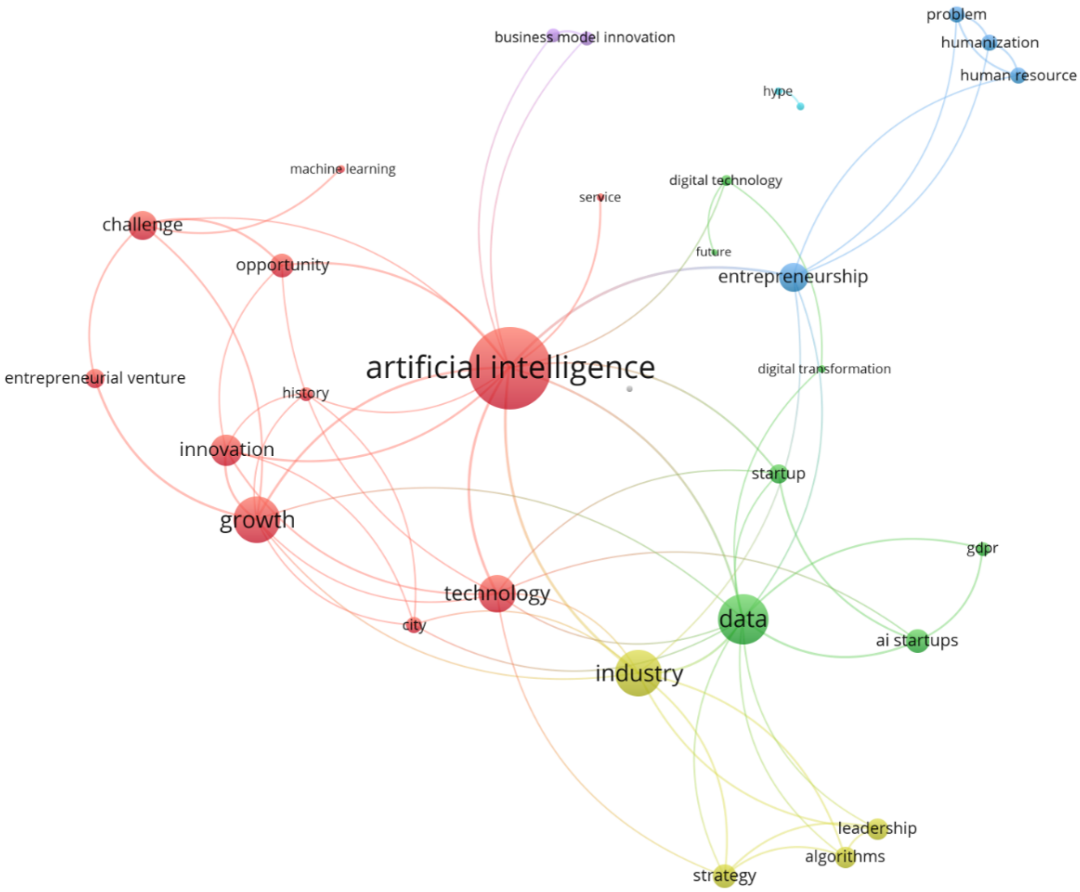


Figure 8.4: Bibliometric analysis word-co occurrence in abstract and title, own figure

The same analysis was performed as in the graph in Figure 8.3, but based on the abstract and title field of the papers. The more often word sequences are occurring, the bigger the node which is displayed, this is the same for Figure 8.4.

You can see, that the most important words are "artificial intelligence", "industry", "data", "growth", "innovation" and "entrepreneurship". As I'm investigating service and platform startups, it's also interesting to look up what terms are associated with those: "service" is only connected to "artificial intelligence" and "platform" has unfortunately not occurred enough and therefore is missing an association.

To show the difference between the views of journals and researchers on this topic, I compare the clusters of each visualization to each other. Here are only 6 clusters, presented in Table 8.3, for which I am comparing the previous Figure 8.4:

Table 8.3: Content description of clusters: word co-occurrence

Cluster	Description
Red	This information systems view cluster has become more precise, focusing on technological aspects. Geographical aspects have shifted to focus on cities, and innovation is now linked to technology. The technological cluster from previous classifications is merged here, supporting the view that information systems topics should integrate technological know-how. Machine learning is mainly discussed in terms of challenges.
Purple	This cluster is about business model innovation and scaling AI, though it has reduced in size.
Green	This is the data cluster, grouping perspectives on data related to AI and digital transformation, including legal perspectives. Data-related topics are now a standalone cluster previously associated with technology. Now, they are directly linked to AI and other startups.
Blue	Focuses on a distinct view of entrepreneurship in AI, discussing issues such as human resource challenges.
Light Blue	A very small cluster about hype and life cycles of AI technology, standing alone with few abstracts and titles having this perspective.
Dark Yellow	This cluster is about industry-related topics like strategy and leadership. The previous orange cluster on leadership and the light blue cluster on organizational assessment have merged here, which makes sense as leadership topics fit well in industry contexts.

So, in short, if you want to know what the bibliometric analysis has to say about AI-Startups and Artificial Intelligence, we can zoom into the relevant graphs, as shown in Figure 8.5.

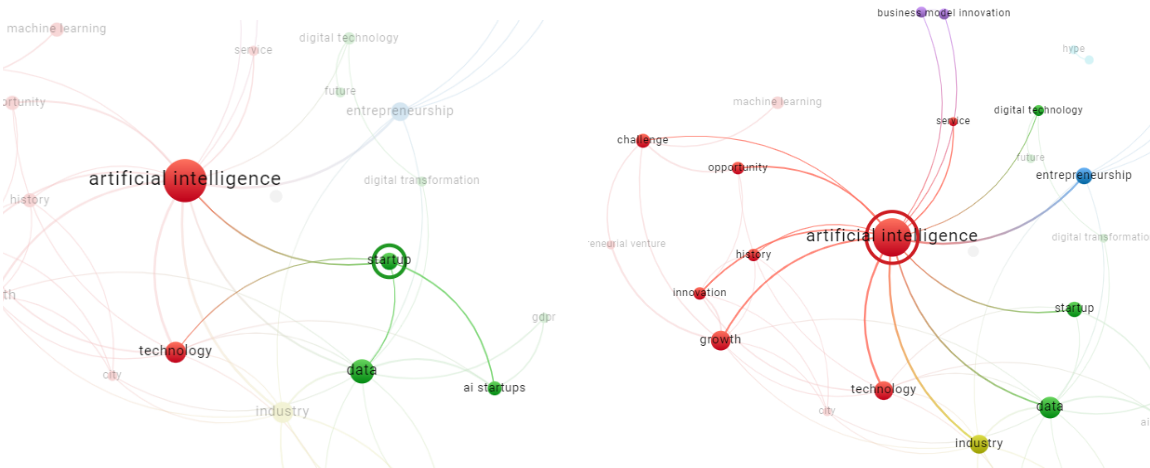


Figure 8.5: Focus on startup and AI topics, own figure

You can say that the startup and AI-startup terms are in the green data cluster, and the red AI cluster with technology relation is also important to investigate. Especially the "Artificial Intelligence" term is viewed from an information systems view (red cluster) in this master thesis, adding small insights from the technical data approach (green data cluster) and business administration (purple, blue, dark-yellow clusters).

So let us define the term "artificial intelligence" based on this bibliometric research, and how the authors perceive the term:

Artificial Intelligence (AI) is a technology closely linked to the fields of data, innovation, and growth, providing critical opportunities and challenges for entrepreneurial ventures. AI is based on algorithms and is used in various industries to promote the growth and development of both start-ups and established companies. It supports strategic decisions, the humanization of resources, and digital transformation. In the modern economy, AI is regarded as a central element that drives the competitiveness of companies through data-driven processes and technological innovations.

You can see this definition is not very useful for my analysis. That is why I decided to keep the first definition from the keyword analysis.

Also, the bibliometric analysis handles all papers equally important, which is not always beneficial. A case study on a successful AI startup is not as important as a quantitative overview study, for example, one of the business models for AI startups.

8.2 Dataset Visualization

8.3 Data Procurement Query

The correct and full query is here:

```
AI = "Artificial Intelligence"
    OR "Data Analytics"
    OR "Artificial Intelligence & Machine Learning" (vertical)
    OR "Artificial Intelligence & Machine Learning"
```

This AI category is now reused in the Platform and Service categories:

```
1) AI-enabled Platform = AI
    AND (platform
        OR Broker
        OR Intermediary Service
        OR Marketplace OR Digital Ecosystem
        OR Social/Platform Software (vertical)
        )

2) Non-AI Platforms = NoT AI
    AND (platform
        OR Broker
        OR Intermediary Service
        OR Marketplace OR Digital Ecosystem
        OR Social/Platform Software (vertical)
        )

3) AI-enabled Service = AI
    AND (Digital Services
        OR Commercial Services (Vertical)
        OR Services (Non-Financial) (vertical)
        OR Business Products and Services (B2B) (vertical)
        OR Consumer Products and Services (B2C) (vertical)
        OR Financial Services (vertical)
        )
    AND Information Technology (vertical)
    AND NoT (platform
        OR Broker
        OR Intermediary Service
        OR Marketplace
        OR Digital Ecosystem
        OR Social/Platform Software (vertical)
    )
```

```
4) Non-AI Service = NoT AI
    AND (Digital Services
        OR Commercial Services (vertical)
        OR Services (Non-Financial) (vertical)
        OR Business Products and Services (B2B) (vertical)
        OR Consumer Products and Services (B2C) (vertical)
        OR Financial Services (vertical)
        )
    AND Information Technology (vertical)
    AND NoT (platform
        OR Broker
        OR Intermediary Service
        OR Marketplace
        OR Digital Ecosystem
        OR Social/Platform Software (Vertical)
        )
```

Continents of Startups

Figure 8.6 displays the continents (hqRegions) in which the startups are located.

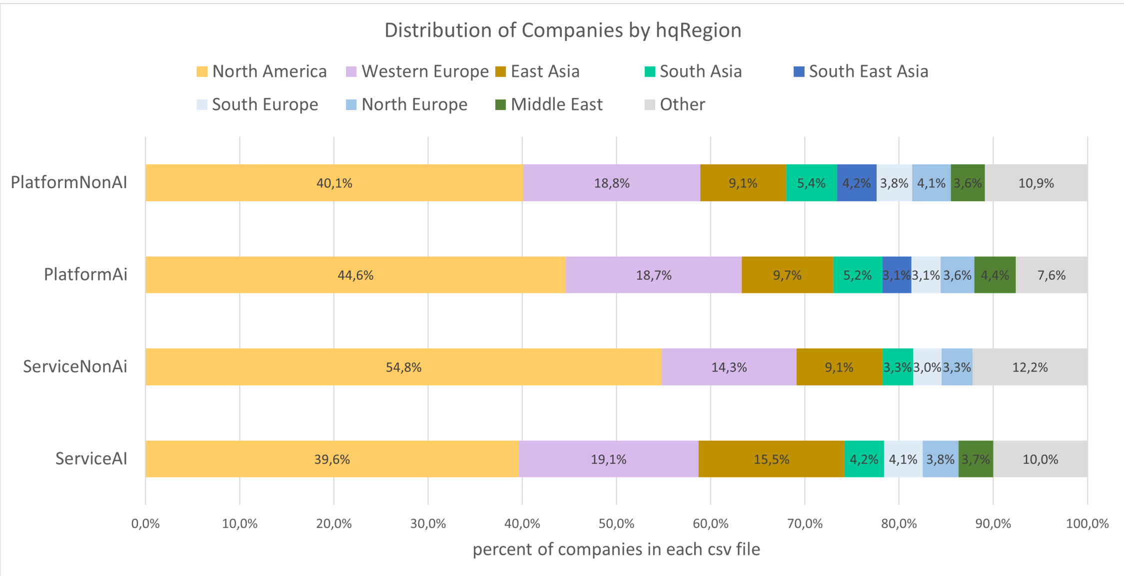


Figure 8.6: Headquarters of 156,732 startups in PitchBook by startup types on different continents, own figure

Preparation for Analysis: Number of Employees Restriction

One of the optional requirements is the number of employees limited to 100. In the following Figure 8.7, I show how the condition influences the number of data points for all four business types.

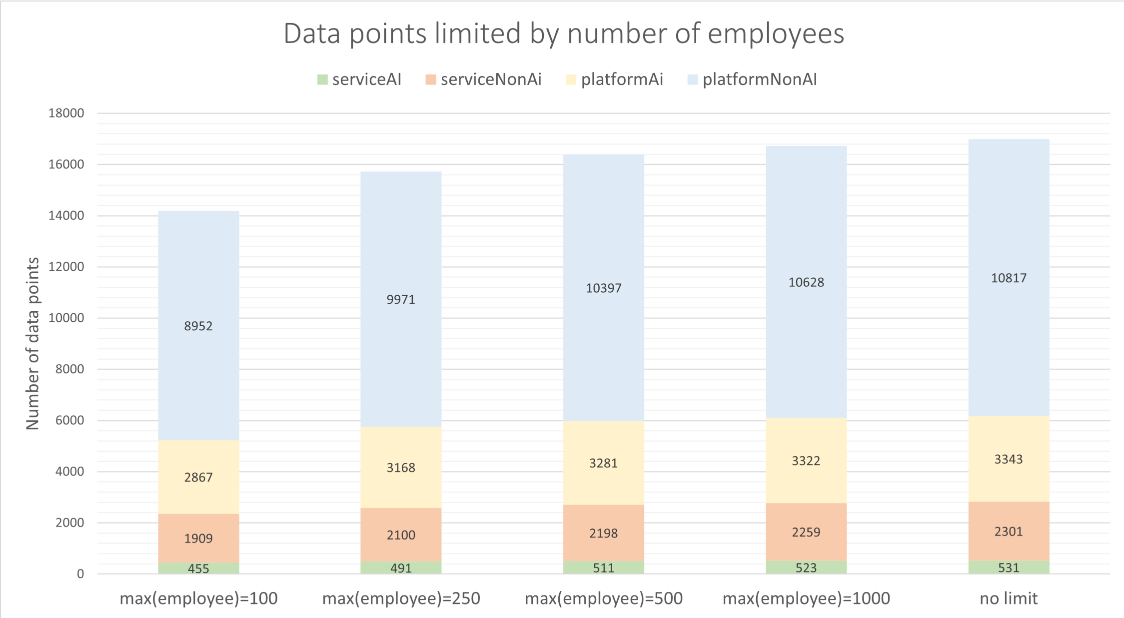


Figure 8.7: Number of data points regarding which maximum of employees is set, own figure

The category with the smallest amount of data points is the service AI startup group. You can see that if the bar of 100 employees, as a maximum, is raised to 1000 employees, this group can grow up to 523 data points. This growth in data points is up to 13%, or if no limit is set, this group can grow up to 14%, resulting in 531 companies. This small visualization and discussion led to the maximum employee restriction being left ignored in this master thesis.

8.3.1 Attributes and Choose of Columns

1	financial data	company data	www data	industry data	science data
2	activeInvestorCount	activeInvestors	companyWebsite	allIndustries	clinicalTrials
3	competeGrowthScore	businessStatus	majesticGrowthScore	competitors	inactivePatents
4	competeGrowthScorePercentile	companyLinkedIn	majesticGrowthScorePercentile	primaryIndustryCode	totalClinicalTrials
5	competeScore	companyPbId	majesticRefDomains	primaryIndustryGroup	totalPatentDocuments
6	competeScorePercentile	employeeHistory	majesticRefDomainsChange	primaryIndustrySector	totalPatentFamilies
7	competeVisitors	employees	majesticRefDomainsPercentChange	verticals	
8	competeVisitorsChange	familiarName	majesticScore		
9	competeVisitorsPercentChange	formerInvestors	majesticScorePercentile		
10	ebit	formerName	twitterFollowers		
11	ebitda	hqAddressLine1	twitterFollowersChange		
12	enterpriseValue	hqAddressLine2	twitterFollowersPercentChange		
13	exchange	hqCity	webGrowthScore		
14	financingStatus	hqCountry	webGrowthScorePercentile		
15	financingStatusNote	hqLocation	webSizeScore		
16	firstFinancingDealClass	hqPhone			
17	firstFinancingDealType2	hqRegion			
18	firstFinancingDealType3	hqStateProvince			
19	firstFinancingValuationStatus	hqSubRegion			
20	fiscalPeriod	hqZipCode			
21	grossProfit	keywords			
22	growthScore	legalName			
23	growthScoreChange	mnaProbability			
24	growthScorePercentChange	primaryContactEmail			
25	ipoProbability	PrimaryContactPBID			
26	lastFinancingDealClass	primaryContactPhone			
27	lastFinancingDealType	primaryContactTitle			
28	lastFinancingSizeStatus	registrationNumber			
29	lastFinancingStatus	successClass			
30	lastFinancingValuationStatus	universe			
31	lastKnownValuationDealType	yearFounded			
32	lastKnownValuationStepUp				
33	lastValuationStepUp				
34	marketCap				
35	netDebt				
36	netIncome				
37	noExitProbability				
38	opportunityScore				
39	predictedExitClass				
40	profileDataSource				
41	revenue				
42	sizeScore				
43	sizeScoreChange				
44	sizeScorePercentChange				
45	sizeScorePercentile				
46	successProbability				
47	ticker				
48	vcRaised				

Figure 8.8: Overview of all columns provided in the Pitchbook data, own figure

From all the columns provided in the Pitchbook dataset, as shown in Figure 8.8, I decided to use the view presented in Figure 8.9.

1	financial data	company data	industry data	use in master thesis
2	revenue	employees		<- for the scaling analysis
3				
4				
5	netIncome	yearFounded	primaryIndustryCode	<- for descriptive statistics and subanalysis
6	netDebt	businessStatus	primaryIndustryGroup	
7	ebit	hqCity	primaryIndustrySector	
8	ebitda	hqCountry		
9	enterpriseValue	hqLocation		
10	financingStatus	hqRegion		
11	vcRaised	hqStateProvince		
12		hqSubRegion		
13				
14				
15		keywords	verticals	<- already used for categorization of startups
16		discription		

Figure 8.9: Overview of columns used from the Pitchbook data, own figure

8.3.2 Visualization and statistical Evaluation of the Data for the four Startup Types.

To approach the data, it is interesting to look at the used variables presented in Table 8.4. Here are the four datasets compared to a variable are listed.

Table 8.4: Descriptive statistics across different metrics and datasets

	Metrics	Service AI	Service Non-AI	Platform AI	Platform Non-AI
Employees	Count	3050	12853	19481	66872
	Mean	57.0	81.0	44.0	49.0
	Std Dev	613.0	1181.0	705.0	709.0
	Min	2.0	2.0	2.0	2.0
	25%	5.0	5.0	5.0	5.0
	Median	11.0	11.0	12.0	10.0
	75%	28.0	28.0	29.0	26.0
	Max	26000.0	85000.0	75000.0	114000.0
Revenue	Count	531	2301	3343	10817
	Mean	56.43	78.72	47.03	77.60
	Std Dev	629.26	1225.97	775.13	1257.36
	Min	0.01	0.01	0.01	0.01
	25%	0.15	0.22	0.12	0.14
	Median	0.95	1.22	0.76	0.90
	75%	3.75	6.30	2.86	4.50
	Max	13158.90	52611.86	32345.25	65938.05
VC Raised	Count	1727	5616	12016	35736
	Mean	20.40	19.19	16.73	16.22
	Std Dev	182.50	208.02	135.52	98.58
	Min	0.01	0.01	0.01	0.01
	25%	0.46	0.36	0.55	0.42
	Median	1.87	1.63	2.37	1.80
	75%	6.96	6.25	8.64	6.81
	Max	6780.69	11371.21	10490.86	6006.74
Net Income	Count	135	681	547	2105
	Mean	5.19	11.29	7.87	11.08
	Std Dev	29.26	166.89	75.76	170.60
	Min	0.01	0.01	0.01	0.01
	25%	0.04	0.04	0.01	0.02
	Median	0.21	0.22	0.07	0.11
	75%	0.98	1.00	0.40	0.62
	Max	263.47	4266.67	1163.41	5729.73
Gross Profit	Count	70	378	300	1226
	Mean	22.71	32.24	12.32	16.58
	Std Dev	126.15	215.06	49.72	69.53
	Min	0.01	0.01	0.01	0.01
	25%	0.12	0.25	0.14	0.14
	Median	0.76	1.40	0.81	0.99
	75%	7.74	6.62	5.78	5.90
	Max	1053.53	3321.49	617.33	979.55

After reading all the lines, we can plot the histograms for the startup types.

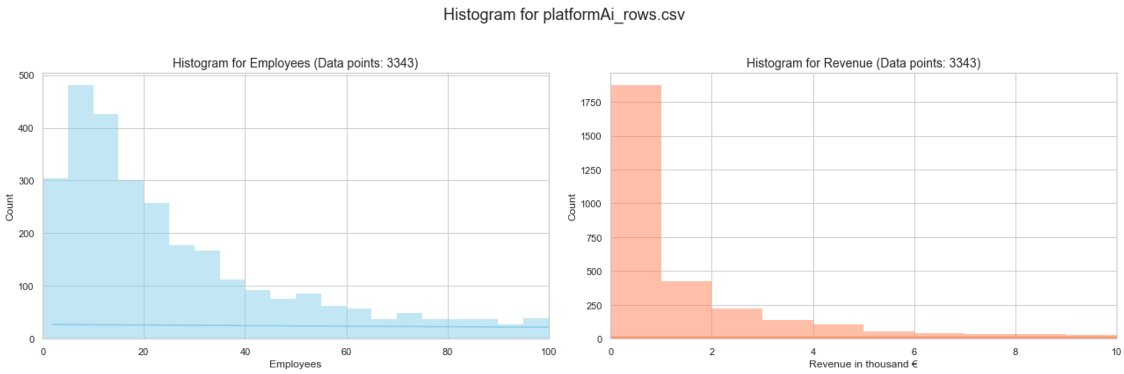


Figure 8.10: Histogram for platform AI startups, own figure

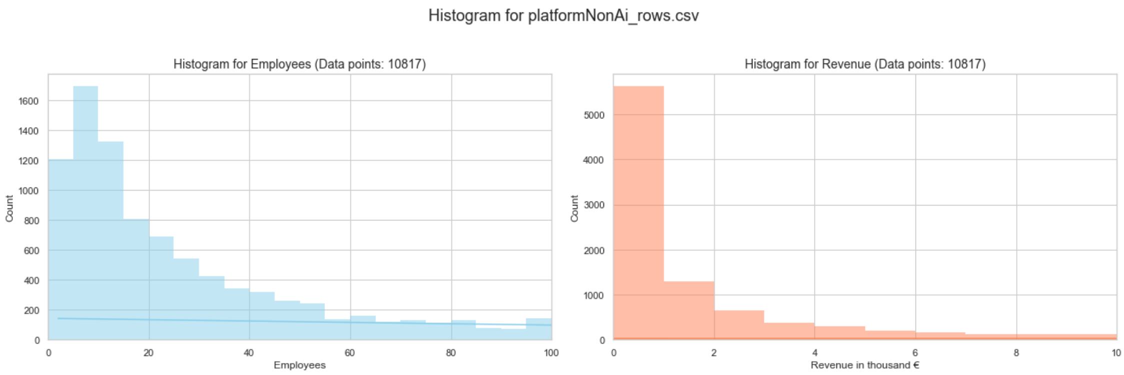


Figure 8.11: Histogram for platform non-AI startups, own figure

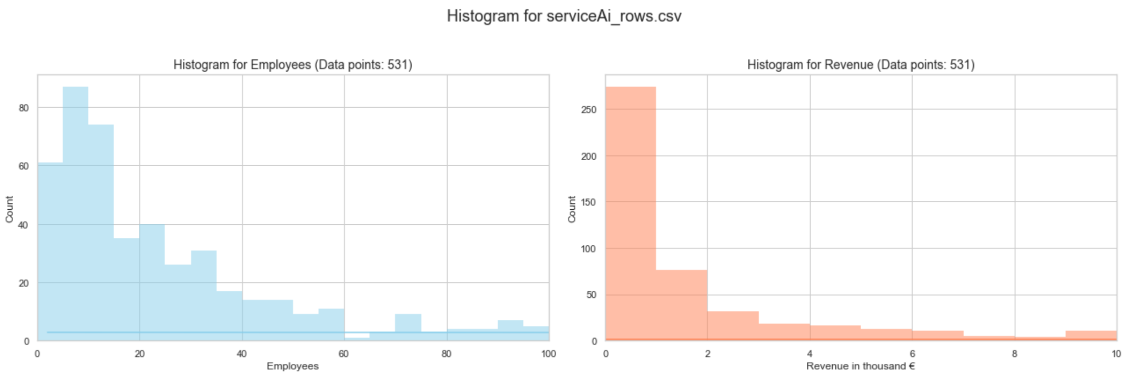


Figure 8.12: Histogram for service AI startups, own figure

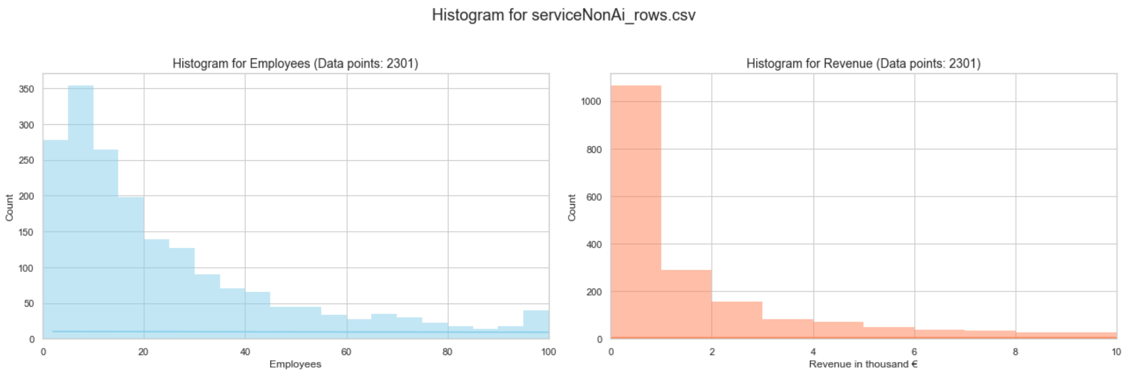


Figure 8.13: Histogram for service non-AI startups, own figure

It's interesting that most startups in all categories have between 5 and 15 employees. Additionally, many startups generate very low revenue, less than €1,000 per year. The relevant Histograms are shown in Figures 8.10, 8.11, 8.12, 8.13.

8.4 Drawing the Baseline: West's Scaling Analysis - more Details

The next step is to perform the same analysis with the same variables like West (2019) used, but from Pitchbook data with a focus on the four types of startups.

He analyzed the incomes and gross profits of nearly 30,000 companies and plotted them logarithmically against the number of employees. These financial metrics are key indicators of a company's fiscal health and performance.(West 2019).

As these analyses demonstrate, companies scale following a simple power law. "So in this statistical sense, companies are approximately scaled, self-similar versions of one another: Walmart is an approximately scaled-up version of a much smaller, modest-size company" (West 2019, p. 395).

So, now that the expectations towards the scaling analysis are set, I will try to replicate those results with startups.

8.4.1 Net Income and its average values

For each employee number the average netincome is calculated and connected, see Figure 8.14

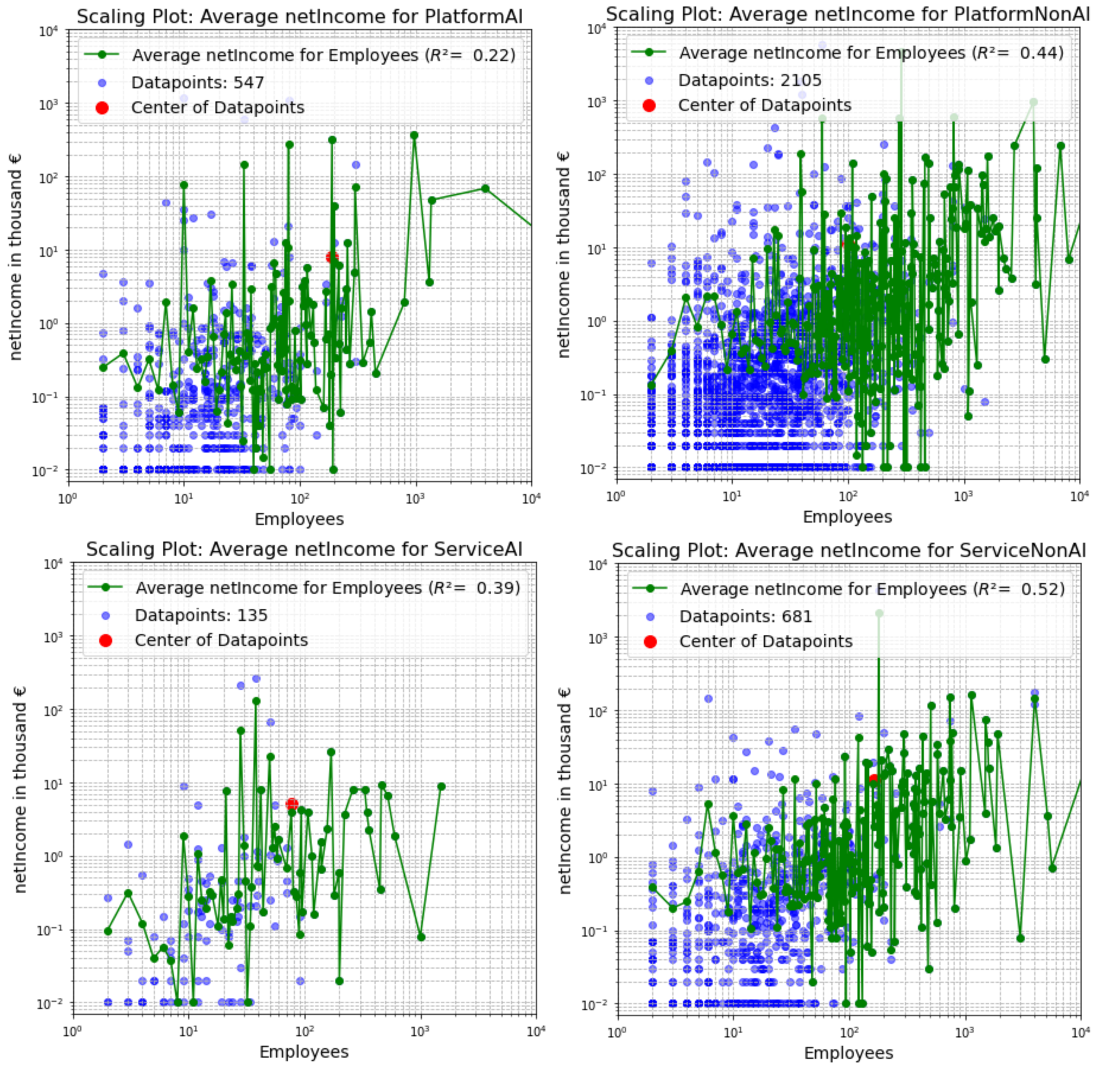


Figure 8.14: Average net income based on employees, scaling analysis using West´s methodology, own figure

As you can see, this analysis does not provide the same "line" as West (2019) has in his plots. The R^2 values are also not that high, as his.

8.4.2 Net Income with West Methodology

Using the West methodology, I obtained the following results. West split the employee axis into eight bins and calculated the mean for each bin. Afterward, he connected the means, forming a log-log scale line.

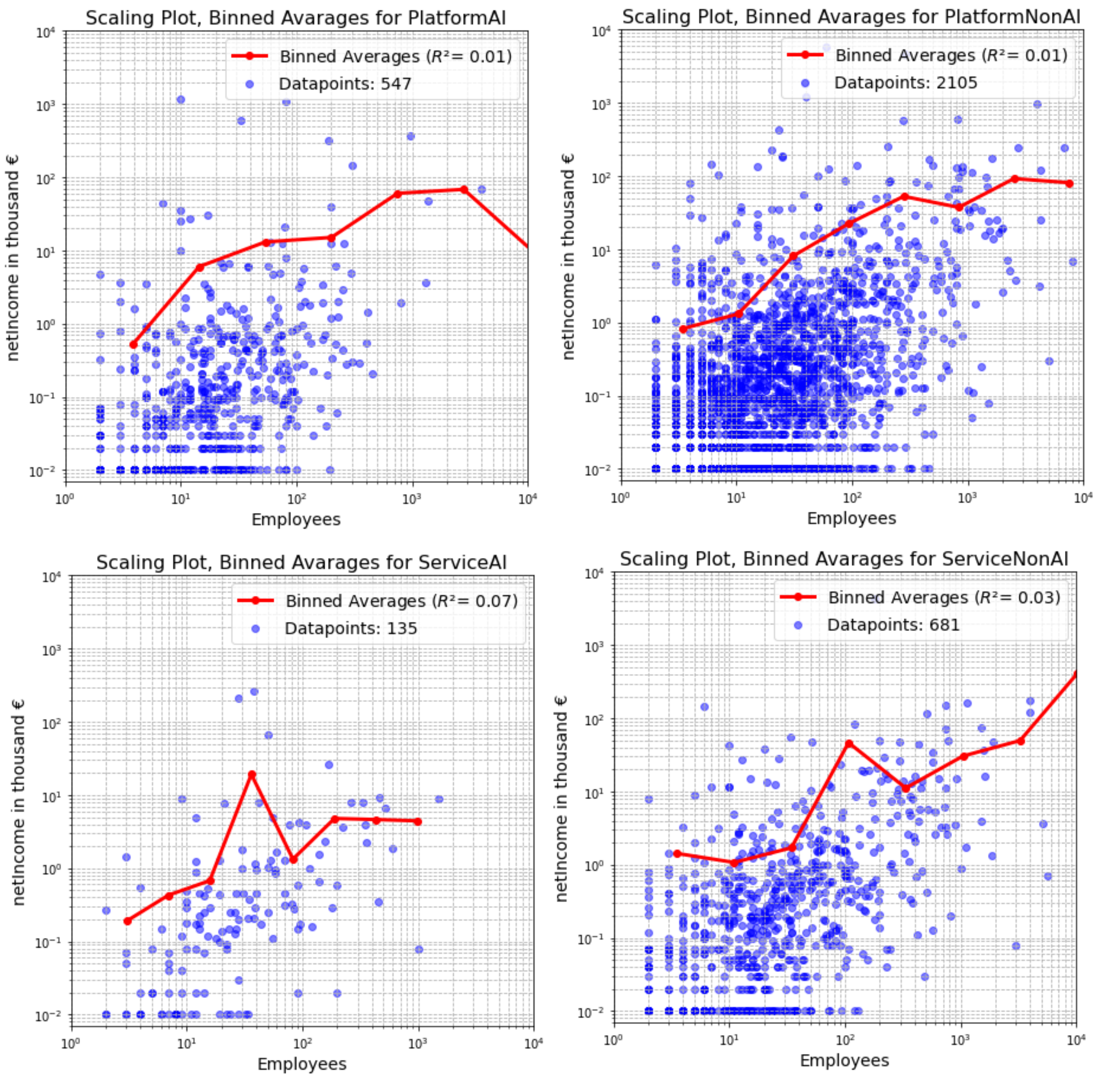


Figure 8.15: Average net income based on employees - scaling analysis using West´s methodology, own figure

You can see that the R^2 is small. These Bins connected do not form a straight line.

8.4.3 Net Income

The first analysis that West performed was to analyze, with a power function, the net income to employees for companies. The results of the scaling analysis inspired by West (2019) but performed on the PitchBook data are shown here:

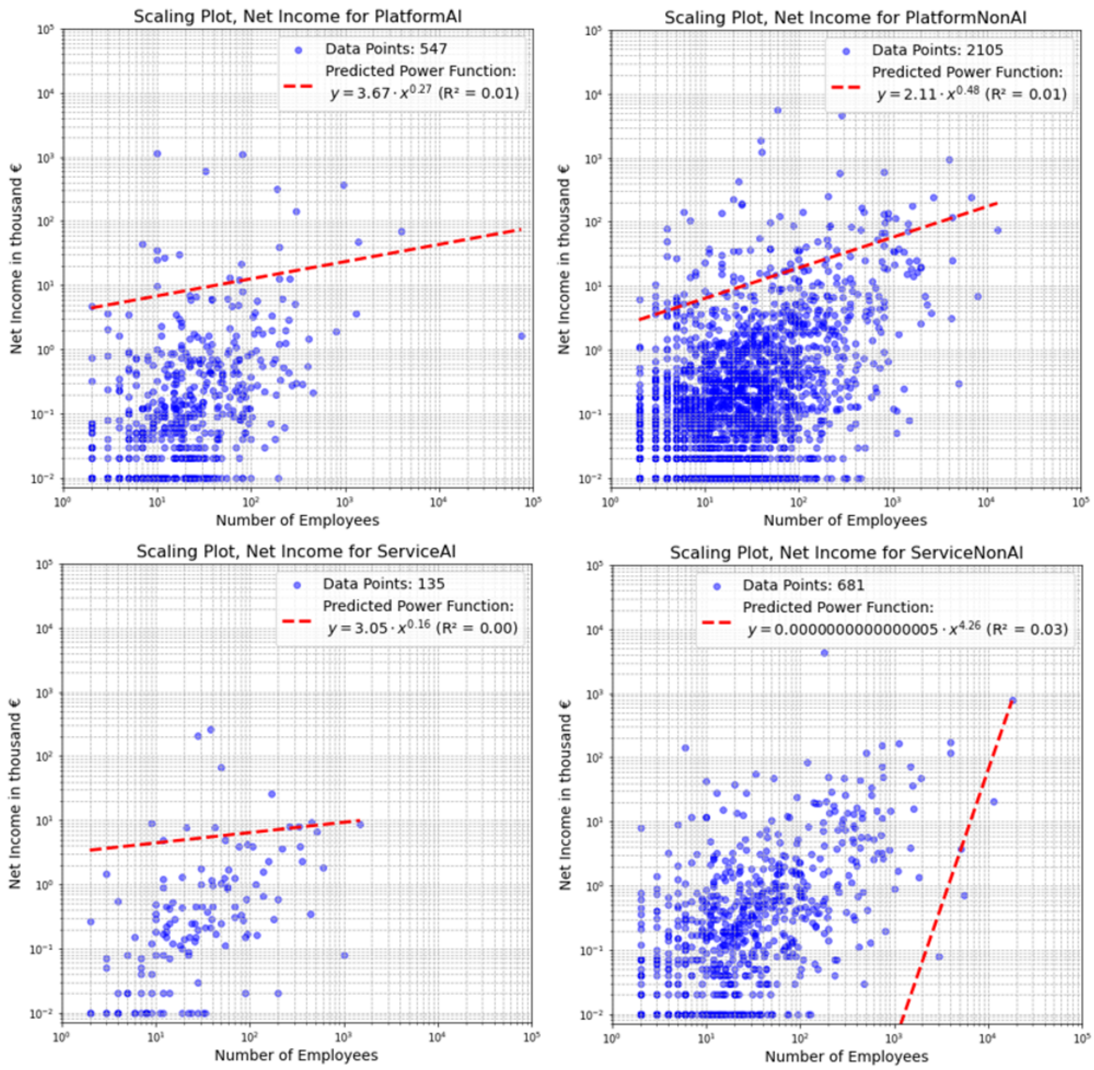


Figure 8.16: Net income based on employees, scaling analysis, own figure

You can see that the R^2 is near zero, which means the model is not good for this kind of relationship. This could be because of the startup nature that they are not really generating income yet, or the scaling-power-law function is not applicable because many startups are on the way to failing but still counted here. Potentially the universal law of scaling is only applicable to big companies.

8.4.4 Gross Profit

The second analysis that West performed was to analyze, with an power function, the gross profit to employees.

8.4.5 Gross Profit and average values

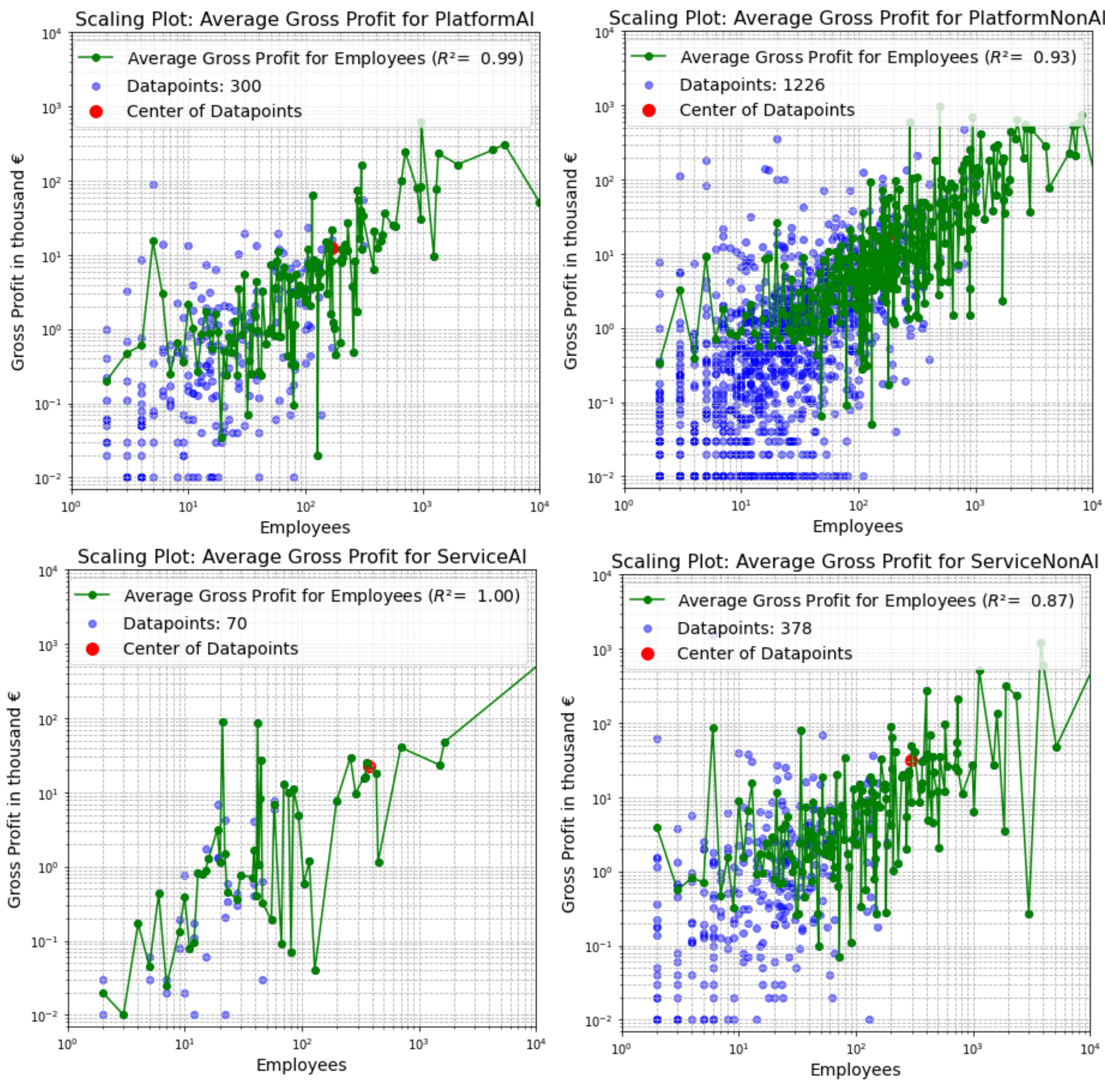


Figure 8.17: Gross Profit based on employees, scaling analysis using West´s methodology, own figure

As you can see, this analysis does not provide the same "line" as West (2019) has in his plots.

8.4.6 Gross Profit with West Methodology

Using the West methodology, I obtained the following results. West split the employee axis into eight bins and calculated the mean for each bin. Afterward, he connected the means, forming a log-log scale line.

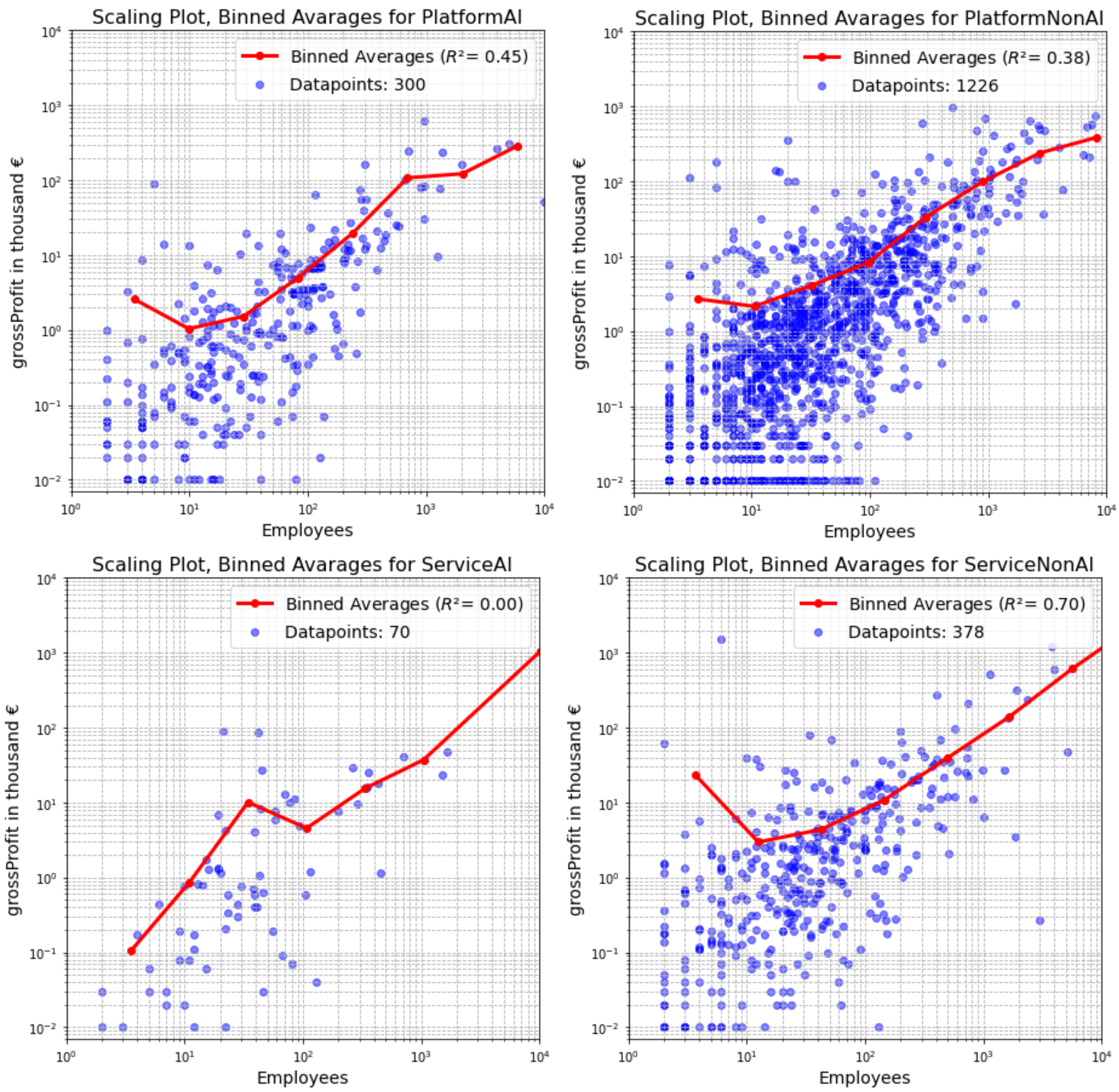


Figure 8.18: Gross Profit based on employees, scaling analysis using West´s methodology, own figure

As you can see, this analysis does not provide the same "line" as West (2019) has in his plots. But in service non-AI, it would somehow if you start looking above 12 employees. But that's the only case.

8.4.7 Gross Profit

You can see in Figure 8.22 that the R^2 values are quite good for service AI. But the three other values are not really near the $R^2 = 0.92$, like in graphs from West (2019). Also, for service AI startups, there are not enough data points to have a good analysis. The exponent for service AI of 1.24 is still remarkable because an exponent over 1 means the company is extremely innovative, as following the interpretation from (West 2019).

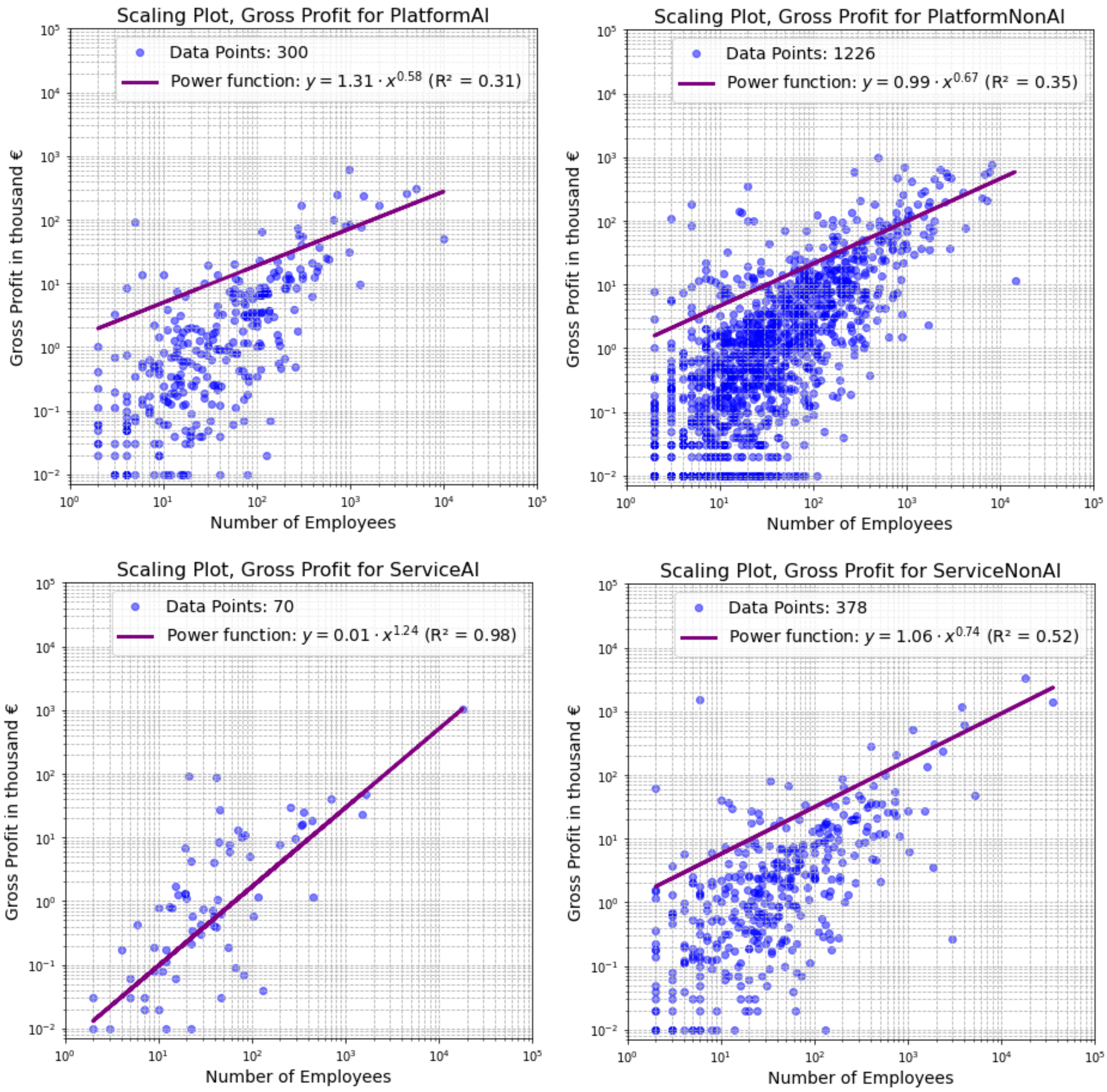


Figure 8.19: Gross profit based on employees, scaling analysis, own figure

8.5 Drawing the Baseline: Schulte-Althoff's Scaling Analysis - more Details

Like in the paper of Schulte-Althoff et al. (2021), I will examine the revenue and VC raised variables because these two attributes have far more data completeness, so far more data points for analysis.

8.5.1 VC Raised - West Methodology

Using the West methodology, I obtained the following results. West split the employee axis into eight bins and calculated the mean for each bin. Afterward, he connected the means, forming a log-log scale line.

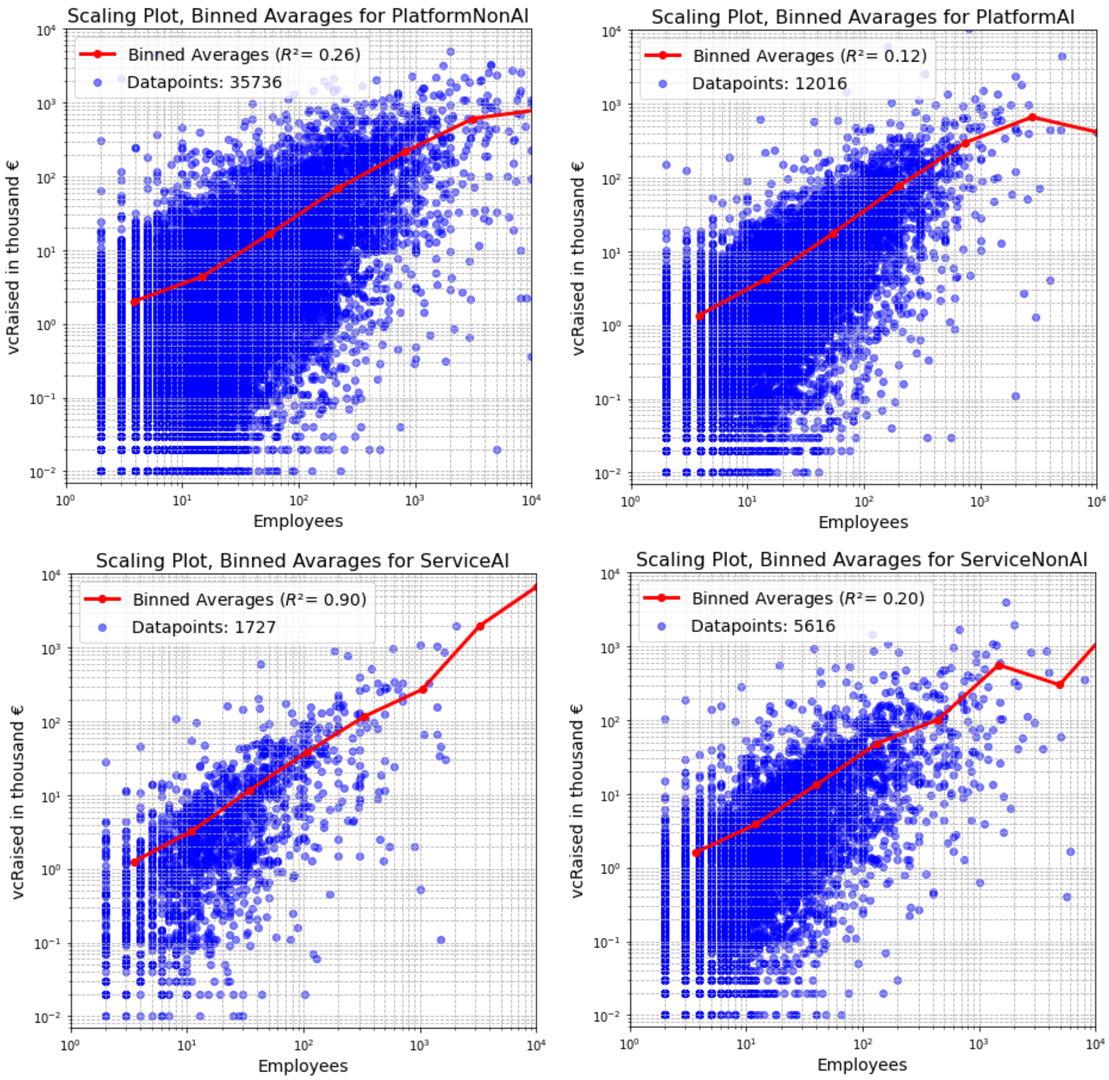


Figure 8.20: VC raised based on employees, West Methodology, own figure

This curve somehow forms a line, but the R^2 values are relatively low. Only the service AI startups following this line stand out.

8.5.2 VC Raised - Scaling Anaysis

The same mathematical power functions from the study 4.1 above have been used to produce Figure 8.21 containing four scatter plots.

In figure 8.21, the data point dots are transparent, and if many are in the same spot, the color becomes dark. The plots contain more companies or data points, and all axes scale to 10.000 in thousand € units.

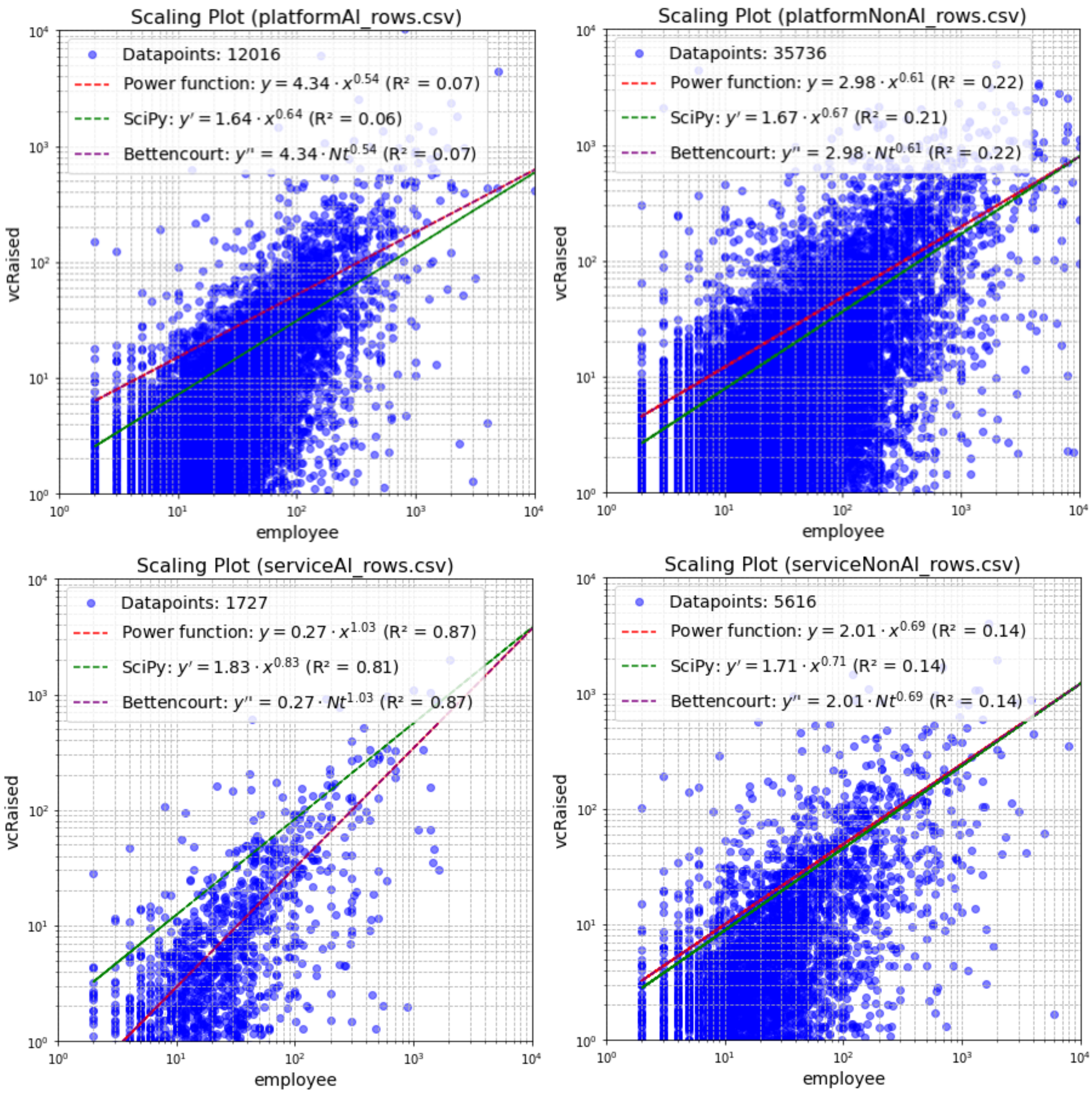


Figure 8.21: VC raised based on employees, scaling analysis, own figure

The graphs show that a lot of startups in every category have very low funding. Only the service AI startups seem to follow the power-law growth or scaling pattern with a relatively strong R^2, and regarding a large number of companies, this context is remarkable. This strongly indicates there is an power-law relationship between employees and the VC funding volume for service AI startups. For the other three categories, this analysis doesn't show this strong relationship.

8.5.3 Revenue - West Methodology

sing the West methodology, I obtained the following results. West split the employee axis into eight bins and calculated the mean for each bin. Afterward, he connected the means, forming a log-log scale line.

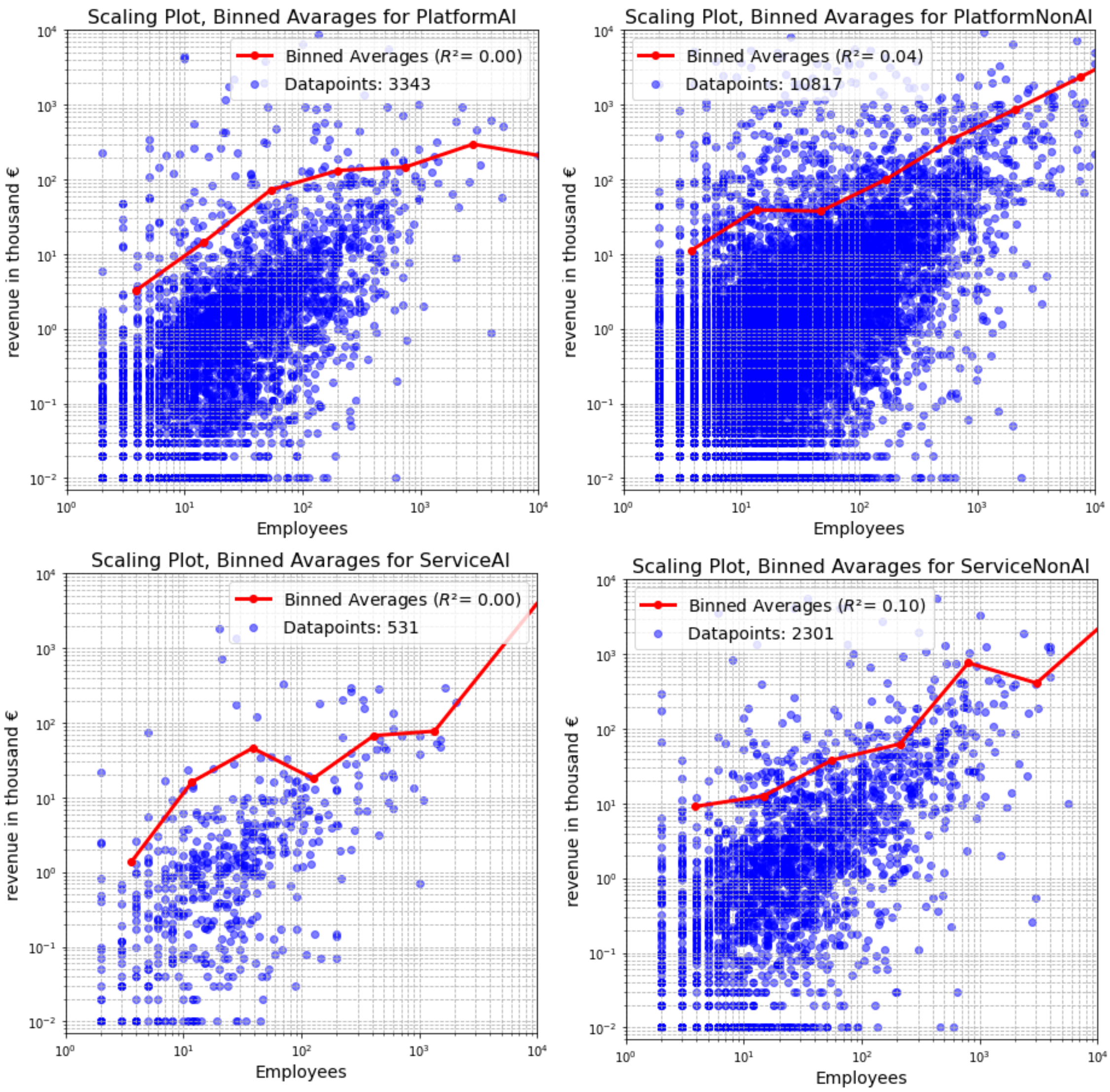


Figure 8.22: Revenue based on employees, West Methodology, own figure

This curve does not form a line, and the R^2 values are relatively low.

8.5.4 Revenue - Scaling Analysis

The same mathematical power formulas from Chapter 4.1 have been used to lay a function over the data in Figure 8.23 for the four scatter plots.

The plots in Figure 8.23 have far fewer companies presented than the VC-raised ones. And again, service AI startups stand out - even though the number of companies analyses is not as high as for VC raised, the R^2 is still high, and the beta or exponent in the power function is extremely high. Also, it should be mentioned that the exponent in the service Non-AI startups is relatively high, even though the R^2 value is small, so the mathematical model is not very meaningful for this category.

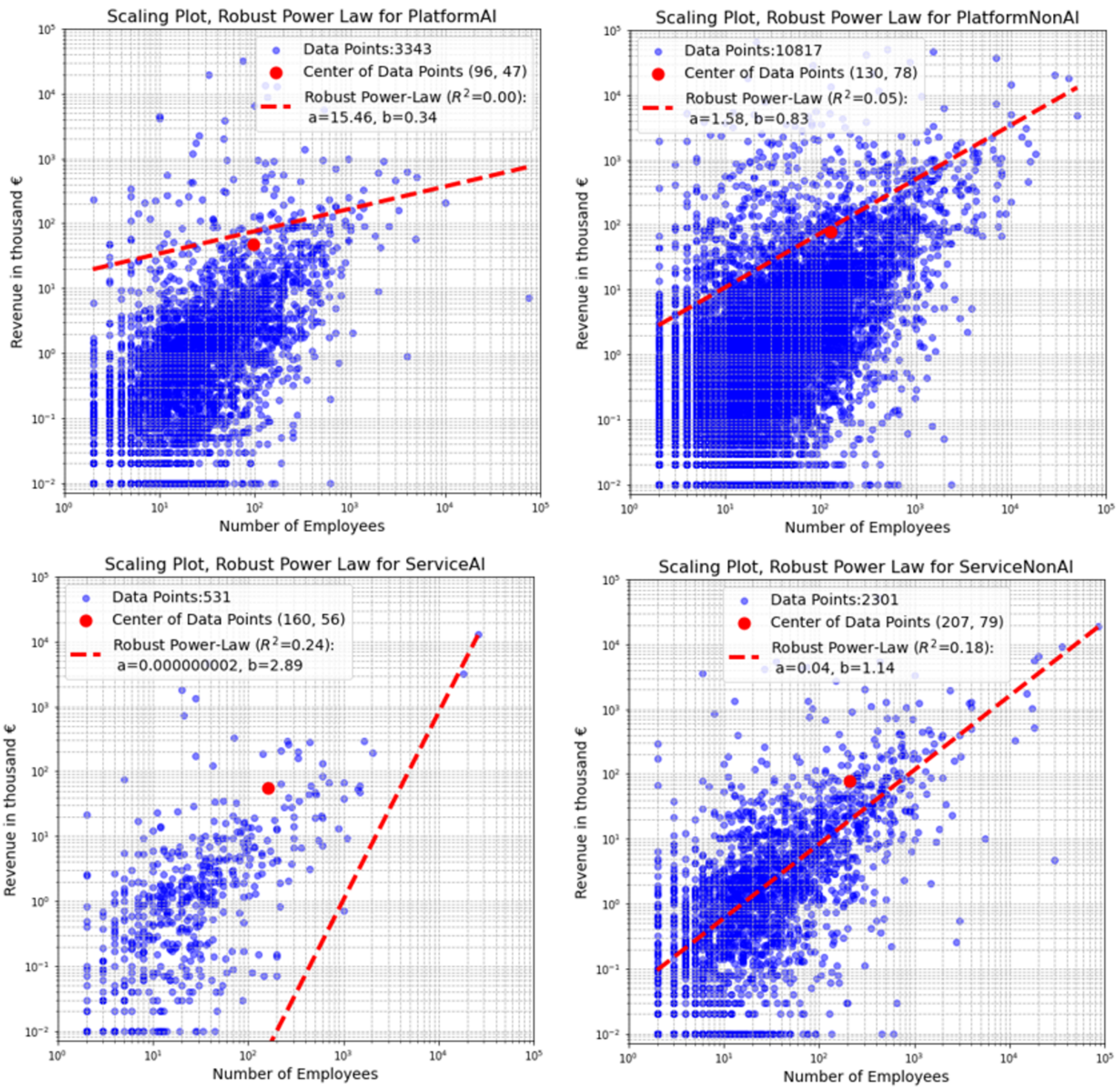


Figure 8.23: Revenue based on employees using least squares to determine power function, scaling analysis, own figure

The service AI function in Figure 8.23 may look suspicious, but for these plots, the very high revenue achievers are not shown. To keep the axis scaling the same and the plots comparable, with normal axes, this curve looks as shown in Figure 8.24

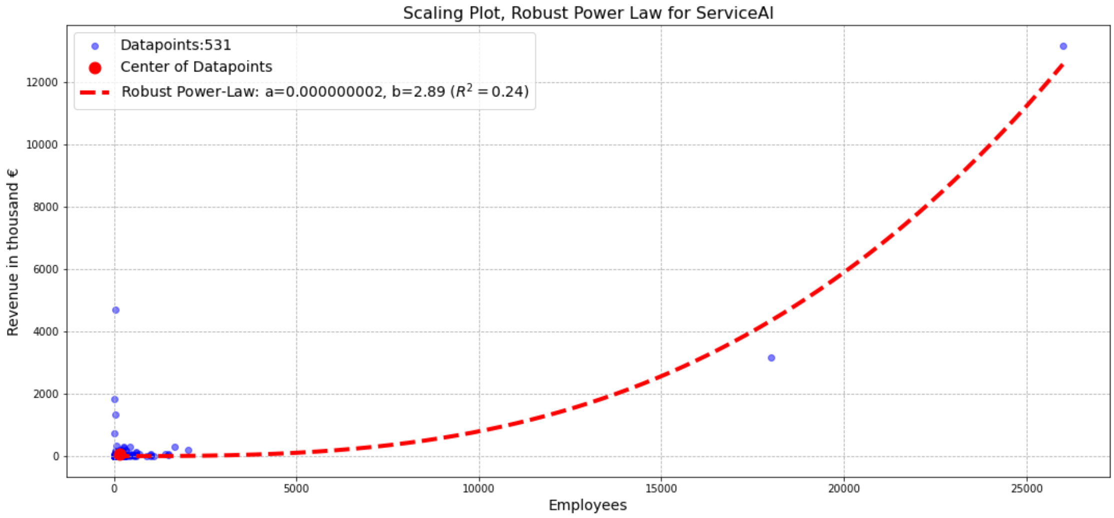


Figure 8.24: Revenue based on employees, scaling analysis, normal axis, own figure

8.6 Tables with Geographical Data

8.6.1 Platform AI

In Figure 8.25 the location data is presented:

hqCountry

Top 10 (r_squared) hqCountry mit Revenue per Employee: platformAI_rows.csv

hqCountry	average_revenue_per_employee	count	beta	r_squared
Estonia	0.044	36	2.71	0.97
Ireland	0.133	16	1.13	0.94
Japan	0.133	20	2.02	0.89
China	0.169	35	1.07	0.73
Finland	0.043	55	1.21	0.73
Poland	0.030	17	0.82	0.46
Canada	0.080	78	0.91	0.37
Sweden	0.085	64	0.93	0.36
South Korea	0.116	136	2.24	0.36
Italy	0.063	58	0.44	0.36

hqCity

Top 10 (r_squared) hqCity mit Revenue per Employee: platformAI_rows.csv

hqCity	average_revenue_per_employee	count	beta	r_squared
Pune	0.013	28	2.96	0.99
Sao Paulo	0.035	16	1.91	0.99
Los Angeles	0.144	19	1.62	0.99
Tallinn	0.046	31	2.69	0.98
Dublin	0.133	16	1.57	0.96
Hyderabad	0.030	30	1.55	0.94
Tokyo	0.138	18	2.04	0.89
Gurgaon	0.013	22	1.18	0.82
Helsinki	0.062	27	1.15	0.80
Madrid	0.073	23	0.73	0.46

hqRegion

Top 10 (r_squared) hqRegion mit Revenue per Employee: platformAI_rows.csv

hqRegion	average_revenue_per_employee	count	beta	r_squared
Oceania	0.140	32	0.23	0.04
Asia	0.151	776	0.75	0.02
Europe	0.499	711	0.30	0.00
Americas	1.273	1049	0.34	0.00
Middle East	13.985	76	0.00	0.00

hqStateProvince

hqStateProvince	average_revenue_per_employee	count	beta	r_squared
Telangana	0.033	25	1.56	0.94
Illinois	0.107	16	0.73	0.94
Colorado	0.096	28	0.97	0.60
Delaware	0.077	26	2.21	0.47
Virginia	0.856	25	1.11	0.38
Pennsylvania	0.067	18	0.42	0.33
Singapore	0.042	19	0.48	0.20
Ontario	0.091	38	0.71	0.20
England	0.730	121	0.60	0.15
Karnataka	0.024	124	0.86	0.13

Figure 8.25: Platform AI geographical data, own figure

8.6.2 Platform non-AI

In Figure 8.26, the location data is presented.

hqCountry

Top 10 (r_squared) hqCountry mit Revenue per Employee: platformNonAI_rows.csv

hqCountry	average_revenue_per_employee	count	beta	r_squared
Malaysia	0.071	70	2.58	0.99
Hungary	0.029	89	2.95	0.96
Austria	2.734	19	1.00	0.82
Switzerland	0.444	41	1.40	0.81
Thailand	0.089	35	2.77	0.81
Mexico	1.548	38	1.03	0.79
Norway	0.111	113	1.43	0.65
Singapore	0.153	313	1.10	0.46
Canada	0.479	183	0.84	0.45
Turkey	0.212	20	1.13	0.44

hqCity

Top 10 (r_squared) hqCity mit Revenue per Employee: platformNonAI_rows.csv

hqCity	average_revenue_per_employee	count	beta	r_squared
Hyderabad	0.014	33	3.32	1.00
Petaling Jaya	0.079	20	1.63	1.00
Turin	0.076	20	1.77	1.00
Prague	0.080	25	4.58	1.00
Mexico City	0.362	26	1.94	1.00
Jakarta	0.054	35	1.28	0.99
Noida	0.021	61	1.12	0.98
Palo Alto	0.251	19	1.10	0.96
Austin	0.471	78	1.82	0.86
Pune	0.056	52	1.39	0.83

hqRegion

Top 10 (r_squared) hqRegion mit Revenue per Employee: platformNonAI_rows.csv

hqRegion	average_revenue_per_employee	count	beta	r_squared
Africa	1.194	84	1.01	0.45
Oceania	0.801	122	0.50	0.06
Americas	2.407	2573	0.74	0.04
Middle East	1.135	108	0.30	0.02
Europe	1.198	2649	0.64	0.01
Asia	1.800	2311	0.51	0.00

hqStateProvince

Top 10 (r_squared) hqStateProvince mit Revenue per Employee: platformNonAI_rows.csv

hqStateProvince	average_revenue_per_employee	count	beta	r_squared
Michigan	0.116	22	1.92	1.00
Quebec	0.085	24	1.52	1.00
Telangana	0.025	30	3.32	1.00
Selangor	0.068	25	1.32	1.00
Nevada	0.086	18	2.30	0.99
Uttar Pradesh	0.019	70	1.15	0.98
Minnesota	0.083	23	5.02	0.98
Virginia	0.167	34	4.79	0.65
England	0.588	384	0.86	0.62
Tamil Nadu	0.031	46	0.93	0.59

Figure 8.26: Platform non-AI geographical data, own figure

8.6.3 Service AI

In Figure 8.27, the location data is presented.

hqCountry

Top 10 (r_squared) hqCountry mit Revenue per Employee: serviceAI_rows.csv

hqCountry	average_revenue_per_employee	count	beta	r_squared
India	0.015	43	10.36	0.97
United States	1.055	101	1.30	0.74
Sweden	0.076	18	0.79	0.64
China	0.400	19	0.56	0.29
South Korea	0.277	28	0.11	0.00

hqCity

Top 10 (r_squared) hqCity mit Revenue per Employee: serviceAI_rows.csv

hqCity	average_revenue_per_employee	count	beta	r_squared
Seoul	0.397	18	0.14	0.0

hqRegion

Top 10 (r_squared) hqRegion mit Revenue per Employee: serviceAI_rows.csv

hqRegion	average_revenue_per_employee	count	beta	r_squared
Americas	0.900	120	1.35	0.74
Asia	0.162	114	0.63	0.26
Europe	1.835	122	0.25	0.00

hqStateProvince

Top 10 (r_squared) hqStateProvince mit Revenue per Employee: serviceAI_rows.csv

hqStateProvince	average_revenue_per_employee	count	beta	r_squared
California	2.956	32	-0.09	0.0

Figure 8.27: Service AI geographical data, own figure

8.6.4 Service non-AI

In Figure 8.28, the location data is presented.

hqCountry

Top 10 (r_squared) hqCountry mit Revenue per Employee: serviceNonAI_rows.csv

hqCountry	average_revenue_per_employee	count	beta	r_squared
South Korea	0.155	68	1.67	0.98
Hungary	0.023	24	0.79	0.69
United States	1.397	312	0.86	0.68
Australia	0.175	24	0.96	0.66
Finland	0.066	45	1.06	0.66
Germany	0.441	18	0.81	0.48
Italy	0.162	43	0.45	0.27
United Kingdom	0.553	54	0.90	0.26
France	0.186	82	1.13	0.23
Canada	0.433	44	2.31	0.23

hqCity

Top 10 (r_squared) hqCity mit Revenue per Employee: serviceNonAI_rows.csv

hqCity	average_revenue_per_employee	count	beta	r_squared
San Francisco	0.152	21	9.12	0.93
Bangalore	0.023	23	0.59	0.65
Mumbai	0.050	21	1.85	0.51
Helsinki	0.093	17	0.87	0.43
Seoul	0.145	45	1.27	0.43
Milan	0.244	18	0.27	0.23
London	0.667	38	0.87	0.23
Stockholm	0.070	35	0.32	0.16
Oslo	0.058	18	0.54	0.09
Singapore	0.288	45	0.46	0.05

hqRegion

Top 10 (r_squared) hqRegion mit Revenue per Employee: serviceNonAI_rows.csv

hqRegion	average_revenue_per_employee	count	beta	r_squared
Oceania	0.155	28	1.06	0.64
Americas	1.209	379	0.79	0.56
Middle East	1.027	18	0.34	0.08
Europe	0.368	526	0.51	0.07
Asia	0.572	384	0.49	0.03

hqStateProvince

Top 10 (r_squared) hqStateProvince mit Revenue per Employee: serviceNonAI_rows.csv

hqStateProvince	average_revenue_per_employee	count	beta	r_squared
Colorado	0.081	20	1.40	1.00
Texas	0.248	19	1.51	0.98
Haryana	0.088	20	9.16	0.54
Maharashtra	0.067	29	1.33	0.45
England	0.559	53	0.91	0.26
Florida	1.093	18	0.45	0.11
California	2.290	87	0.54	0.10
New York	0.172	35	0.37	0.06
Karnataka	0.094	46	0.24	0.01

Figure 8.28: Service non-AI geographical data, own figure

8.7 Tables with Industry Data

I define an industry as a collection of more than 20 companies from an area with the same vertical label. In Table 8.5, the industry analysis data is presented.

VC Raised

I define an industry as a collection of more than 20 companies from an area with the same vertical entry. In Table 8.6, the industry analysis data is presented.

8.7.1 Verticals

Verticals are like industries categorization, but less formal Pichbook (2024). Here, new industries are also present. I define an industry as a collection of more than 20 companies from an area with the same vertical entry. $R^2 > 0.78$ and $\geq$1.

Revenue

I define an industry as a collection of more than 20 companies from an area with the same vertical entry. $R^2 > 0.78$ and $\beta > 1$, summarized in Table 8.7.

Table 8.5: Overview of verticals for all four startup categories

Verticals	**R-squared**	**Count**	**Beta**	**Average Revenue per Employee**
platformAI_rows.csv				
Advanced Manufacturing	0.680	51	0.808	0.076240
FinTech	0.563	34	1.000	0.050835
serviceAI_rows.csv				
Advanced Manufacturing	0.883	22	0.954	0.092814
Artificial Intelligence & Machine Learning	0.002	276	0.263	0.777022
serviceNonAI_rows.csv				
Internet of Things	0.997	62	13.078	0.142789
Big Data	0.964	30	0.623	0.244135
Marketing Tech	0.842	26	4.089	0.046247
CleanTech	0.796	88	1.018	0.336289
Gaming	0.470	27	1.252	0.693513
Manufacturing	0.378	59	0.537	0.142784
platformNonAI_rows.csv				
AudioTech	0.927	54	1.436	0.787700
Advanced Manufacturing	0.850	55	1.018	0.185392
Climate Tech	0.845	29	0.970	0.287218
AgTech	0.639	102	0.865	0.180279
Industrials	0.423	110	0.615	0.224324
Manufacturing	0.386	21	0.554	0.086170
CleanTech	0.362	147	0.694	0.497310
Supply Chain Tech	0.321	38	0.804	0.345317
LOHAS & Wellness	0.313	75	2.266	0.489869
HealthTech	0.281	114	1.079	0.212836

Table 8.6: Overview of verticals for VC raised

Verticals	**R-squared**	**Count**	**Beta**	**Average VC Raised per Employee**
		platformAI_rows.csv		
AdTech	0.996125	94	2.745475	0.206690
SaaS	0.945354	46	1.863924	0.300471
Advanced Manufacturing	0.819304	51	0.810534	0.272617
		serviceAI_rows.csv		
Advanced Manufacturing	0.249231	22	0.408426	0.213229
Artificial Intelligence & Machine Learning	NaN (Optimal parameters not found)	276	0.359577 (average)	0.370237 (average)
		serviceNonAI_rows.csv		
Big Data	0.990127	30	0.668619	0.297304
CleanTech	0.982585	88	2.157811	0.311417
Robotics and Drones	0.932940	21	2.647596	0.265068
SaaS	0.931695	77	0.907983	0.167015
Gaming	0.827579	27	1.709789	0.456834
		platformNonAI_rows.csv		
Advanced Manufacturing	0.991808	55	1.986101	0.249188
Climate Tech	0.968612	29	1.196466	0.638393
Esports	0.959413	96	2.641297	0.343533
AudioTech	0.948331	54	1.349601	0.274072
B2B Payments	0.859707	66	1.462428	0.372551
Cannabis	0.847977	35	1.229082	0.404583
Cybersecurity	0.833480	158	1.739326	0.346284
HR Tech	0.762204	279	1.039650	0.155992

Table 8.7: Overview of verticals for revenue

Verticals	**R-squared**	**Count**	**Beta**	**Average Revenue per Employee**
		serviceNonAI_rows.csv		
Internet of Things	0.997	62	13.078	0.142789
Marketing Tech	0.842	26	4.089	0.046247
CleanTech	0.796	88	1.018	0.336289
		platformNonAI_rows.csv		
AudioTech	0.927	54	1.436	0.787700
Advanced Manufacturing	0.850	55	1.018	0.185392

VC Raised

I define an industry as a collection of more than 20 companies from an area with the same vertical entry. $R^2 > 0.78$ and $\beta > 1$, summarized in Table 8.8 You can see that service startups are not present in this table.

I) Verticals for AI Startups: Platform AI & Service AI

In this section, I investigate if the usage of AI in an industry field makes a difference in the scaling analysis. Like in the previous sections, I used employees and revenue as the key variables for the scaling Analysis. Table 8.9 shows that scaling analysis for different industries with the AI-Startups.

Table 8.8: Verticals overview of average VC raised for platforms

Verticals	**R-squared**	**Count**	**Beta**	**Average VC Raised per Employee**
platformAI_rows.csv				
AdTech	0.996125	94	2.745475	0.206690
SaaS	0.945354	46	1.863924	0.300471
platformNonAI_rows.csv				
Advanced Manufacturing	0.991808	55	1.986101	0.249188
Climate Tech	0.968612	29	1.196466	0.638393
Esports	0.959413	96	2.641297	0.343533
AudioTech	0.948331	54	1.349601	0.274072
B2B Payments	0.859707	66	1.462428	0.372551
Cannabis	0.847977	35	1.229082	0.404583
Cybersecurity	0.833480	158	1.739326	0.346284

Table 8.9: Top 10 (R^2) verticals for Revenue: Platform AI & Service AI

Verticals of Platform AI & Service AI	Average Revenue per Employee	Count	Beta	R^2
FinTech	0.051	35	0.990708	0.563096
Advanced Manufacturing	0.081	73	0.785320	0.490467
SaaS	2.258	47	0.445311	0.023639
AdTech	1.896	105	0.316552	0.009274
Artificial Intelligence & Machine Learning	0.874	2395	0.322842	0.001561
AgTech	1.320	77	0.149023	0.001319
Big Data	8.288	74	0.039153	0.000100

In the following Table 8.10, the same scaling Analysis for the Non-AI Startups is shown.

Table 8.10: Top 10 (R^2) verticals for VC raised: Platform AI & Service AI

Verticals of Platform AI & Service AI	Average VC raised per Employee	Count	Beta	R^2
AdTech	0.215	105	2.885	0.991
SaaS	0.295	47	1.863	0.945
Advanced Manufacturing	0.255	73	0.657	0.509
AgTech	0.331	77	0.563	0.245
Big Data	0.254	74	0.475	0.201
FinTech	0.475	35	0.382	0.174
Artificial Intelligence & Machine Learning	0.319	2395	0.439	0.086

II) Verticals for Non-AI Startups: Platform Non-AI & Service Non-AI

I define an industry as a collection of more than 20 companies from an area with the same vertical label for $R^2 > 0.78$ and $\beta > 1$.

The following tables contain a list of all verticals of non-AI startup industries. The first Table 8.11 is for revenue, while the second Table 8.12 is for VC raised.

Table 8.11: Top 10 (R^2) verticals: Platform Non-AI & Service Non-AI (Non-AI startups)

Verticals of Platform Non-AI & Service Non-AI	Average Revenue per Employee	Count	Beta	R^2
AudioTech	0.634	69	1.383	0.919
Climate Tech	0.248	38	3.313	0.815
Advanced Manufacturing	0.156	85	0.941	0.612
CleanTech	0.437	235	0.892	0.505
Robotics and Drones	0.058	41	0.784	0.476
AgTech	0.466	121	1.115	0.467
Industrials	0.221	144	0.623	0.410
Manufacturing	0.127	80	0.568	0.378
Supply ChAIn Tech	0.321	42	0.795	0.354
LOHAS & Wellness	0.426	91	2.255	0.317

In the following table, is shown the same scaling Analysis for the Non-AI Startups. You can see that the verticals differ. The tables are ranked by the R^2, so the precision of the model, or as West (2019) would interpret it, the innovations.

Table 8.12: Top 10 (R^2) Verticals VC raised for Platform Non-AI and Service Non-AI

Verticals of Platform Non-AI & Service Non-AI	Average VC raised per Employee	Count	Beta	R^2
Esports	0.339	108	2.637	0.956
Climate Tech	0.542	38	2.774	0.955
AudioTech	0.322	69	1.374	0.946
Advanced Manufacturing	0.248	85	1.228	0.861
Cannabis	0.537	41	1.210	0.848
B2B Payments	0.409	69	1.204	0.828
Cybersecurity	0.332	172	1.987	0.824
HealthTech	0.191	121	1.122	0.759
HR Tech	0.238	302	1.055	0.744
CloudTech & DevOps	0.309	176	0.844	0.676

8.7.2 I) primaryindustryCode for AI Startups: Platform AI & Service AI)

I define an industry as a collection of more than 20 companies from an area with the same vertical label for $R^2 > 0.78$ and $\beta > 1$.

Table 8.13 is showing revenue, while Table 8.14 shows VC raised. Both are ranked by R^2 and show the top 10 results.

Table 8.13: Top 10 (R^2) Revenue primaryindustryCode: Platform AI & Service AI

Primary Industry Code (Platform AI & Service AI)	Average Revenue per Employee	Count	Beta	R^2
Other Healthcare Technology Systems	0.047	37	0.785	0.980
Information Services (B2C)	0.521	37	0.803	0.769
IT Consulting and Outsourcing	0.088	38	0.512	0.505
Communication Software	0.167	33	0.817	0.233
Multimedia and Design Software	0.102	65	0.857	0.167
Database Software	0.341	40	0.444	0.127
Human Capital Services	0.090	57	0.351	0.094
Educational Software	0.066	96	0.466	0.067
Financial Software	0.626	245	0.550	0.012
Entertainment Software	1.285	43	-0.430	0.009

Table 8.14: Top 10 (R^2) VC raised primaryindustryCode: Startups for Platform AI & Service AI

Primary Industry Code (Platform AI & Service AI)	Average VC raised per Employee	Count	Beta	R^2
Information Services (B2C)	0.191	37	1.968	0.999
Other Hardware	0.376	56	1.826	0.953
Software Development Applications	0.329	47	2.465	0.886
Other Healthcare Technology Systems	0.247	37	0.513	0.751
Educational Software	0.174	96	0.827	0.726
Electronic Equipment and Instruments	0.203	76	0.777	0.675
Financial Software	0.497	245	1.098	0.571
Human Capital Services	0.144	57	0.607	0.459
Database Software	0.343	40	0.429	0.441
Business/Productivity Software	0.271	1443	1.213	0.431

8.7.3 II) primaryindustryCode for Non-AI Startups: Platform Non-AI & Service Non-AI

I define an industry as a collection of more than 20 companies from an area with the same vertical label for $R^2 > 0.78$ and $\beta > 1$. The first table is revenue, the second is VC raised. Both are ranked by R^2.

Table 8.15: Top 10 (R^2) Revenue primaryindustryCode: Non-AI Startups for Platform Non-AI & Service Non-AI

Primary Industry Code (Platform Non-AI & Service Non-AI)	Average Revenue per Employee	Count	Beta	R^2
Database Software	0.107868	48	1.449069	0.994309
IT Consulting and Outsourcing	0.242260	86	1.044987	0.968623
Electronic Equipment and Instruments	0.377309	146	1.263098	0.962989
Internet RetAIl	0.899059	66	0.878925	0.825552
Publishing	0.699262	46	2.933920	0.780445
Education and Training Services (B2B)	0.169246	53	2.369579	0.732349
Other Healthcare Technology Systems	0.173229	66	1.312432	0.710285
Other Commercial Services	2.585162	58	1.988032	0.575024
Other Services (B2C Non-Financial)	0.634077	202	0.687612	0.288251
Social/Platform Software	1.684044	195	1.550229	0.223391

Table 8.16: Top 10 (R^2) VC raised primaryindustryCode: Non-AI Startups for Platform Non-AI & Service Non-AI

Primary Industry Code (PlatformNon AI & Service Non-AI)	Average VC raised per Employee	Count	Beta	R^2
IT Consulting and Outsourcing	0.177	86	1.563	0.996
Electronic Equipment and Instruments	0.251	146	0.844	0.962
Automotive	0.311	212	2.195	0.924
Communication Software	0.315	67	3.323	0.831
Other Hardware	0.268	106	1.291	0.830
Application Software	0.154	143	0.718	0.813
Clinics/Outpatient Services	0.148	72	1.359	0.809
Internet Retail	0.427	66	1.517	0.756
Software Development Applications	0.338	108	1.514	0.719
Education and Training Services (B2B)	0.172	53	1.079	0.696

8.7.4 I) primaryindustryGroup for AI Startups: Platform AI & Service AI

I define an industry as a collection of more than 20 companies from an area with the same vertical label for $R^2 > 0.78$ and $\beta > 1$. The first table is revenue, the second is VC raised. Both are ranked by R^2.

Table 8.17: Top 10 (R^2) Revenue primaryindustryGroup: Platform AI & Service AI

Primary Industry Group	Average Revenue per Employee	Count	Beta	R^2
Healthcare Technology Systems	0.154	55	0.628	0.815
Media	0.449	46	0.835	0.716
Other Financial Services	0.075	34	0.897	0.709
Commercial Products	0.119	31	0.621	0.432
IT Services	0.082	47	0.509	0.422
Services (Non-Financial)	0.308	41	0.272	0.018
Commercial Services	3.690	324	0.417	0.007
Software	0.632	2164	0.348	0.005
Computer Hardware	4.081	144	-0.027	0.000

Table 8.18: Top 10 (R^2) VC raised primaryindustryGroup: Platform AI & Service AI

Primary Industry Group	Average VC raised per Employee	Count	Beta	R^2
Media	0.183	46	1.966	0.998
Commercial Products	0.270	31	2.315	0.968
Healthcare Technology Systems	0.237	55	0.527	0.755
Commercial Services	0.352	324	1.056	0.546
Computer Hardware	0.274	144	0.760	0.427
Other Financial Services	0.742	34	0.490	0.322
Services (Non-Financial)	0.234	41	0.461	0.240
Software	0.297	2164	0.464	0.142
IT Services	0.144	47	0.293	0.108

8.7.5 II) primaryindustryGroup for Non-AI Startups: Platform Non-AI & Service Non-AI

I define an industry as a collection of more than 20 companies from an area with the same vertical label for $R^2 > 0.78$ and $\beta > 1$.

Table 8.19: Top 10 (R^2) Revenue primaryindustryGroup: Platform Non-AI & Service Non-AI

Primary Industry Group (Platform Non-AI & Service Non-AI)	Average Revenue per Employee	Count	Beta	R^2
Computer Hardware	0.791	280	1.216	0.957
IT Services	0.544	121	1.070	0.951
Healthcare Technology Systems	0.180	105	1.373	0.760
Consumer Non-Durables	0.128	41	0.417	0.315
Insurance	0.086	58	0.672	0.203
Retail	1.024	113	0.690	0.202
Commercial Products	0.201	91	0.363	0.192
Consumer Durables	0.304	158	0.655	0.184
Media	1.145	325	0.941	0.176
Healthcare Services	0.083	106	0.727	0.132

Table 8.20: Top 10 (R^2) VC raised primaryindustryGroup: Platform Non-AI & Service Non-AI

Primary Industry Group (Platform Non-AI & Service Non-AI)	Average VC raised per Employee	Count	Beta	R^2
Communications and Networking	0.330	47	3.670	0.998
IT Services	0.265	121	1.536	0.993
Computer Hardware	0.264	280	0.746	0.902
Transportation	0.310	222	1.207	0.830
Healthcare Services	0.134	106	1.392	0.786
Semiconductors	0.374	37	0.574	0.767
Retail	0.361	113	0.543	0.427
Consumer Non-Durables	0.167	41	0.791	0.415
Media	0.268	325	0.702	0.347
Commercial Services	0.224	1140	0.655	0.311

8.7.6 I) primaryindustrySector for AI Startups: Platform AI & Service AI

I define an industry as a collection of more than 20 companies from an area with the same vertical label for $R^2 > 0.78$ and $\beta > 1$.

Table 8.21: Top 10 (R^2) Revenue primaryindustrySector: Platform AI & Service AI

Primary Industry Sector (Platform AI & Service AI)	Average Revenue per Employee	Count	Beta	R^2
Healthcare	0.129	77	0.686	0.795
Financial Services	0.194	58	0.875	0.526
Consumer Products and Services (B2C)	0.284	156	0.968	0.269
Business Products and Services (B2B)	3.306	363	0.429	0.006
Information Technology	0.828	2368	0.341	0.004

Table 8.22: Top 10 (R^2) VC raised primaryindustrySector: Platform AI & Service AI

Primary Industry Sector (Platform AI & Service AI)	Average VC raised per Employee	Count	Beta	R^2
Healthcare	0.248	77	0.533	0.708
Business Products and Services (B2B)	0.349	363	0.921	0.549
Consumer Products and Services (B2C)	0.265	156	0.697	0.391
Financial Services	0.656	58	0.591	0.369
Information Technology	0.295	2368	0.468	0.146

8.7.7 II) primaryindustrySector for Non-AI Startups: Platform Non-AI & Service Non-AI

I define an industry as a collection of more than 20 companies from an area with the same vertical label for $R^2 > 0.78$ and $\beta > 1$.

Table 8.23: Top 10 (R^2) Revenue primaryindustrySector: Platform Non-AI & Service Non-AI

Primary Industry Sector (Platform Non-AI & Service Non-AI)	Average Revenue per Employee	Count	Beta	R^2
Healthcare	0.144	238	1.954	0.572
Consumer Products and Services (B2C)	0.786	1548	0.748	0.041
Financial Services	2.159	408	0.554	0.020
Information Technology	1.768	5723	0.725	0.016
Business Products and Services (B2B)	1.942	1267	0.461	0.013

Table 8.24: Top 10 (R^2) VC raised primaryindustrySector: Platform Non AI & Service Non-AI

Primary Industry Sector (Platform Non-AI & Service Non-AI)	Average VC raised per Employee	Count	Beta	R^2
Consumer Products and Services (B2C)	0.263	1548	0.840	0.364
Information Technology	0.316	5723	0.775	0.347
Business Products and Services (B2B)	0.256	1267	0.561	0.243
Financial Services	0.639	408	0.618	0.069
Healthcare	0.184	238	NaN	NaN

8.8 MLP Regressor: Optimal Hidden Layers

The MLP model has the tuning parameter of hidden layers. If the optimal hidden layers that are extracted from the above calculations are used in the cross-validation, there are quite bad R^2. This just means that these "optimal hidden layers" are not universally optimal. Therefore, it is beneficial to look at different hidden layer sizes and cross-validate every generated MLP-Regressor- model afterward.

I plotted the hidden layer sizes regarding 5-fold-cross-validated R^2 in Figure 8.29 to show that it is complicated to cleverly guess the number of hidden layers:

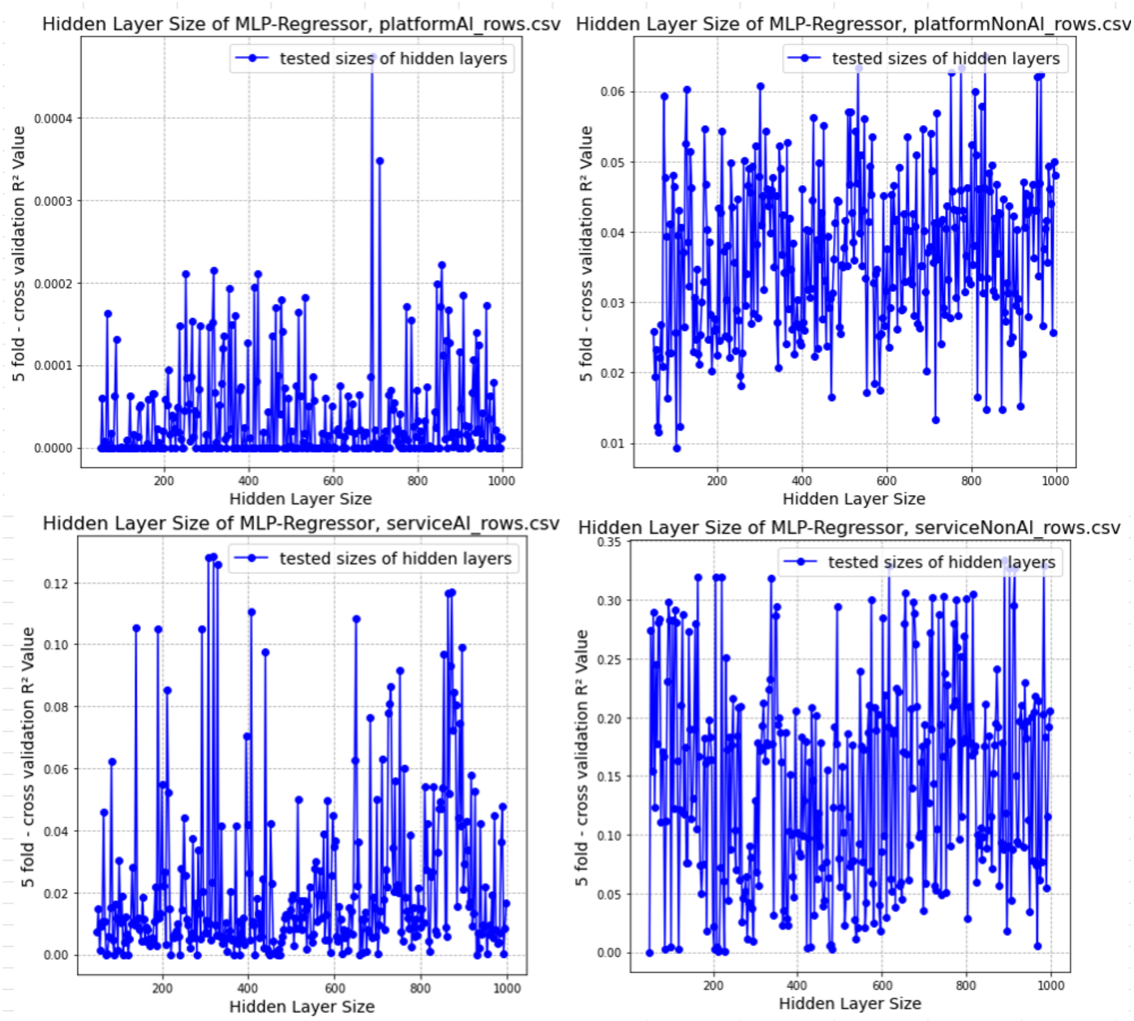


Figure 8.29: MLP: hidden layers cross-validation tested, own figure

It's chaotic, and the cross-validated R^2 is not automatically better if more hidden layers are used. But still, the overall best R^2 results I tested are presented in the main table below.

That's why I decided to look at my smallest data point category, service AI, with 531 data points, and use half of it as the upper limit for the MLP network hidden layer size. So, only half of all data points could be maximum "memorized" in the structure.

So, the maximum of 265 hidden layers could be sufficient for the service AI test case. I'm assuming that the complexity of the revenue to employees is not higher in the other cases; that's why the 265-hidden layer limit is also applied to them; of course, the percentage calculation is much lower for the different test cases because 265 data points are not half of the dataset: service non-AI, platform AI and platform non-AI should have therefore lower percentages "memorized" in the NN. The results are shown in Figure 8.30.

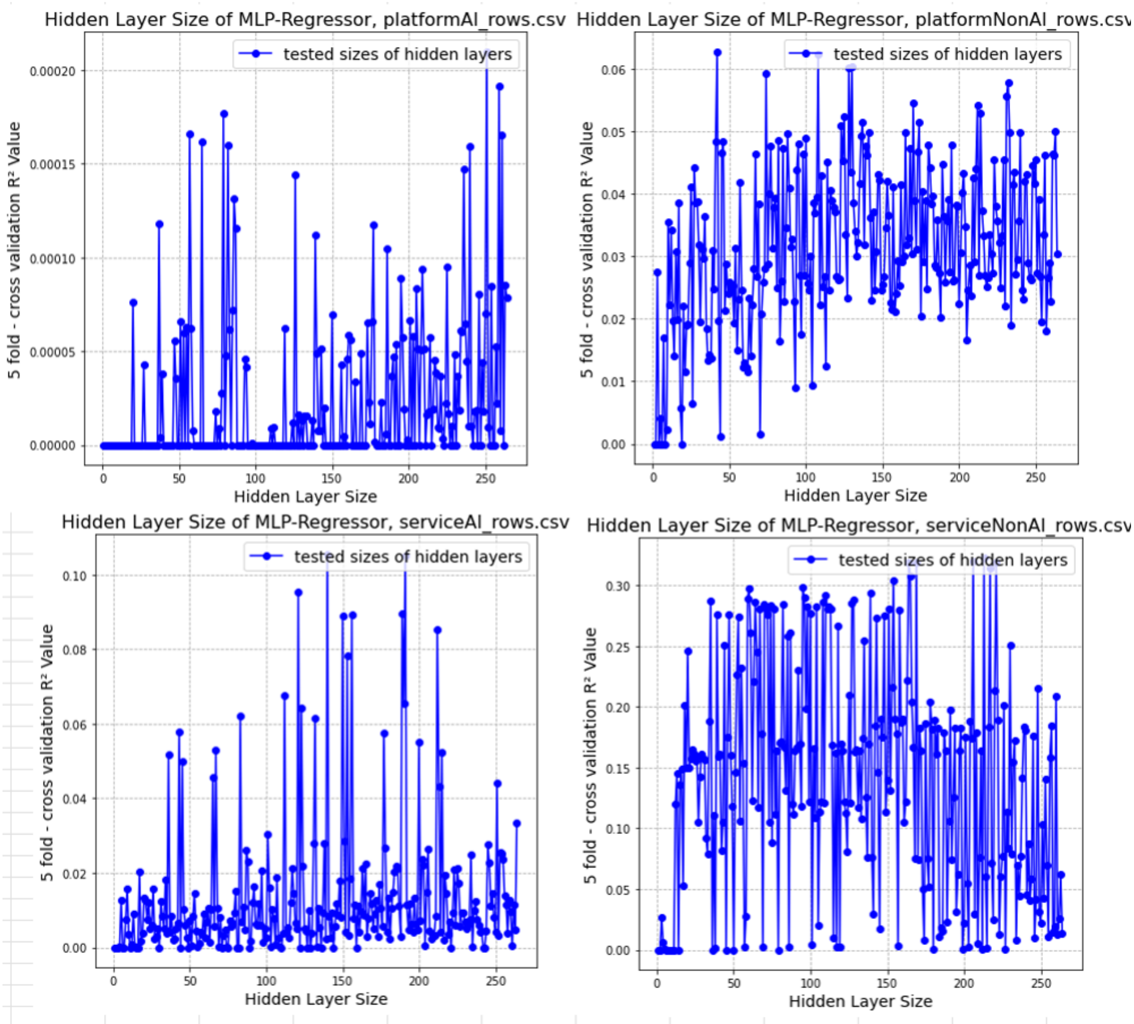


Figure 8.30: MLP: limited to 265 hidden layers, cross-validation tested R^2, own figure

In Table 8.25, the best models with their best-hidden layer sizes are presented. The graphs in Figure 8.30 should demonstrate the context of the best values - mainly that the values are not approximated or guessed but instead calculated, and the neighbor values are not necessarily good.

Table 8.25: Best hidden layer sizes and R^2 values for hidden layer size limit = 1000

Robust Model	**Service AI**	**Service Non-AI**
Best hidden layer size	320	890
R^2 Value	0.13	0.33
	Platform AI	**Platform Non-AI**
Best hidden layer size	692	833
R^2 Value	0.00	0.06

Overall, the mentioned concerns regards the over fitting MLP-regressor, you can

Table 8.26: Best hidden layer sizes and R^2 values for hidden layer size limit = 265

Robust Model	**Service AI**	**Service Non-AI**
Best hidden layer size	140	213
R^2 Value	0.11	0.32
	Platform AI	**Platform Non-AI**
Best hidden layer size	251	42
R^2 Value	0.00	0.06

say, based on the big testing runs with maximum to 1000 layers, is not supporting the "memorizing" theory. On the other hand, the reduced hidden layer size to 265 shows comparable results, as can be seen in Figure 8.26, but with far fewer computing resources needed.

8.9 Visualization of Mathematical Models: Not Robust and Robust

8.9.1 Ordinary Least Squares in log-log Axis

In contrast to applying the power scaling function to the original startup data, another approach is to calculate the logarithm first and then perform a linear regression analysis on log-log data. This procedure has the following consequences.

First, the power function is applied to the original data. In this case, the optimization will yield a lower R^2 because it penalizes outliers quadratically. However, if the ordinary least squares (OLS) optimization is performed on the logarithmic data, the conversion will result in a higher R^2 in the log model. This is because the quadratic penalty is now applied to the logarithmic data, which reduces the impact of outliers on the R^2 value.

The linear function derived from the OLS regression of the log-log model, as illustrated in Figures 8.31 and 8.32, accurately represents the logarithmized data and shows results that are visually comparable to the analysis conducted by West (2019). Additionally, the R^2 values are relatively high. However, when we exponentiate the log-log model using base e to transform it back, the function does not align as well with the original data, as the power function previously did.

This difference between optimizing the log-log model and optimizing a power function for the original data does not mean that the results are not comparable.

In fact, the exponent b can be interpreted similarly in both approaches. Since this work emphasizes scaling analysis, the exponent is the primary variable of interest in these formulas.

Figure 8.24, on the other hand, shows that optimizing for the original data is a good idea, not for the log-transformed one. It is also crucial to consider what the data represents in the real world. The outliers in the data and in Figure 8.24 are large companies with high income and many employees. It is reasonable to assume that such huge companies will likely have reliable key figures and, therefore, should be kept in the dataset. In summary, the work with the R^2 on the original data - which is not transformed - is better.

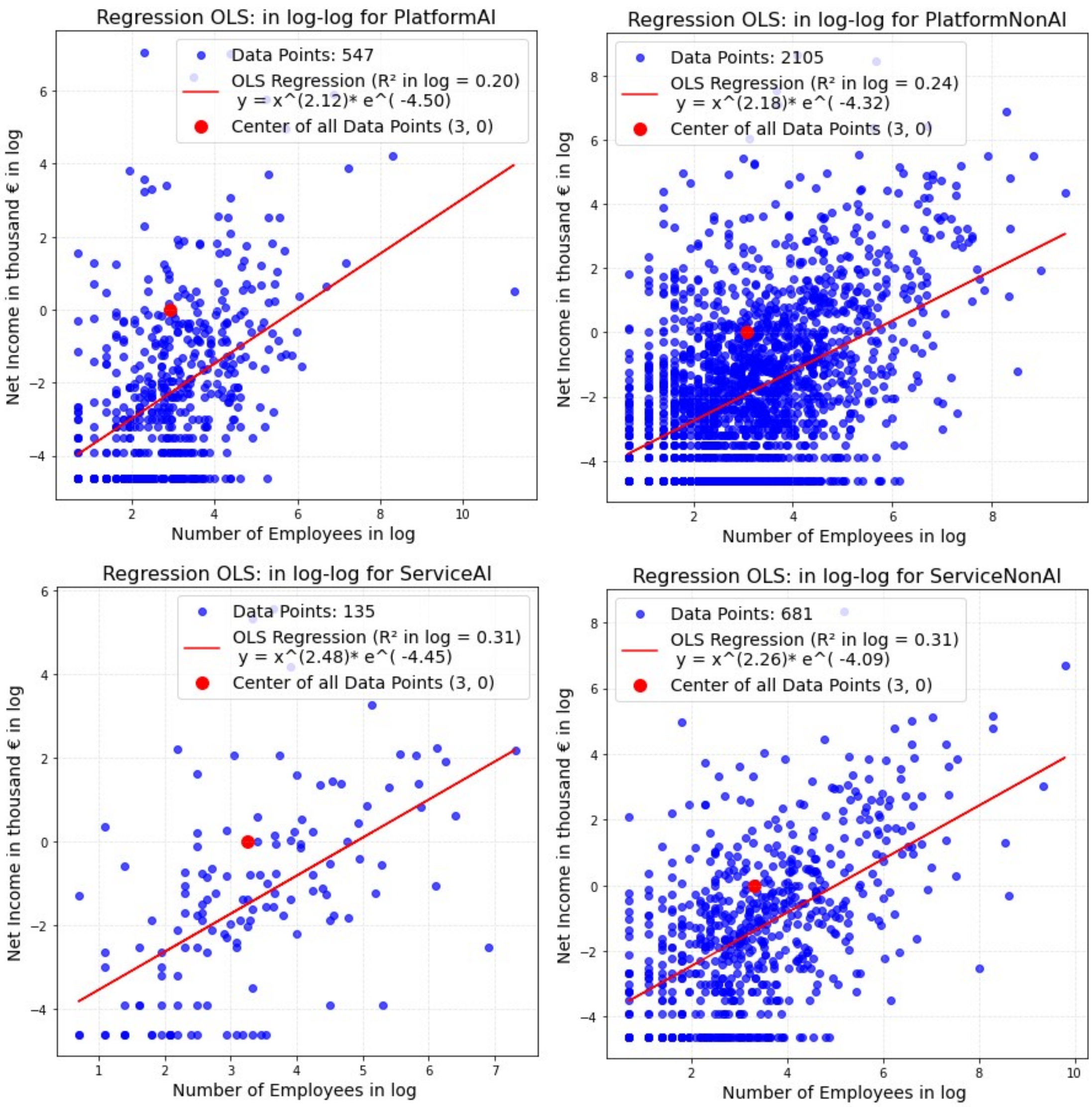


Figure 8.31: Net Income based on employees, OLS Regression, with log-log axes, own figure

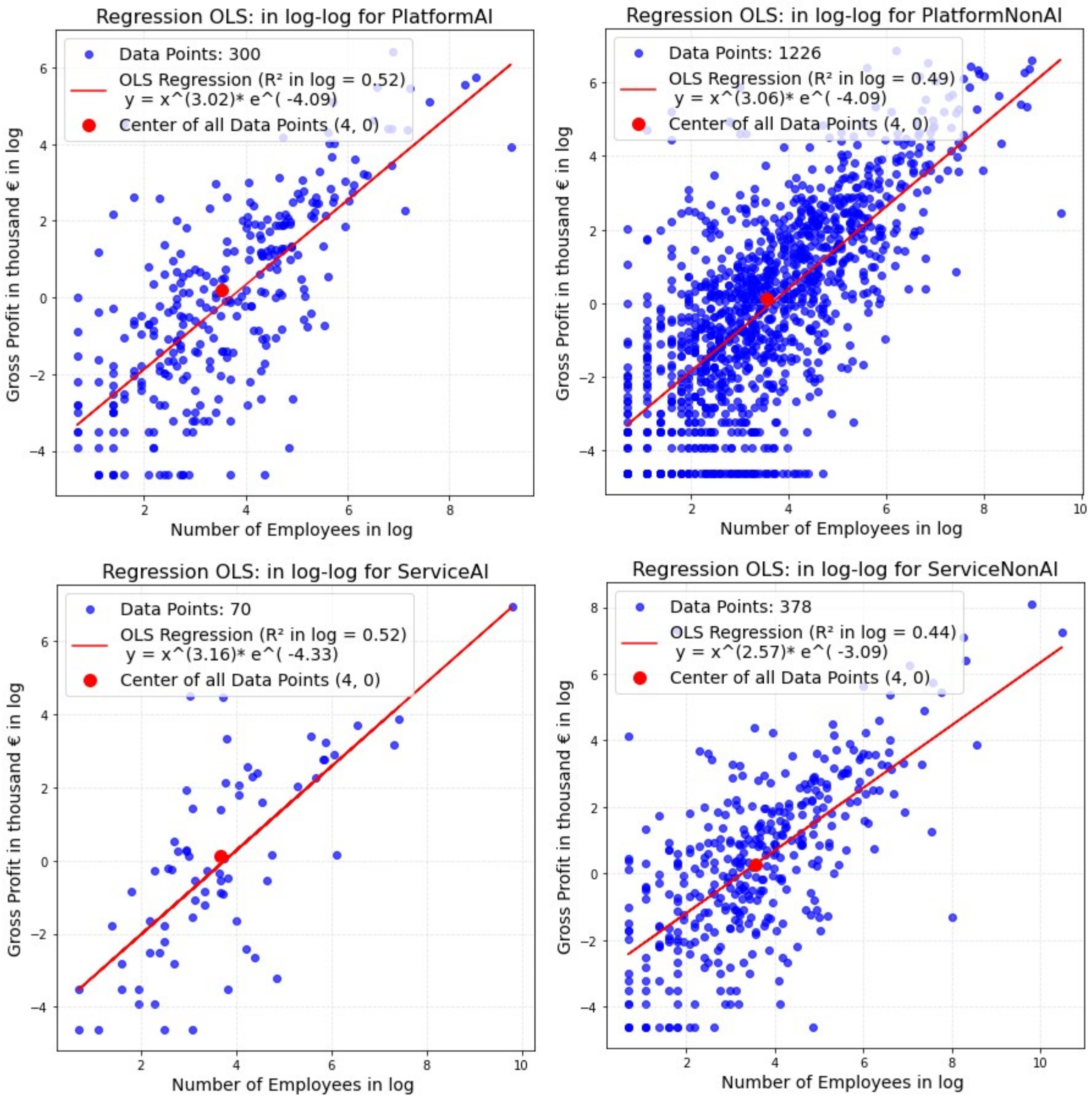


Figure 8.32: Gross Profit based on employees, OLS Regression, with log-log axes, own figure

8.9.2 Support Vector Regression

The non-robust SVR is shown in Figure 8.33. You can see the curve differs from the startup category.

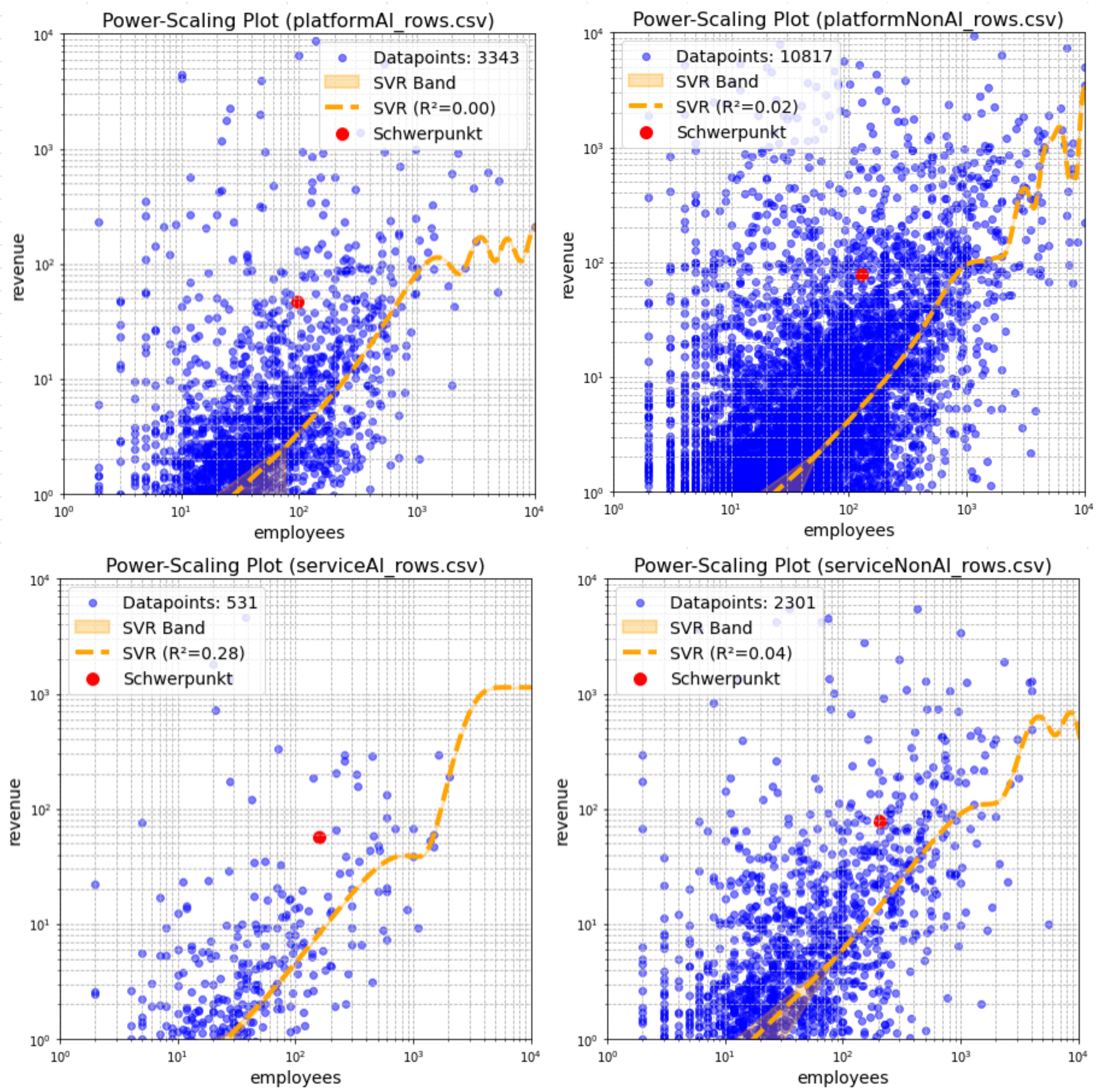


Figure 8.33: Revenue based on employees, SVR, own figure

Here the robust function is shown in Figure 8.34. To make the R^2 higher, the top 1% revenue achievers are excluded.

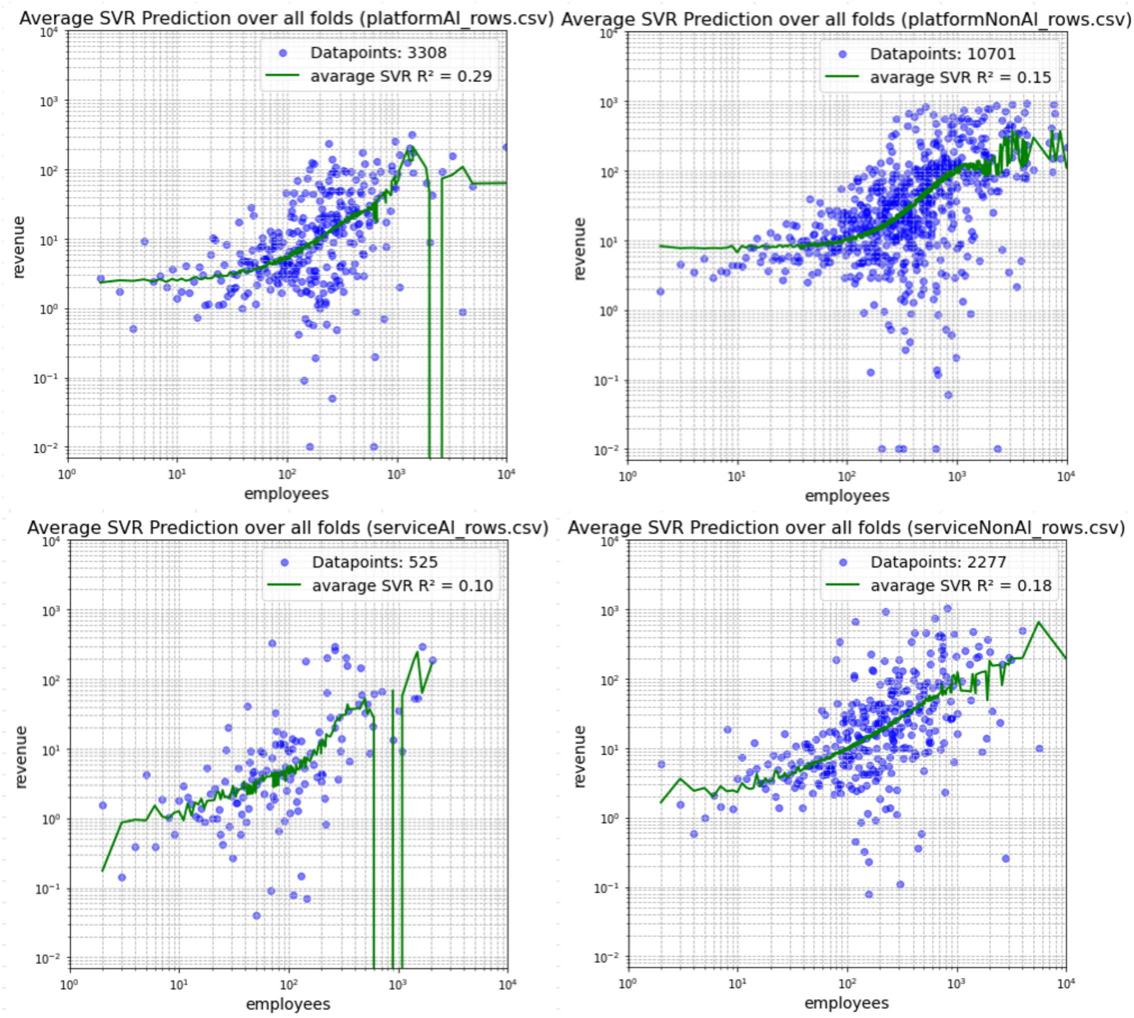


Figure 8.34: Revenue based on employees, SVR, robust average prediction, own figure

8.9.3 Polynomial Regression

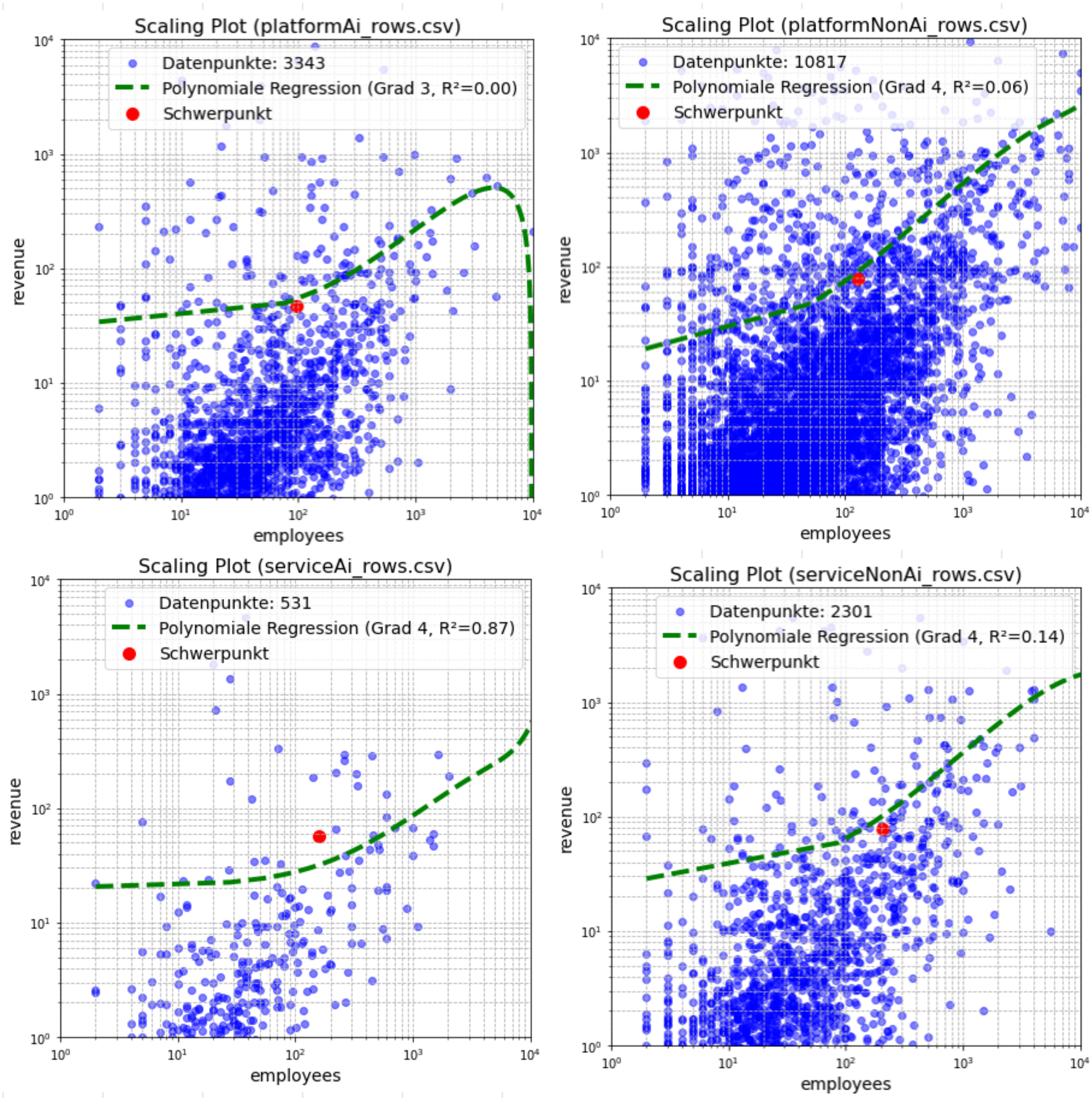


Figure 8.35: Revenue based on employees, polynomial regression, own figure

If you try to make a robust version of this mathematical prediction of the results in Figure 8.35, the graphical representation of the results can be seen in Figure 8.36

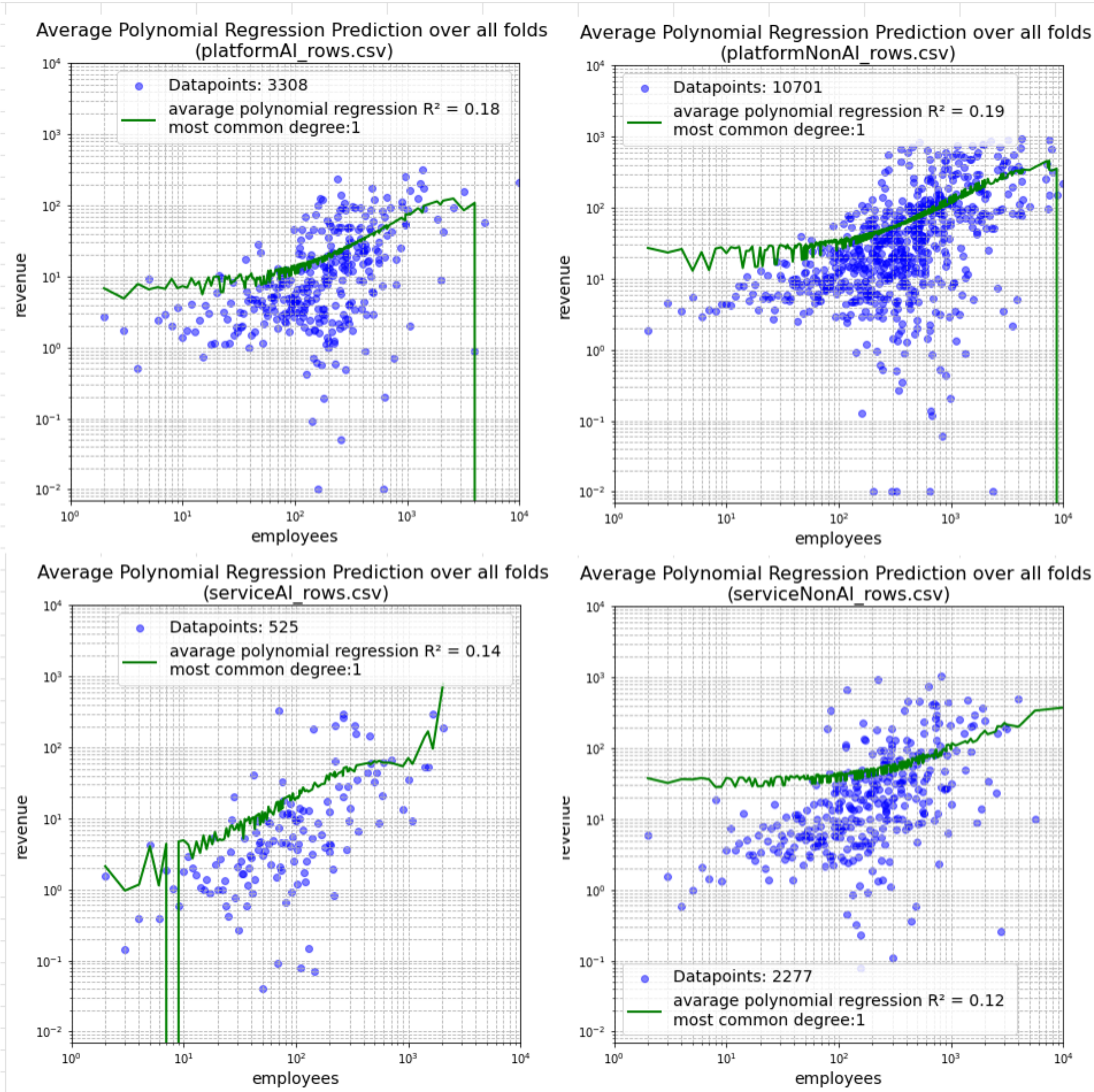


Figure 8.36: Revenue based on employees, polynomial regression, average estimates with cross-validation, own figure

8.9.4 Linear Interpolation

One of the mathematical models is Linear Interpolation. It has the following form:

$$y = y_0 + \frac{(x-x_0)(y_1-y_0)}{x_1-x_0}$$

Figure 8.37 shows plots of the model, where in the case of double occupancy, the average was calculated from the multiple values. The blue dots are the not manipulated data, and the R^2 value was calculated based on them. Afterward, the interpolated values undergo the robustness procedure. For the visualization, I plotted the interpolation numbers, over all folds, the average value after all folds are calculated, and the R^2 is also the average of all folds. So basically, you see an average function with its average R^2.

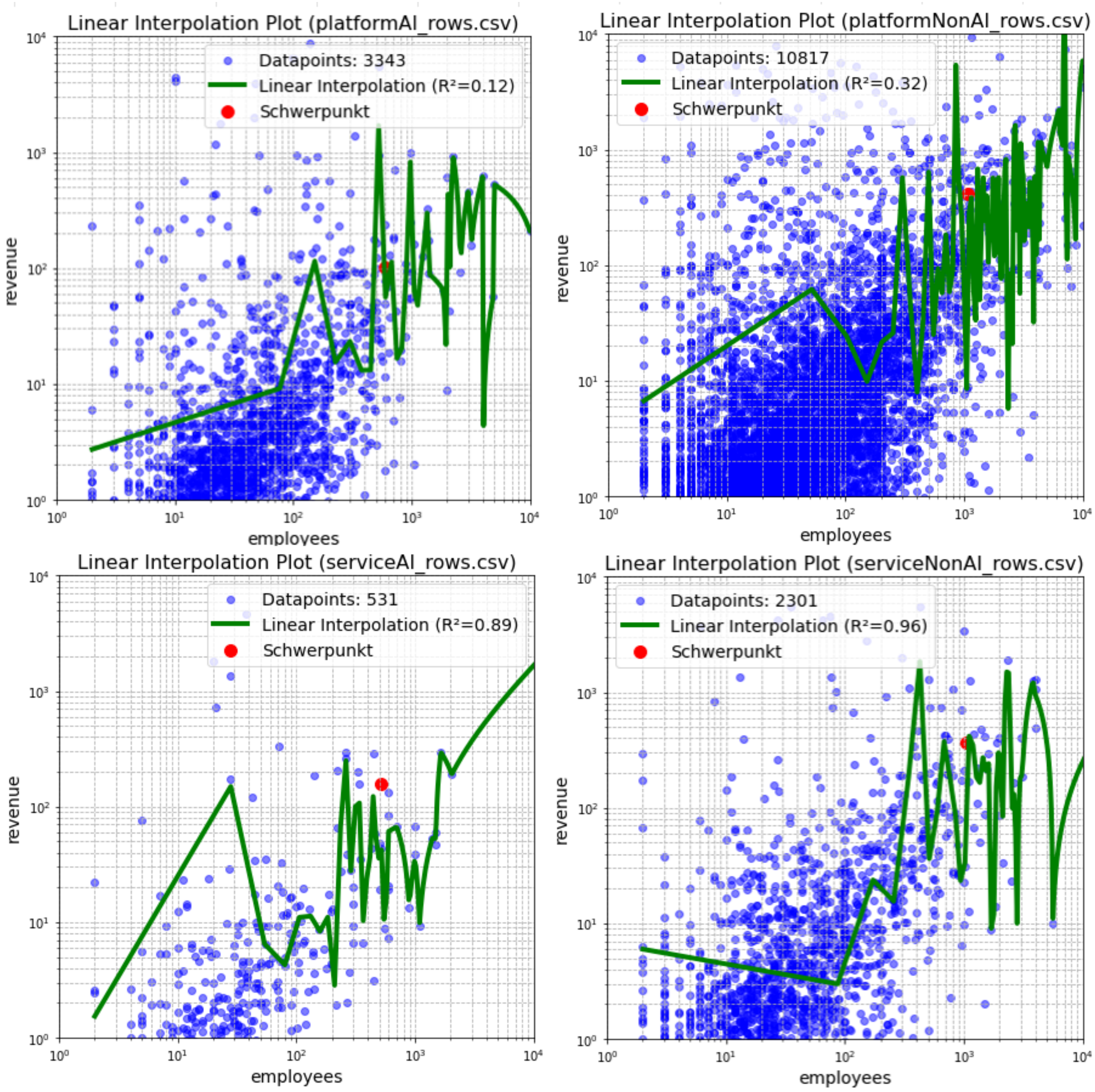


Figure 8.37: Revenue based on employees, linear interpolation, own figure

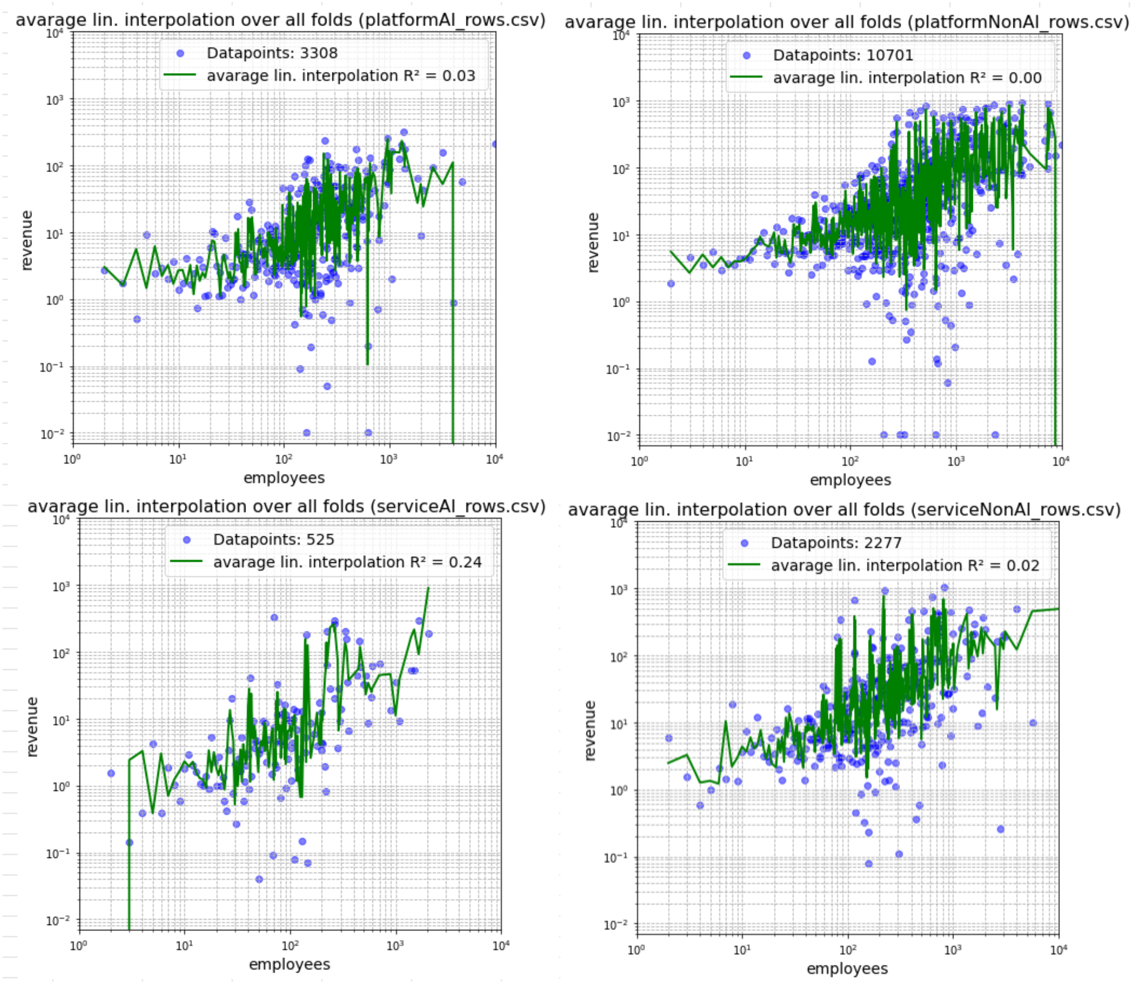


Figure 8.38: Revenue based on employees, robust linear interpolation, own figure

The functions shown in both Figure 8.37 and Figure 8.38 really look like overfitting the data.

8.9.5 Neuronal Network: MLP Regressor

Figure 8.39 shows a curve laid through all data points, and the R^2 was calculated for all data points to the curve.

Figure 8.40 shows R^2 calculated on a 20% random picked test set.

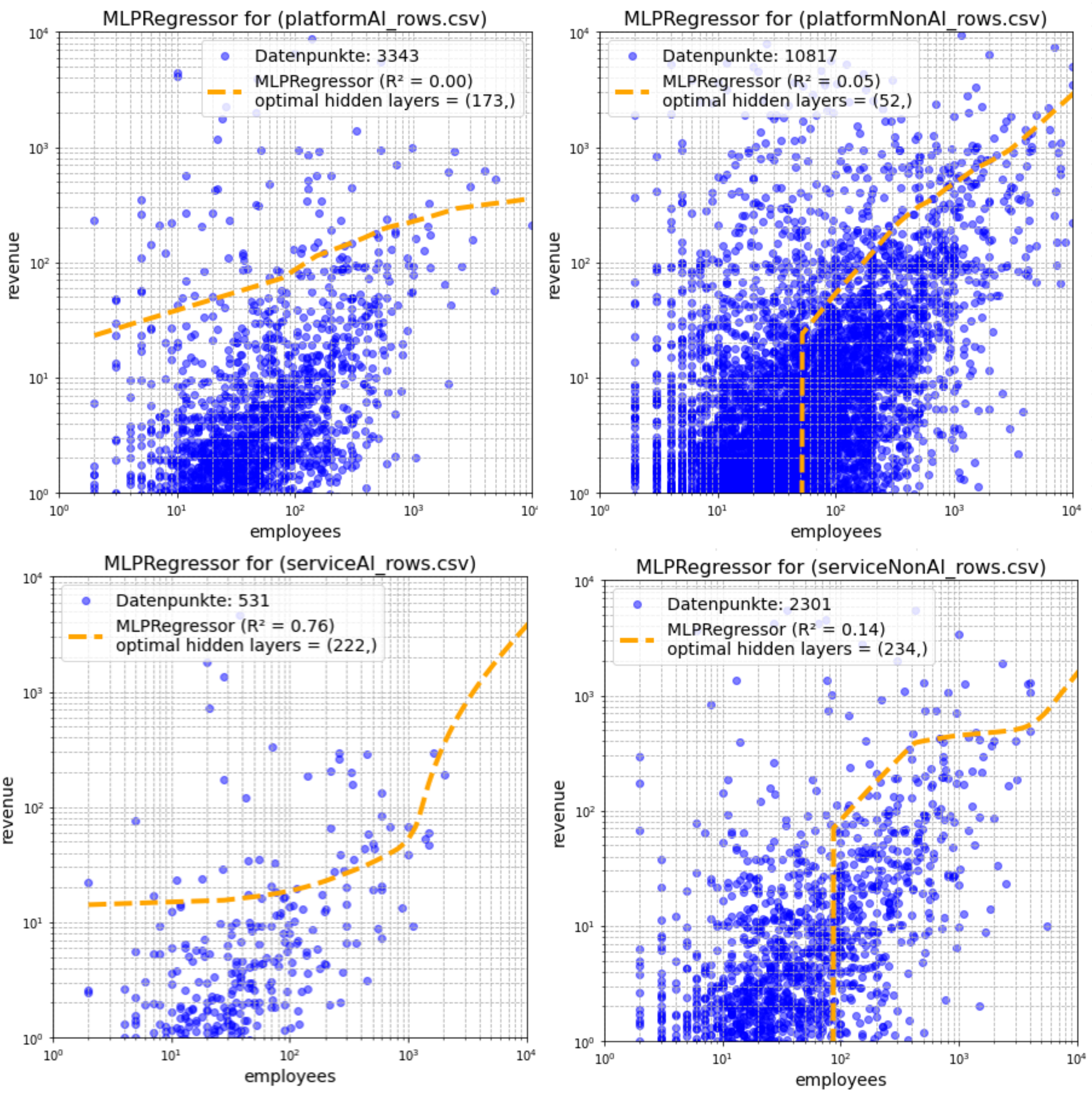


Figure 8.39: Non-robust MLP regression, R^2 calculated an all datapoints, own figure

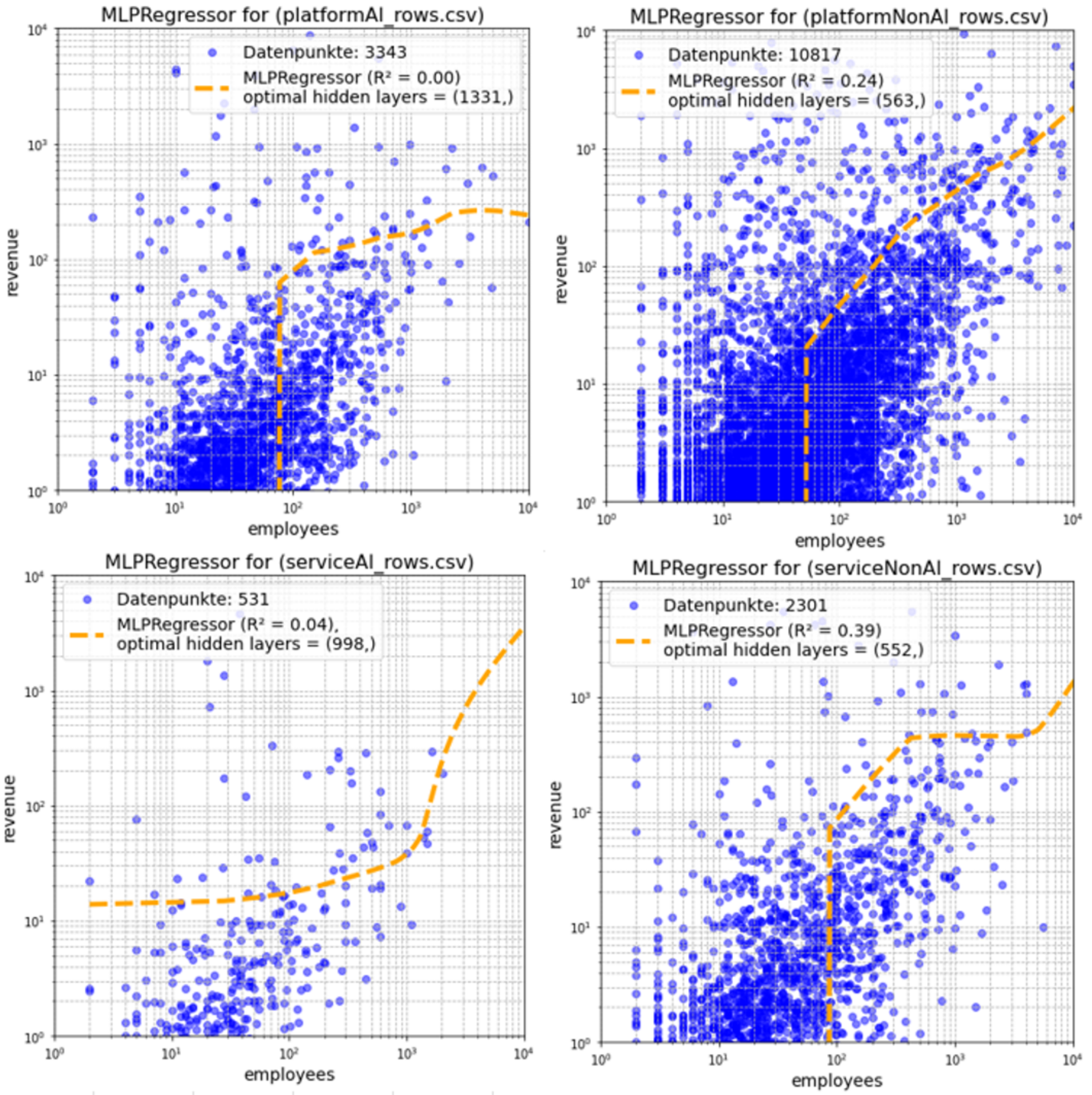


Figure 8.40: Robust (20% test data points) revenue based on employees, MLP Regressor, own figure

The last plot shows the MLP Regressor after the 5-fold cross-validation process.

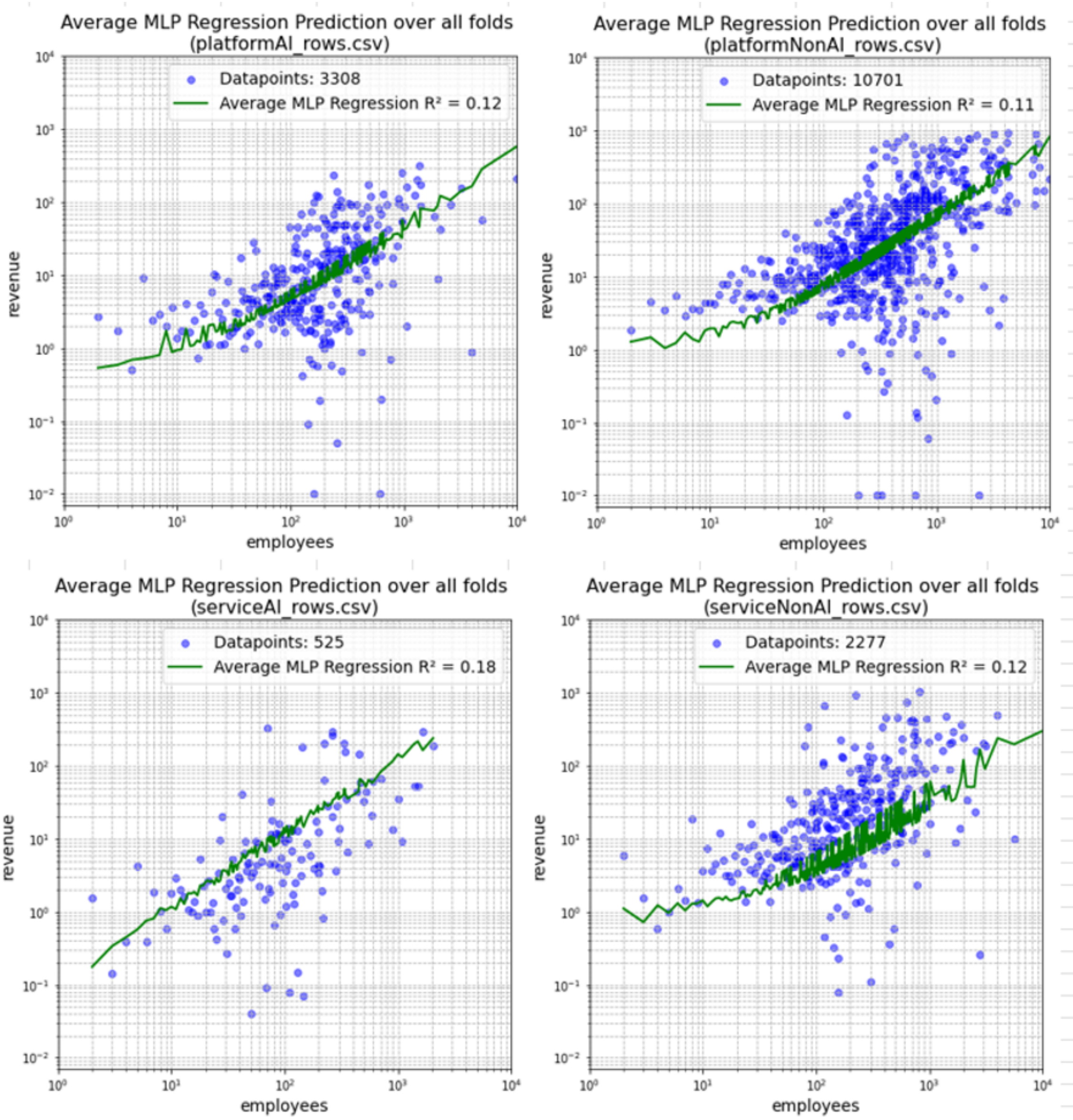


Figure 8.41: 5 fold cross validation: Revenue based on employees, power scaling robust with average values over all folds, own figure

You can see in the labels of the scatter plots that for each startup business model, there are different optimal hidden layer sizes. That means it is not optimal to train one model for all four startup types. Instead, it is better to do 4 different models.

8.9.6 Sub-analysis 2: Excluding top 1% Revenue Achievers, Power Function

Here is the overview of the 4 startup types, but without the top 1% revenue achievers. The presented function is the robust power function:

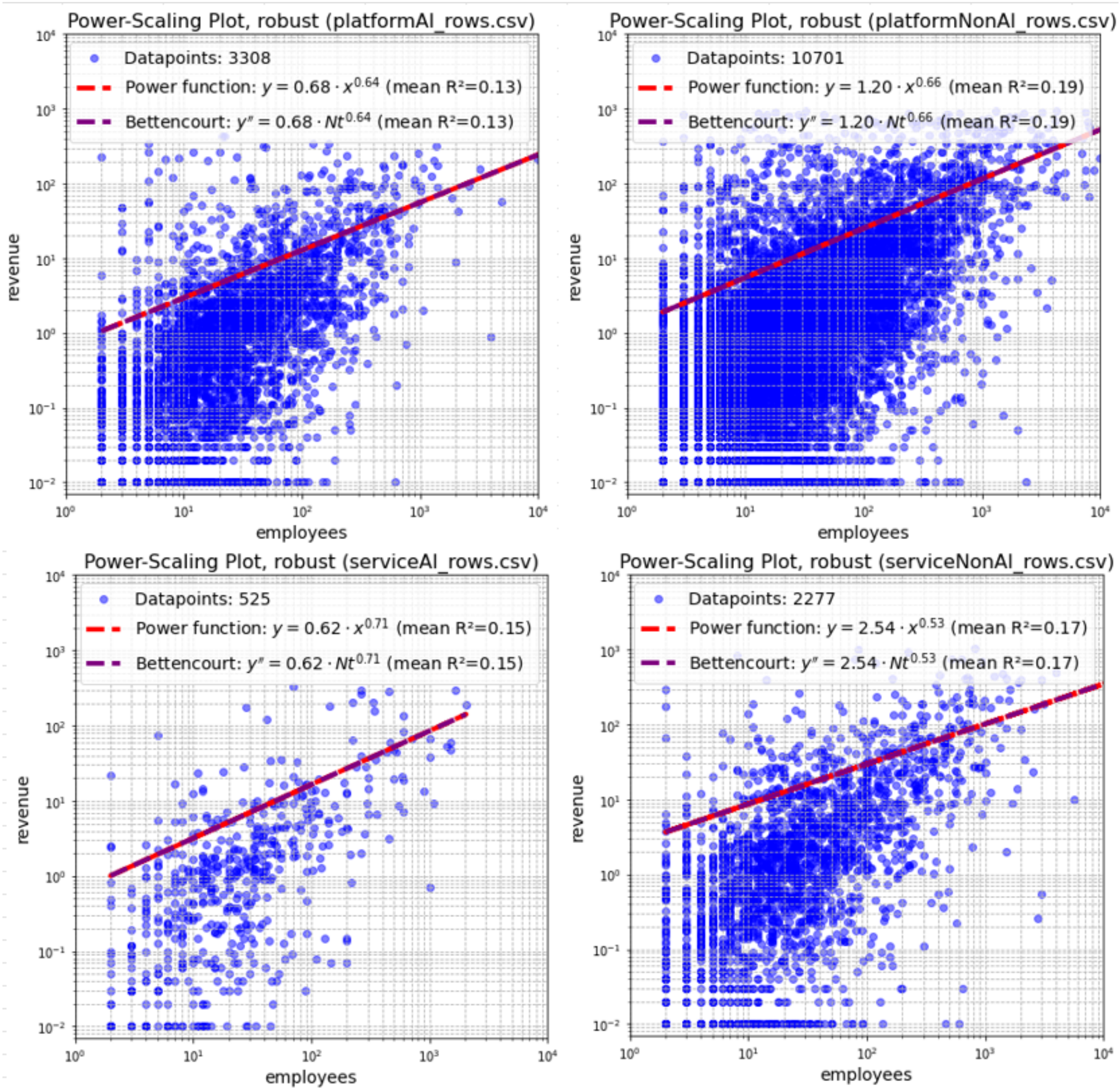


Figure 8.42: Robust model power-law scaling analysis, without top 1 Percent, own figure

You can see that the robust R^2 is quite similar for each category of startups, and the slope is also similar. You can clearly say that this 1% of companies, based on revenue, are having a big impact on the scaling analysis.

8.10 Overview Robust Models with Details

Table 8.27: R^2 values of different models across various categories, with MLP Regressor hidden-layer set to limit = 1000

Models (Robust)/R^2	Platform AI	Platform Non-AI	Service AI	Service Non-AI
Power / Bettencourt Function	0.00	0.05	0.12	0.31
P. Func. (without top 1%)	0.13	0.19	0.15	0.17
Support Vector Regressor	0.00	0.04	0.13	0.25
SVR (without top 1%)	0.29	0.15	0.10	0.18
Polynomial Regression	0.00	0.10 (degree = 1)	0.29 (degree = 3)	0.30 (degree = 2)
Pol. Reg (without top 1%)	0.18 (often degree = 1)	0.19 (often degree = 1)	0.14 (often degree = 1)	0.12 (often degree = 1)
Linear Interpolation	0.00	0.04	0.15	0.01
Lin. Interp. (without top 1%)	0.03	0.00	0.24	0.02
MLP Regressor	0.00 (layers = 692)	0.06 (layers = 833)	0.13 (layers = 320)	0.33 (layers = 890)
MLP Reg. (without top 1%)	0.12 (layers = 692)	0.11 (layers = 833)	0.18 (layers = 320)	0.12 (layers = 890)

8.10.1 Robust Models Gross Profit and Net Income for Scaling Factor

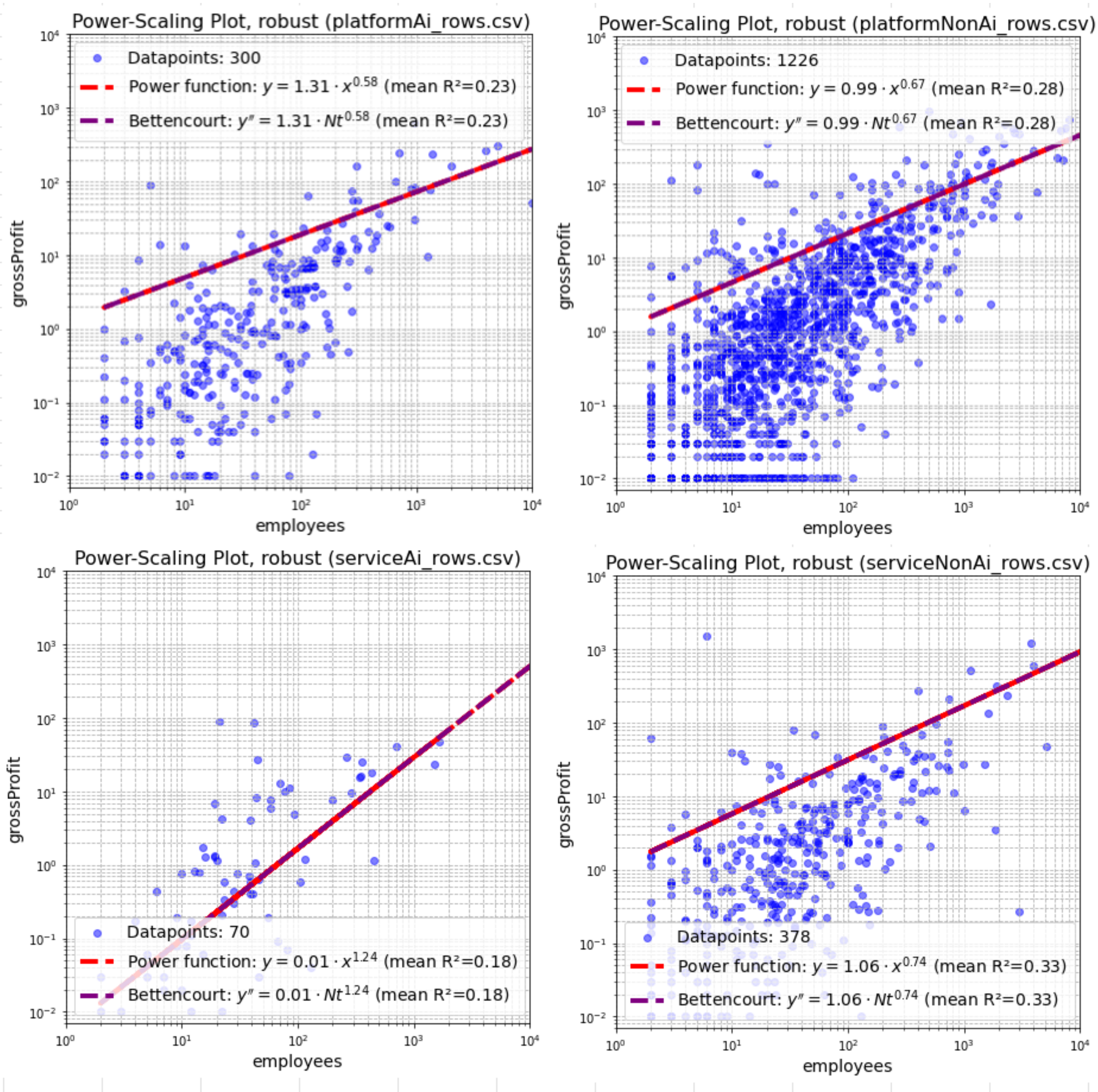


Figure 8.43: Robust model power law scaling analysis gross profit, own figure

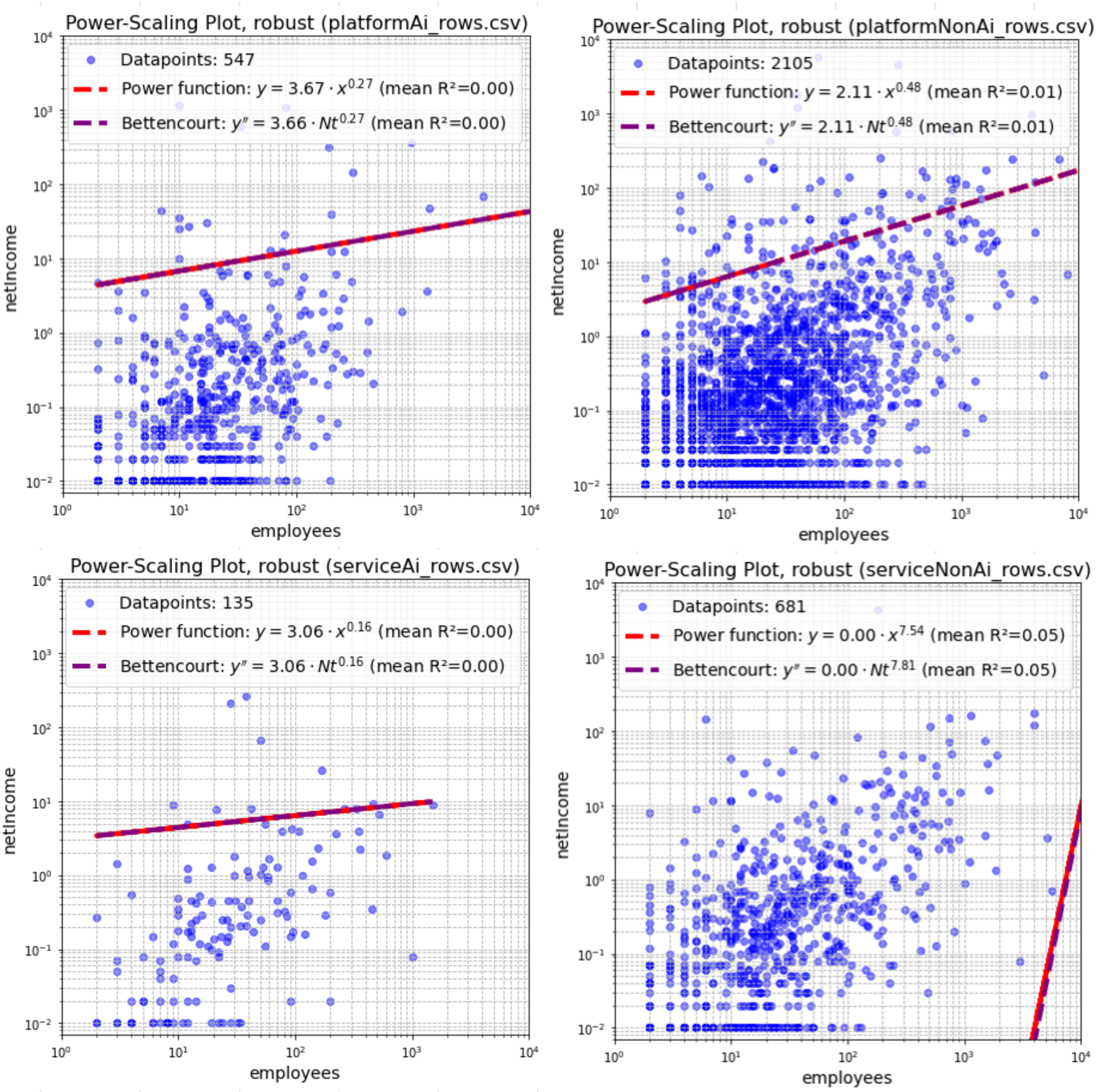


Figure 8.44: Robust model power law scaling analysis net income, own figure

8.11 Startup Classification Check Tables

8.11.1 Platform AI

Table 8.28: Platform AI summary of 100 validated startups

General information:		
Data points overall:	3343	
Number of checked startups, who have revenue and employees	100	
Number Yes	84	
Number No	14	
Number Unclear	2	

Table 8.29: List with 100 startups, random sample picked from the Platform AI CSV file and checked if classified correctly

Check	Reason	Legal name
Yes	AI-enabled HR platform, talent matching	10 BY 10 Inc.
Yes	AI-enabled platform, you can invest in AI startups, but can also get a pre-trained AI	100.co LLC
Yes	AI-enabled platform, Personnel matching platform	100Digital.AI LLC
Yes	AI-enabled platform for insurance to find and quit, matching insurance and consumer	10Life Hong Kong Limited
Yes	AI-enabled marketing strategy platform	10th Man Media Ltd
Yes	AI-enabled HR platform, talent matching	10x Psychology Ltd.
Yes	AI-enabled Platform, for sensors+services matching the project	12 Bridges, Inc.
Yes	AI 3D model platform for influencer	1337AI, Inc.
Yes	AI-enabled platform, matchmaking sales optimization Applications, and AI services are also there	19th Mile Capability Solutions Pvt. Ltd.
Yes	pot. AI-enabled platform, matchmaking building management suppliers and owner of buildings	1AIm GmbH
Yes	AI-enabled platform for climate compliance	1Climate, Inc.
Yes	pot. AI-enabled platform, matchmaking in death sector and foto services	1Director, Inc.
Yes	AI-enabled HR platform, talent matching	1Dossier Holdings Limited
Yes	AI-enabled platform, finance product matching	1Konto Inc.
Yes	AI-enabled platform, client-lawyer matching	1LAW, PC.
Yes	AI-enabled platform, matching jobs and leverage student skills	1Mentor Inc.

Yes	AI platform, matchmaking car rents and costumer	1Rent SL
Yes	AI Platform for IT solutions, automated matching of software solutions and AI influencers	1sec Co., Ltd.
Yes	AI-enabled HR platform, talent matching	1z Labs Co., Ltd.
Yes	platform for IT solutions, automated matching of software solutions	2 Twelve Solutions LLC
Yes	An enabled platform, matchmaking suppliers of risky software components and governments	202 Group LLC
Yes	AI-enabled platform matching buyers and sellers of real estate, many 3D features	202up GmbH
Yes	AI-enabled platform, compliance checks and resolve	2040 Sp. z o.o.
Yes	AI-enabled website builder, platform for templates like WordPress	21 Labs Inc
Yes	AI-enabled Platform matching buyers and sellers of obese woman clothing	21Squared, Inc.
Yes	AI-enabled platform, communication AI updated for social networking	Aurhum Networks Pvt Ltd
Yes	Platform for AI training sets, with matching index, could be AI-enabled	Beijing Zero One Everything Technology Co., Ltd.
Yes	AI-enabled Platform matching buyers and sellers	Close Comms Ltd.
Yes	AI-enabled platform, for microservice-architecture service matching	Devtenx Technologies Private Limited
Yes	AI-enabled platform and AI service, matchmaking for defense logistics, optimizing engineering parts	Safeflights, Inc.
Yes	AI-enabled website builder, platform for templates like WordPress	TenWeb, Inc.
Yes	Platform for deaf IT solutions, possibly automated matching of software solutions	Transcense, Inc.
Yes	AI-enabled Platform, social network	Two Two Two, Inc.

Yes	AI-enabled HR platform, talent matching	Yiwei Credit Service Co., Ltd.
No	AI-enabled service, for go-to-market Bot, connects and supervises section of companies	11X Limited
Yes	AI-enabled platform, matching translators to costumers	Berba Translations S.L.
Yes	pot. AI-enabled platform, offers 3D products for Metaverse to costumers	Mona Gallery, Inc.
No	platform like datev	144 Impact & Innovation Ventures LTD
No	AI-enabled service, data scientists label data	1715 Labs Ltd
No	AI aasS für cyber attaken	1Strike Sp z.o.o
No	Saas, klassich	24 Proof, LLC
No	heavy engineering product	Kangrim Co., Ltd.
No	AIaaS, datasets for AI, industry processes for AI	Beijing Jianyizhi Technology Co., Ltd.
Yes	AI-enabled platform, matchmaking (protection and hack-backs) costumers, AI-driven software sellers	Baffin Bay Networks AB
Yes	AI-enabled platform: matchmaking customer with visualization AIs	EXO Unlimited Inc.
Unclear	?	Melotech SociÃ©tÃ© Par Actions SimplifiÃ©e
Yes	AI-enabled platform: matchmaking health costumers, with App developer	Passio Inc.
Yes	AI-enabled platform: matchmaking investors and value creation offering companies	Plunk, Inc.
No	AIaas, ship tracking and arrival predictions, working on supply chain optimization -> not yet AI platform	Portcast Pte. Ltd.
Yes	AI-enabled platform, matchmaking insurance, products to health costumers	Artivatic Data Labs Private Limited
Yes	AI-enabled platform, matching energy producers and traders	SmartPulse Global Limited
Yes	AI-enabled platform, matching open source offerer and consumers	Debricked AB

Yes	AI-enabled platform, matching sensor-driven car insurance and car owners	Jooycar LLC
Yes	AI-enabled platform, matching car insurance, dealership, car sharing and car owners	Click-Ins, Ltd.
Yes	AI-enabled platform, matching data providers with data costumers for supply chain optimization, optimizes unstructured data with AI	Ioxio Oy
Yes	AI-enabled platform, matching supplier of object detection in images with costumers	PomVom Ltd.
Yes	AI-enabled platform, matching mining costumers with sensor anomaly detection and experts to fix it	Razor Labs Ltd
No	AIaas, offers market insights on pricing for producers	ShopGrok Pty Ltd
No	AIaas, offers smart home cyber protection	Zobi Limited
No	service, sells cooling systems	Smart Joules Pvt. Ltd.
Yes	AI-enabled platform, matching freelancers and website building-jobs	GIG Co., Ltd.
Yes	pot. AI-enabled platform, matches small ventures with software solutions, financing, or full integration with selling platforms	Osome Pte. Ltd.
No	AIaaS, elimination of data silos	Sigma Squares (Beijing) Technology Co., Ltd.
Yes	AI-enabled platform, enables customers to make their own AI, by AI ...	Worlds Enterprises, Inc.
No	AI service, data protection	Privacera, Inc.
Yes	AI-enabled platform: matchmaking HR resources	Future of Work Ltd.
No	AI aaS, selling website optimization	DemandJump, Inc.
Yes	AI-enabled platform, matches mortgages taker and giver	Homewise Solutions, Inc.

Yes	pot. AI-enabled platform, matching companies with growth-inducing financial or It products	Revio Analytics, Inc
Yes	AI-enabled platform, matching trucks and costumers	Trukky Logistics Services Pvt. Ltd.
Yes	AI-enabled platform, HR requirement	ShelfFlip, Inc.
Yes	AI-enabled platform, matching vessels and power grids	Rimot, Inc.
Yes	pot. AI-enabled platform, matching money donors with poor families	Goodnation Philanthropy Advisors, Inc.
Yes	pot. AI-enabled platform, matching truck fleets with costumers, like transportation or material producers, construction site	Tread Technologies Inc.
Yes	pot. AI-enabled platform, matching pre-trained AI sellers and buyers	Privitar Limited
Yes	AI-enabled platform, matching financial products with costumers	Olyv, Inc.
Yes	AI-enabled platform, matching hardware producer with consumers (with provided standard designs from startup)	Exceenis SAS
Yes	AI-enabled platform, transaction service	Cloud Asset Oy
Yes	AI aaS, cleanup company data	Integration Alpha GmbH
Yes	AI-enabled platform, matching private transport needs and transport enabler	Padam Mobility SAS
Yes	AI platform, social platform matching travelers	Blinkoo Srl
Yes	AI platform, matching influences with brands	Hypefactors AS
Yes	AI-enabled platform, matching clients with microbiome problems with Health Innovators	Phyla Technologies Inc.
Yes	AI enabled platform, matching truckers Accounting with funding supplier	Duke-AI, LLC.
Yes	AI enabled platform, social platform for students	Element451, Inc.

Yes	pot. AI enabeld platform, dokument based collaboration platform for companies	PT Feedloop Global Teknologi
Yes	AI enabled platform, matching brands with influencers	Perfpie Ltd.
Yes	AI enabled platform, matching beauty products with costumers	HAIrmod Bilgi Teknolojileri A S
Yes	AI enabeld platform, matching agrarbetriebe mit finazierung	Adea Grow Tecnologia E Intermediacao SA
Unclear	?	Dashible, Inc.
Yes	AI-enabled platform, a social learning platform for children, parents...	Hodoo Labs Co., Ltd.
Yes	AI-enabled platform, social platform for friends	TASDELEN PTE. LTD.
Yes	AI-enabled platform, matches financiers and borrower	Sales Boomerang, LLC
Yes	AI-enabled platform, matching employer and staff	StaffWRX, LLC
Yes	AI-enabled platform, matching employer and staff	Chefaa Inc.
Yes	AI-enabled platform, matching patients with local pharmacies and medicaments	LocoBuzz Solutions Pvt. Ltd.
Yes	AI-enabled platform, supply chain optimization and platform for data sharing in and between industries	Habu, Inc.
Yes	AI-enabled platform, matching trip-related interests, like compliance, eco-friendly..	WorkTrips Sp. Z O.o.
Yes	AI-enabled platform, matching fleets with fuel, and insurance	Tourmaline Labs, Inc.
Yes	AI-enabled platform, matching employees and health insurance	HealthMetrics Sdn Bhd
Yes	AI-enabled platform, education platform matching knowledge, students, and teacher, social platform	Edique Solutions Pvt Ltd

8.11.2 Platform Non-AI

Table 8.30: Platform Non-AI summary of 100 validated startups

General information:		
Data points overall:	10817	
Number of checked startups, who have revenue and employees	100	
Number Yes	70	
Number No	27	
Number Unclear	3	

Table 8.31: List with 100 startups, random sample picked from the Platform Non-AI CSV file and checked if classified correctly

Check	**Reason**	**Legal name**
Yes	platform matching shopping malls and shoppers	WE ARE APPA S.A.
Yes	platforms for games	Huya, Inc.
Yes	platforms for games	Cashgrail Private Limited
Yes	platforms home mortages	Pluto Labs Ltd
Yes	platforms for business payments, with hardware	myPOS World Ltd
Yes	platforms for the energy sector, transaction of energy and data between parties	Correla Ltd
Yes	platforms like taxi	zTrip Inc
Yes	platform for Mini jobs in russia	Ventra LLC
Yes	platform or online shop for electronics	Zebit, Inc.
No	AI-enabled platform, credit score, and finance product mathing	Clear Score Technology Ltd.
No	secure online platform, abstraction of infrastructure	Coder Technologies, Inc.
Yes	platform for food	JumbotAIl Technologies Pvt. Ltd.
Yes	platform for friends and dates, Noch kein Ki	Bumble, Inc.
Yes	platform for real estate investment	Quadro Partners, Inc.
Yes	platform for gaming	Galactus Funware Technology Private Limited
Yes	platforms for groceries	AyGoods Systems LLC

No	construction service	Beijing Urban Construction Intelligent Control Co., Ltd.
No	construction service	ShangHAI V-Test Semiconductor Co., Ltd.
No	service, digitalization of banks	M2P Solutions Private Limited
Yes	online uni with matchmaking students employer	Incanus Technologies Private Limited
Yes	platform for store founding	Vananam Ventures Pvt. Ltd.
Yes	platforms for equipment sharing	EquipmentShare.com Inc.
Yes	platforms for restaurant optimization and food delivery	OrderMark Inc
No	electronics company	Sourceability NA LLC
Yes	platform, online shop for shoes and small gadgets	StockX LLC.
No	website optimization service	ThriveCart LLC
Unclear		Capiter, LLC
Yes	platform, matchmaking investors and sports athletes, like players and horses	Commonwealth Markets, Inc.
Yes	platform for real estate investment	Square Yards Consulting Pvt. Ltd.
Yes	platform like bank, matching shares buyers and seller	Trade Republic Bank GmbH
Unclear		Peel Foods Limited
Yes	gaming platform	Blockchain Game Partners, Inc.
Yes	education platform	Go1 Pty Limited
No	IT service	Raintank Inc.
Yes	job platform	Stryder Corp.
No	buero management service	Nuvolo Technologies Corp
No	buero management service	Rippling People Center Inc.
No	restaurant digitalization service	Restaurante Muy S.A.S.
No	IT service	StoreConnect Pty Ltd
No	IT service	Sunrate Pte. Ltd.
Yes	LGBTQ marriage platform	NoRD DDB Cph
No	IT service	TXN Solutions, Inc.
Yes	platform for startup founders, matchmaking diverse services	Capbase Inc
No	IT service	Cato Networks Ltd.
No	IT service	Zulie Venture Inc
Yes	platform, matching bitcoin seller and buyer	CoinFlip Solutions, Inc.

Yes	platform, online shop for industry stuff	Ouyeel Co., Ltd.
Yes	platform, online shop grocery	Shopper Comercio Alimenticio Ltda
No	digitalization service super markets	SoluM Europe GmbH
No	blockchain service for luxury brands	Yaliyomo GmbH
Yes	platform, matchmaking gamers and professional coaches	The RoCo Group Inc.
No	IT service	AdFlare Inc.
Unclear		New BKH Corp.
Yes	platform with AI algorithm, but would function without as well	onlyfans
No	IT insurance service	Blue Zebra Insurance Pty Ltd
Yes	platform, online shop grocery	KiranaKart Technologies Pvt Ltd.
Yes	platform, crypto exchange	Quickbit.eu AB
Yes	platform, used car mobile marketplace	Instamotor Inc.
No	IT service	Fireblocks LLC
No	IT service	Que Processing Services Pvt. Ltd.
No	IT service	Axonius Inc.
No	IT service	Visible Alpha, LLC
Yes	platform for real estate investment	McMakler GmbH
Yes	platform for used smartphones	Swappie Oy
Yes	platform for stock market speculation	Wintermute Trading Limited
No	IT service, making games	Compliant Gaming LLC
Yes	platform for building materials	Capp Mobile BB SRL
Yes	platform, crypto exchange	BAM Trading Services Inc.
Yes	platform, HR talent matching	Teamway
Yes	platform, utility cost optimization	Homebox.io Ltd
Yes	platform, trash handling, matchmaking material, transporter, trash buyer	Schuttflix GmbH
Yes	platform, gas and power trade	Tate S.r.l.
Yes	platform, investment and entepreneurs	PT Goto Gojek Tokopedia Tbk
Yes	platform, online shop grocery	PT Moka Teknologi Indonesia
No	AI-enabled platform uses 'useless' phone data to calculate and convey loans	Branch International Financial Services Private Limited
Yes	Social media platform	Yappa World Inc.
No	Chinese nudist(?) holiday place	nackedHub
Yes	platform, apartment matching	Casalova, Inc.
Yes	platform for loans, insurances ...	Navi Technologies Ltd.

Yes	online shop for virtual bicycle	Zwift, Inc.
Yes	platform for discount vouchers	Double NC Co., Ltd.
Yes	platform, No code it solutions trade	Unqork, Inc.
Yes	platform, cryptocurrency	Bithyve UK Limited
Yes	platform, intra-city logistics platform	GOGOX Holdings Ltd.
Yes	platform, online pharmacy	Metabolic Healthcare Ltd
Yes	platform, bike renting	BLS Bikeleasing-Service GmbH & Co. KG
Yes	platform, like PayPal	Revolut Ltd
Yes	platform, online stock broker	TradeZero, Inc.
Yes	platform for discount vouchers	Upside Services, Inc.
Yes	platform, car rentals	Kovi Tecnologia Ltda.
Yes	platform, shippers carriers matching	Next Trucking, Inc.
Yes	platform, Chinese taxi	Dida Inc.
No	AI-enabled platform car-to-cargo matching	Full Truck Alliance Co. Ltd.
Yes	platform, No code it solutions trade	Commerce7
Yes	platform, online shop for consumer goods	Dastgyr Technologies (Pvt.) Limited.
Yes	platform, online shop for travel related services	Shouyue Technology (Beijing) Co., Ltd.
Yes	platform, online shop for clothing	Godsent AB
Yes	platform, online shop, art boxes, and cooking stuff	Artesane SAS
Yes	platform, sell and book warehouse spaces	OneVAST Ltd
Yes	platform, matching charging stations and fleets	Beep Technologies Pte Ltd

8.11.3 Service AI

Table 8.32: Service AI summary of 100 validated startups

General information:		
Data points overall:	531	
Number of checked startups, who have revenue and employees	100	
Number Yes	94	
Number No	5	
Number Unclear	1	

Table 8.33: List with 100 startups, random sample picked from the Service AI CSV file and checked if classified correctly

Check	**Reason**	**Legal name**
Yes	AI service optimizing robots	Bright Machines, Inc.
Yes	AI service powered image recognition	Basemark Oy
Yes	AI service sentiment analysis	Instyle.AI S.R.O.
No	AI platform, matching Influencers and brands	Epidemic d.o.o.
Yes	AI service, digitalization service with closed data platform with AI	AscentCore Technology SRL
Yes	AI service, digitalization service with closed data platform with AI	Federal Soft Systems Inc.
Yes	AI service, digitalization service with closed data platform with AI financing sector	Quantexa Limited
Yes	AI service, digitalization service with closed data platform with AI marketing sector	Axiata Digital Advertising Sdn Bhd
Yes	AI service, autonomous logistics optimization with robots	Standard Robots (Shenzhen) Co., Ltd.
Yes	AI service, digitalization service with closed data platform with AI	TheMathCompany Pvt Ltd
Yes	AI service, software für Kuka Arme, auch komplette Roboter Verkauf	Unchained Robotics GmbH
No	Apple watch bands	Shenzhen Ouxin International Holdings Co., Ltd.
Yes	AI service sentiment analysis	Seedtag Advertising S.L.

Yes	AI service, digitalization service with closed data platform with AI	NowVertical Group Inc.
Yes	AI service, connects different APIs ecosystem to custom solutions, with AI prediction in ...	iCiDIGITAL, Inc.
Yes	Service, education on AI workforce, without AI	Corndel Ltd.
Yes	AI service, data analysis for e-commerce	Assiduus Global Inc.
No	AI platform, Takeaways delivery system, matching restaurants, and hungry people	OrderYOYO A/S
Yes	AI service, autonomous defense robots	Ghost Robotics Corporation
Yes	AI service, cyber security	Bespot I.K.E.
Yes	AI service, hardware-near software for Android manufacturers, 90% of Android users use this (likely includes intelligent solutions)	Wanka Online Inc.
Yes	AI service, sensors and robots	Wayz Intelligent Manufacturing Technology Co., Ltd.
Yes	AI service and AI platform, AI service: autonomous cars, AI platform for Robo taxis	Chenqi Technology Limited
Yes	AI service, fintech	Cleo AI Ltd
Yes	AI service, digitalization service with closed data platform with AI	Quantgroup Technology Limited
Yes	AI service, sensors and robots for space mission	Spacebit Technologies Ltd.
Yes	AI service, digitalization service with closed data platform with AI	Datasys Group, Inc.
Yes	AI service, digitalization service in healthcare with closed data platform with AI	Purple Labs Healthcare Co., Ltd.
Yes	AI service, a closed data platform for DDDM	BizCents, LLC
Yes	AI service, health insurance and logistics in emergency situations	SOS Information Technology Co., Ltd.
Yes	AI service, smart home cameras	ShanghAI iMiLabs Technology Co., Ltd.

Yes	AI service, AI lawyer	Della AI Ltd
Yes	AI service, AI pilot for drones	Shield AI, Inc.
Yes	AI service, AI face recognition	Vintra, Inc.
Yes	AI service, autonomous robots for defense	Anduril Industries, Inc.
Yes	AI service, optimized healthcare storage logistics	Mingdu Zhiyun (Zhejiang) Technology Co., Ltd.
Yes	AI service, autonomous robots for cleaning floors	Narwal Intelligent Technology (Dongguan) Co., Ltd.
Yes	AI service, farming predictions	Vinsight, Inc.
Yes	AI service, sensor-driven waste sorting	AMP Robotics Corp.
Yes	AI service, cyber security	Moadata Co., Ltd.
No	Products like Lego robot for girls, No AI	SmartGurlz A/S
Yes	AI service, cyber security	Zyston LLC
Yes	AI service, autonomous cars, sensors	iMotion Automotive Technology (Suzhou) Co., Ltd.
Yes	AI service, sensor and robotics	Jiangsu New Vision Automotive Electronics Co., Ltd.
Yes	AI service, sensor and robotics	Morpheos S.r.l.
Yes	AI service, camera sensor for contAIners	AllRead Machine Learning Technologies S.L.
Yes	AI service, LED and IoT sensors for homes	Shenzhen Govee Technology Co., Ltd.
Yes	AI service, digitalization service with closed data platform with AI	GE Digital LLC
No	AI platform, matching all healthcare specialists and patients	Welltory Inc.
Yes	AI service, marketing tool, customer analysis	The Newco S.r.l.
Yes	AI service, digitalization service with closed data platform with AI	HARTB
Yes	AI service, similar to ChatGPT	Wing AI Technologies, Inc.
Yes	AI service, like PayPal but optimized with AI (for faster transactions)	Daniel Finance Ltd
Yes	AI service, object and people recognition for stores	Advertima Vision AG
Yes	AI service, personal finance manager	Coinscrap Finance S.L.
Yes	AI service, cyber security	PT David System Group

Yes	AI service, sensor-driven fitness	MOTUSI Corporation
Yes	AI service, robot for smoothies	6D Bytes Inc.
Yes	AI service, AI robot for chess, bot for interaction	Bryght Labs, Inc
Yes	AI service, AI-based stock supervisor	Pound Investment Co., Ltd.
Yes	AI service, DDDM for construction	LivSYT
Yes	AI service, digitalization service with potential closed data platform with AI	B2DIGIT SAS
Yes	AI service, closed geo data platform with AI	Spottitt Ltd.
Yes	AI service, AI lawyer	Trusli Inc
Yes	AI service, AI-optimized video distribution	VidCrunch, LLC
Yes	AI service, ChatGPT for education	Edusim LLC
Yes	AI service, AI classification of objects with camera	Delvitech SA
Yes	AI service, AI classification of objects on roads	EyeVi Technologies OÜ
Yes	AI service, autonomous robot for snow cleaning	Left Hand Robotics, Inc.
Yes	AI service, call center routine optimization and malpractice detection	Sistemas De Inteligencia De Negocios S.L.
Yes	AI service, AI bitcoin investing	Sentimenti Sp. z o.o.
Yes	AI service, chatbot for recruiting	Zmash AB
Yes	AI service, food-serving robot	Bear Robotics, Inc.
Yes	AI service, sports game analysis	AISpotter Oy
Yes	AI service, digitalization service with closed data platform with AI	Cien, Inc.
Yes	AI service, digitalization service DDDM	Firedesktop Srl
Yes	AI service, digitalization service with closed data platform with AI	AIgorithmics Sp. z o.o.
Unclear	Chinese company, website seems gone	Autognity S.R.L.
Yes	AI service, text generation for LinkedIn posts	Datananas SAS
Yes	AI service, digitalization service DDDM	Annova Solutions Pvt. Ltd.
Yes	AI service, energy usage optimization for data centers and industry	QiO Technologies Ltd

Yes	AI service, digitalization service and potential DDDM	Reiwa S.r.l
Yes	AI service, pre-trained AI algorithms	SciGood LLC
Yes	AI service, ChatGPT-like	Contextere Corporation
Yes	AI service, bankruptcy risk	Enin AS
Yes	AI service, ChatGPT for skin problems	Art Lab Co., Ltd.
Yes	AI service, AI analysis for DNA and chemicals	Consonance Sp. z o.o.
Yes	AI service, AI-powered focus assist for filmmakers	Moonlighting Industries AB

8.11.4 Service Non-AI

Table 8.34: Service Non-AI summary of 100 validated startups

General information:		
Data points overall:	2301	
Number of checked startups, who have revenue and employees	100	
Number Yes	75	
Number No	13	
Number Unclear	12	

Table 8.35: List with 100 startups, random sample picked from the Service Non-AI CSV file and checked if classified correctly

Check	**Reason**	**Legal name**
Yes	chinas Rocket startup	Beijing ZeroG Technology Co.,Ltd
Yes	football game streaming news	Beijing Dela Technology Co., Ltd.
Yes	shipment Notification app	AllSome Planet Sdn Bhd
No	blood sugar sensor with app	Levels Health, Inc.
Yes	3d printing machine seller	Suzhou Boli New Material Technology Co., Ltd.
Yes	like datev, receive management	Genuiti Software, Inc.
Yes	SOS button for old people, moving sensor	Wearable Technologies Inc.
Yes	light bulb for disinfecting surface	Far UV Technologies, Inc.
Yes	internet with satellites	Leaf Space S.r.l.
Yes	smart tag for motorcycles and bikes	Monimoto, UAB

Yes	upgrade of the faucet in the kitchen	Noa Water Ltd.
No	AI-enabled platform for shipment	Freidesk UAB
Yes	3D camera filming service	4DReplay, Inc.
Yes	It consulting service	Buutti Oy
Yes	telematik service	Civtec Ltd New Zealand
Yes	keynote speaker service	nexxtworks
No	VR headset hardware	Orqa d.o.o.
No	platform for money transactions	Zirtue, Inc.
No	platform for automotive stuff	beep-for-service
Yes	sells printing maschines	Bobst-Jetpack SAS
Yes	solar systems for homes, like panels and batteries	Xi'an MAIn Function Intelligent Technology Co., Ltd.
Unclear	website off	MIXMOVE AS
Yes	food rating service	Mystery Apex Ltd.
Unclear	website off	EGEE Services SAS
Unclear	korean website translation problem	J-Car Co., Ltd.
Yes	marketing compAIn service	17 Sport SAS
Yes	Tshirt printing service	Merchadise, Inc.
Yes	steel producer	Nortech Byggautomasjon AS
Yes	foto printing service	Chatbooks, Inc.
Yes	sports equipment seller	Maxpro Fitness LLC
Yes	Polish Rocket startup	SYDERAL Polska Sp. z o. o.
No	onlineshop for sneaker and streetwear Dropout	S.R.L.
No	social platform	Friendz Enterprise Srl
Yes	survey service	Mantap AB
Yes	flight miles collection service	Chatflights International AB
Yes	consulting and IT service on Biochar in African farms	AIrsmat Limited
No	inline shop for electronics	Metricool Software, SL
Yes	business consulting services	Mowdo, Inc.
Yes	sells mailboxes for packages	Citibox Smart Services, S.L.
Yes	gold-based pawn shop	Flat White Capital Private Limited
Yes	internal accounting digitization service	Innovative Payment Technologies LLC
Yes	hearing device producer	Audeara Limited
No	AI-enabled service, insurance document handling	The Jones Agency, Inc.
Yes	watgr delivery service in Dubai	WATERWA Food Stuff Trading LLC

Yes	games consultant service	PlayStack Ltd
Yes	educational books selling to schools	Mallory
No	HR platform	Helpr, Inc.
Yes	certification service	irse.org
Yes	early stage investors in startups	Pioneer Media Holdings Inc
Yes	sells sensor-equipped waste bins	Reen AS
Yes	warehouse service	Webshippy Kft.
Yes	sale there own electronics	Erised Semiconductor (Shenzhen) Co., Ltd.
Yes	consulting on IT	W Executive S.R.L.
Yes	Ad Production Services	SeenThis AB
Yes	engineering watering service	Blue Control A/S
Yes	cleaning service	CleanNow Inc.
Yes	supplement producer	NutriS d.o.o.
Yes	lawyer service	Advisor AS
Yes	marketing service	ricciardi-group
Yes	education courses	SKILFEKTORY LLC
Yes	education courses	zHero S.R.L.
Yes	producer of telematic parts	Erangtek Co., Ltd.
Yes	consulting and software development company	Trenolab SRLS
Yes	builder of electric charging stations	Jiaxing Zhixing Internet of Things Technology Co., Ltd.
Yes	electronics producer	Avi-on Labs, LLC
Yes	sale of engines	Dawn Aerospace Limited
Yes	virtual restaurant brand	kitchenita
Yes	sells toys	Playper LLC
Yes	sells coffee machines, like nestle	Divergent IP LLC
No	AI-enabled, service, wildlife tracking with drones	Wildlife Drones Pty. Ltd
Yes	software seller	Bluenext S.R.L
Yes	business consulting for startups	Chatwin SRL
Yes	closed platform for facility management	Ecotrak, LLC
Yes	it consulting for banks	EedenBull AS.
Yes	test service for Front ends	emazing-retailing sarl
Yes	credit card startup	PT Honest Financial Technologies
Yes	plumber and cleaning service	Odesyo SAS
Yes	car transport service	Papa Mobility Co., Ltd.

Yes	english lessons	Ringteacher Spain SL
Yes	dancing school	One Million Co., Ltd.
Yes	credit card service	UniOrbit Technologies Pvt Ltd
Yes	new UI for search in internet	Beam SAS
Yes	sports betting news	Ribacka Media AB
No	AI service, pipe inspection on huge landscapes with drones ...	ATLAS Innovative Engineering, S.L.
Yes	sell agricultural technology for togo farmers	Wonderful Togo Inc.
Yes	consultant or education service	Framsikt AS
Yes	roboters for warehouse	ATTAbotics, Inc.
Yes	it consulting for banks	Credere Group, LLC
Yes	investor	Trufin PLC
Yes	wingsAIl on ships	Smart Green Shipping Alliance Limited

8.12 Transparent AI Usage

The flowing tasks and model were used to make this master thesis:

Table 8.36: Overview of tasks and models for this work

Where was it used	**Task**	**Model**	**Was the AI-generated result manipulated by me?**
Outline	First draft of the structure	Chat GPT 4.o and Chat GPT 3.5	Yes
Chapter 2	Summary of the most important scaling mechanisms	hesse.AI	Yes
Chapter 1, Chapter 2 and Chapter 3	Introduction and other first drafts of the first chapters	Chat GPT 3.5, hesse.AI	Yes
Everywhere	Improving reading flow	Chat GPT 4.o, Chat GPT 3.5, Grammarly, online Microsoft Copilots	Yes
Everywhere	Improving Latex format	Chat GPT 4.o, Chat GPT 3.5	Yes
Code	Improving python code and comments in the code	Chat GPT 4.o, Chat GPT 3.5, online Microsoft Copilots and Perplexity (llama-3.1-sonar-large-128k-online)	Yes
Everywhere	Improving English	Google translate, Chat GPT 3.5 Chat GPT 4.o	Yes

8.12.1 Reflection about Usage of AI Tools

As required by the university, I reflect on the use of AI tools. While these tools can provide inspiration, they can also produce logically incorrect results, especially in code writing. The AI often overlooks good coding principles and fails to ensure that every possible state of the program is defined. Nonetheless, having these tools is beneficial, as they increase my productivity by a factor of four. I appreciate receiving responses from these language models, even when my prompts contain spelling errors; they perform better than any search engine.

I am aware that using AI comes with its pitfalls, such as biases and hallucinations, but the positive aspects are significantly more prevalent. If given the choice, I would prefer to use a university-level language model trained on the Primo database, as it would provide a more scientific AI that better meets the needs of any thesis.

8.13 Considered Guideline

This master thesis has partly and voluntarily followed this guideline from the FU department:

`https://www.wiwiss.fu-berlin.de/fachbereich/bwl/management/razinskas/resources/leitfaden.pdf`

Bibliography

Aggarwal, C. (2018). Neural networks and deep learning. Springer International Publishing.

Ahmed, S. M. (2019). Artificial intelligence in saudi arabia: Leveraging entrepreneurship in the arab markets. *Proceedings of the IEEE International Conference on Artificial Intelligence and Computer Vision (AICAI)*, 3. Retrieved from https://doi.org/10.1109/aicai.2019.8701348 (Online, Available: Accessed: 21.5.2024)

Berente, N., Gu, B., Recker, J., & Santhanam, R. (2021). Managing artificial intelligence. *MIS Quarterly*, *45*(3), 1433-1450. Retrieved from https://www.researchgate.net/publication/352400557

Bettencourt, L., Lobo, J., Helbing, D., Kuehnert, C., & West, G. (2007). *Growth, innovation, scaling, and the pace of life in cities.* Retrieved 2024-06-26, from https://www.researchgate.net/publication/6390858_Growth_Innovation_Scaling_and_the_Pace_of_Life_in_Cities

Brynjolfsson, E., & Rock, C., D.and Syverson. (2010). Artificial intelligence and the modern productivity paradox: A clash of expectations and statistics. National Bureau of Economic Research, Number Working Paper Series, 24001.

Carland, J., Hoy, F., & Boultom, J., W.and Carland. (1984). Differentiating entrepreneurs from small business owners. In (pp. 9(2), 354-359). The Academy of Management Review.

Colombelli, A., D'Amico, E., & Paolucci, E. (2023). When computer science is not enough: universities knowledge specializations behind artificial intelligence startups in italy. *The Journal of technology transfer, Springer US*, *48*(5), 1599–1627. Retrieved from https://link.springer.com/article/10.1007/s10961-023-10029-7

Davidson, E., & Vaast, E. (2010). Digital entrepreneurship and its sociomaterial enactment. In (p. 1-10). 43rd Hawaii International Conference on System Sciences, Honolulu, HI, USA.

Díaz-Santamaría, C., & Bulchand-Gidumal, J. (2021). Econometric estimation of the factors that influence startup success. *MDPI Sustainability*. Retrieved 2024-09-17, from https://www.mdpi.com/2071-1050/13/4/2242

Ejsmont, K., Gladysz, B., Roczon, N., Bettoni, A., Mert Barut, Z., Haber, R., & Minisci, E. (2024). *Multisided business model for platform offering AI services.*

Springer Cham, SpringerLink. Retrieved from https://link.springer.com/chapter/10.1007/978-3-031-46452-2_7

Ermakova, T., Blume, J., Fabian, B., Fomenko, E., Berlin, M., & Hauswirth, M. (2021). Beyond the hype: Why do data-driven projects fail? *Hawaii International Conference on System Sciences*. doi: 10.24251/HICSS.2021.619

Giuliano, G., Kang, S., & Yuan, Q. (2019). Agglomeration economies and evolving urban form. *The Annals of Regional Science, Springer-Verlag GmbH Germany, part of Springer Nature 2019*. Retrieved 2024-09-25, from https://link.springer.com/article/10.1007/s00168-019-00957-4

Haan, K. (2022). Net income: Definition, formula example. *Seeking Alpha*. Retrieved 2024-11-16, from https://seekingalpha.com/article/4459953-net-income#net-income-formula

Henderson, V., Kuncoro, A., & Turner, M. (1995). Industrial development in cities. *Journal of Political Economy*, *103*(5), 1067–1090. Retrieved 2024-09-25, from https://www.journals.uchicago.edu/doi/pdf/10.1086/262013

Jia, P., & Stan, C. (2021). Artificial intelligence factory, data risk, and vcs mediation: The case of bytedance, an ai-powered startup. *Journal of Risk and Financial Management*, *14*(5), 203. Retrieved 2024-03-01, from https://www.mdpi.com/1911-8074/14/5/203 (Num Pages: 19, Place: Basel, Publisher: MDPI, Web of Science ID: WOS:000654109300001) doi: 10.3390/jrfm14050203

Jöhnk, J., Weißert, M., & Wyrtki, K. (2021). Ready or not, ai comes— an interview study of organizational ai readiness factors. *Business and Information Systems Engineering*, *63*(1), 5–20. Retrieved 2024-03-12, from https://doi.org/10.1007/s12599-020-00676-7 doi: 10.1007/s12599-020-00676-7

Lins, S., Pandl, K. D., Teigeler, H., Thiebes, S., Bayer, C., & Sunyaev, A. (2021). Artificial intelligence as a service. *Business and Information Systems Engineering*, *63*(4), 441–456. Retrieved 2024-03-14, from https://doi.org/10.1007/s12599-021-00708-w doi: 10.1007/s12599-021-00708-w

Mishra, S., & Tripathi, A. R. (2024). Ai business model: an integrative business approach. *Journal of Innovation and Entrepreneurship volume 10, Article number: 18 (2021)*(2). Retrieved from https://innovation-entrepreneurship.springeropen.com/articles/10.1186/s13731-021-00157-5

Mittapally, B. (2024). Artificial intelligence and predictive analytics for business growth. *IHRIM International Association for Human Resource Information Management*. Retrieved 2024-11-18, from https://www.ihrim.org/2024/02/artificial-intelligence-and-predictive-analytics-for-business-growth/

Ollig, R. P. M. (2022). Managing digital transformation of pre-digital organizations. *Business and Information Systems Engineering, SpringerLink*. Retrieved from https://link.springer.com/article/10.1007/s12599-021-00708-w

Paluch, S., & Wirtz, J. (2020). Artificial intelligence and robots in the service encounter. *SMR - Journal of Service Management Research*, *4*, 3 – 8. Retrieved from https://www.nomos-elibrary.de/10.15358/2511-8676-2020-1-3/artificial-intelligence-and-robots-in-the-service-encounter-jahrgang-4-2020-heft-1?page=1

Perrault, R., & Clark, J. (2023). *Artificial intelligence index report 2023*. Stanford Institute for Human-Centered Artificial Intelligence (HAI). Retrieved from https://aiindex.stanford.edu/wp-content/uploads/2023/04/HAI_AI-Index-Report_2023.pdf

PichBook. (2024a). Our global data: Financials. *PichBook*. Retrieved 2024-11-16, from https://pitchbook.com/platform-data/financials

PichBook. (2024b). Pichbook example. *PichBook*. Retrieved 2024-11-16, from https://pitchbook.com/profiles/company/242514-19#comparisons

Pichbook. (2024). What are industry verticals? *Pichbook*. Retrieved 2024-11-17, from https://pitchbook.com/what-are-industry-verticals

pitchbook. (2022). pitchbook: Annual data increases. *pitchbook*. Retrieved 2024-11-02, from https://files.pitchbook.com/website/files/pdf/DACH_Annual_Data_Increases_2022.pdf

pitchbook. (2024). pitchbook: Screenshot frontend. *pitchbook*. Retrieved 2024-11-08, from https://pitchbook.com/products/desktop

Schapiro, A., Keutner, I., Friedrich, T., & Sanwald, L. (2024). Projektbericht winf-projekt informationsmanagement wise 23/24, entwicklung skalierbarer ressourcenkonfiguration in ki-startups: Eine qualitative studie. Freie Universität Berlin, Fachbereich Wirtschaftsinformatik, Unpublished work.

Schilling, M. A. (2002). Technology success and failure in winner-take-all markets: The impact of learning orientation, timing, and network externalities. *Academy of Management Journal, 45(2), 387–398.*. Retrieved 2024-11-18, from https://www.researchgate.net/publication/324986371_Technology_Success_and_Failure_in_Winner-Take-All_Markets_The_Impact_of_Learning_Orientation_Timing_and_Network_Externalities

Schlegel, D., Schuler, K., & Westenberger, J. (2023). Failure factors of ai projects: results from expert interviews. *International Journal of Information Systems and Project Management*, *11*(3), 25-40. Retrieved from https://aisel.aisnet.org/ijispm/vol11/iss3/3

Schulte-Althoff, M., Fürstenau, D., & Lee, G. M. (2021). *A scaling perspective on ai startups.* Retrieved 2023-11-21, from https://scholarspace.manoa.hawaii.edu/server/api/core/bitstreams/5ef17e99-8097-49cd-864d-ae3612beaef5/content doi: 10.24251/

HICSS.2021.784

Statista. (2024). Artificial intelligence - worldwide. *Statista.* Retrieved 2024-11-02, from https://www.statista.com/outlook/tmo/artificial-intelligence/worldwide#key-market-indicators

van Eck, N. J., & Waltman, L. (2018). Manual for vosviewer version 1 .6.9. Centre for Science and Technology Studies, Leiden University, The Netherlands. Retrieved from https://www.vosviewer.com/documentation/Manual_VOSviewer_1.6.9.pdf

Vartak, M. (2022). How to scale ai in your organization. *Business management, Havard Business Review.* Retrieved 2024-11-16, from https://hbr.org/2022/03/how-to-scale-ai-in-your-organization

Weber, M., Beutter, M., Weking, J., Boehm, M., & Krcmar, H. (2021). Ai startup business models. *Business and Information Systems Engineering, The International Journal of WIRTSCHAFTSINFORMATIK, SpringerLink*, *64*(1). Retrieved 2024-03-14, from https://aisel.aisnet.org/bise/vol64/iss1/6

West, G. (2019). *Scale - the universal laws of growth, innovation, sustainability and the pace of life, in organisms, cities, economies, and companies.* PENGUIN PRESS An imprint of Penguin Random House LLC. Retrieved from https://www.penguinrandomhouse.com/books/314049/scale-by-geoffrey-west/

Zebhauser, J., Rothe, H., & Sundermeier, J. (2023). Scaling ai ventures: How to navigate tensions between automation and augmentation. *Conference: Hawaii International Conference on System Sciences 2023*. Retrieved from https://www.researchgate.net/publication/365823835 doi: 10.24251/HICSS.2023.688

Zhao, F., & Collier, A. (2016). *Digital entrepreneurship: Research and practice.* STORE - Staffordshire Online Repository. Retrieved 2024-03-14, from https://eprints.staffs.ac.uk/6274/

Öztürk, O., Kocaman, R., & Kanbach, D. (2024). *How to design bibliometric research: An overview and a framework proposal.* SpringerLink. Retrieved 02.10.2024, 11:35 Uhr, from https://link.springer.com/article/10.1007/s11846-024-00738-0

Øverby, H., & Audestad, J. (2021). Multisided platforms: Classification and analysis. *Systems, MDPI*. Retrieved 2024-11-16, from https://www.researchgate.net/publication/356681607_Multisided_Platforms_Classification_and_Analysis

Acknowledgement

I would like to express my heartfelt gratitude to everyone who has supported me throughout the journey of completing this thesis.

My deepest thanks go to my supervisors, Prof. Fürstenau and Dr. Matthias Schulte-Althoff, for their invaluable guidance, constructive feedback, and insightful discussions, all of which significantly contributed to the success of this work.

I also wish to thank my family, especially my husband, Mark Paranskij, and my parents, Olga and Reinhard, for their unwavering patience, understanding, and encouragement during this intense period.

Additionally, I am grateful to my friends and fellow students, who were always there to listen to my questions and motivate me with their kind words of support.

Last but not least, I would like to thank my child, Ada, whose curiosity and joy for life continuously remind me of the importance of embracing new perspectives and tackling challenges with confidence.

This work would not have been possible without the support of all these wonderful people, and I am deeply thankful for their help.

....

(Berlin, November 19, 2024, signature of the author)

Declaration of Authorship/Eigenständigkeitserklärung

Hiermit versichere ich,

- dass ich die von mir vorgelegte Arbeit, selbständig abgefasst habe,
- dass ich keine weiteren Hilfsmittel verwendet habe als diejenigen, die im Vorfeld explizit zugelassen und von mir angegeben wurden,
- dass ich die Stellen der Arbeit, die dem Wortlaut oder dem Sinn nach anderen Werken (dazu zählen auch Internetquellen und KI-basierte Tools) entnommen sind, unter Angabe der Quelle kenntlich gemacht habe und
- dass ich die vorliegende Arbeit noch nicht für andere Prüfungen eingereicht habe.

Mir ist bewusst,

- dass ich diese Prüfung nicht bestanden habe, wenn ich die mir bekannte Frist für die Einreichung meiner schriftlichen Arbeit versäume,
- dass ich im Falle eines Täuschungsversuchs diese Prüfung nicht bestanden habe,
- dass ich im Falle eines schwerwiegenden Täuschungsversuchs ggf. die Gesamtprüfung endgültig nicht bestanden habe und in diesem Studiengang nicht mehr weiter studieren darf und
- dass ich, sofern ich zur Erstellung dieser Arbeit KI-basierter Tools verwendet habe, die Verantwortung für eventuell durch die KI generierte fehlerhafte oder verzerrte (bias) Inhalte, fehlerhafte Referenzen, Verstöße gegen das Datenschutz- und Urheberrecht oder Plagiate trage.

....

(Berlin, 19.11.2024, Unterschrift der Verfasserin)