A Scaling Perspective on AI Startups

Schulte-Althoff, Mattias; Fürstenau, Daniel; Moo Lee, Gene

Link to publication in CBS Research Portal

A Scaling Perspective on AI Startups

Matthias Schulte-Althoff
Freie Universität Berlin
matthias.schulte-althoff@fu-berlin.de

Daniel Fürstenau[1,2]
[1]Copenhagen Business School
dfu.digi@cbs.dk
[2]Freie Universität Berlin

Gene Moo Lee
UBC Sauder School of Business
gene.lee@sauder.ubc.ca

Abstract

Digital startups' use of AI technologies has significantly increased in recent years, bringing to the fore specific barriers to deployment, use, and extraction of business value from AI. Utilizing a quantitative framework regarding the themes of startup growth and scaling, we examine the scaling behavior of AI, platform, and service startups. We find evidence of a sublinear scaling ratio of revenue to age-discounted employment count. The results suggest that revenue-employee growth pattern of AI startups is close to that of service startups, and less so to that of platform startups. Furthermore, we find a superlinear growth pattern of acquired funding in relation to the employment size that is largest for AI startups, possibly suggesting hype tendencies around AI startups. We discuss implications in the light of new economies of scale and scope of AI startups related to decision-making and prediction.

1. Introduction

Artificial Intelligence (AI) unicorns [1], such as Google DeepMind, SenseTime, and UIPath have unlocked growth at an unprecedented pace. Networks and AI are reshaping the operational foundations of firms, enabling digital scale, scope, and learning, and simultaneously erasing deep-seated limits that have constrained firm growth and impact for hundreds of years [1]. AI startups seem to be able to extend the known types of scaling up, e.g., by being able to transfer potent machine learning models to other business use cases [2] or by providing new kinds of services that outperform humans in terms of perception and cognition [3]. We conceptualize *growth* as the process of changing in relevant measures of firm size and *scaling* as the relation of concepts to each other within the

growth processes. Achieving substantial growth and the ability to scale accordingly is a crucial point for startups, especially in regards to obtaining funding, maintaining productivity and enriching the diffusion of new product and technological innovations [4].

Given their potential, it seems important to consider why AI startups do not always scale like traditional software and Software-as-a-Service (SaaS) companies. Current Information Systems (IS) studies stress the potential barriers for creating value with AI technologies, such as unclear business cases for AI implementations, lack of leadership support and limited technological capabilities [5, 6]. They point to both the need for talent and access to data as well as the ability to capture value from AI applications. Indeed, recent industry reports point towards similar scaling problems stressing the expenses for cloud infrastructure usage and the problems of initial AI model setup that stem from data quality issues and unclear specifications [7]. Current observations paint the picture of AI startups as having to put a great deal of time into optimizing their AI models and collecting relevant data within the first two years [8] – effectively rendering their growth rate more similar to that of a traditional service company than to a platform or software company that can rely more on existing ready-to-use frameworks. For instance, many AI startups in the field of digital health focus on service business models before developing a more scalable approach [9].

At the core of these considerations is the question of whether there exist substantial differences between the "AI" and more classical digital startups and if so, how they can also be measured empirically. Hence the research question of this paper is: *How do AI startups scale compared to non-AI startups?* This is important for both management and investors since the current hype around AI falls short in describing the growth process of AI startups; yet and moreover, risks with regards to the initial phase of an AI startup tend to be

---

1 Unicorn is a market term characterizing a newly founded firm that

had rapidly grown to a private valuation of a billion or more US dollars [19]

HICSS

more concealed. Those AI startups are contrasted with, succinctly put, non-AI startups, especially those with a platform or service model, which does not indicate involvement of AI as a core of their business model. Drawing on a rich body of literature regarding platform and service startups, we theorize on how the mentioned types of startups should be able to scale compared to AI startups. Current reports provide indications that AI startups may scale more like traditional service companies tend to despite the big scaling potential of AI startups due to data network effects [8].

Our study uses data sets collected from the large-scale content aggregator Crunchbase[2]. Employing a quantitative predictive framework for growth and scaling, we analyze how size measures scale in relation to selected business measures, i.e. revenues and acquired funds, using OLS regression on a dataset of 12,373 AI startups, 11,839 platform startups and 24,401 professional service startups to examine group differences. We find that all types of startups show sublinear growth in terms of size and revenue and superlinear growth in terms of size and acquired funding. Additionally, taking into account the startups' age, we find that higher revenue is about as closely connected with more employees for AI startups as it is for service startups.

Our methodology allows for the quantification of scaling characteristics for different startup ventures. The empirical results provide critical arguments over phenomena of new economies of scale and scope of AI startups. First, we provide quantifiable differences between AI and non-AI startups. We then link the specific need for use of human resources in the process of scaling up among the different startup types. On average, AI startups seem to need more human resources in order to scale than platform startups and about as many as service startups do.

2. Growth and Scaling Characteristics of AI Startups

Growth of startups has been conceptualized in prior research, and we will briefly examine key findings relevant for AI scaling. We describe both growth and scaling while focusing on the growth characteristics of AI-based startups in comparison to two other startup types, i.e. platform and service. Our aim is to present a quantitative framework that allows for measuring growth and scaling differences between the three aforementioned startup types.

2.1 Digital Startups and Growth

We conceptualize **growth** as the process of changing in relevant measures of size, this includes: sales revenue, employees, and operating profits. Growth is deeply embedded in the scientific interests of entrepreneurship, so much so that it is sometimes included in the distinction between startups and small enterprises [10]. As a result, startups have been conceptualized as young, growth-oriented firms that engage in innovative behavior [11].

From this vantage point, digital startups can be further described as firms that market, deliver, and support a digital product or service online [12]. Digital startups rely on aspects of digital media and IT to pursue market opportunities [13]. They often do this by using emerging digital technologies such as AI, machine learning, deep learning, natural language processing, big data analytics, virtual reality, IoT platforms, 3D printing, or cloud computing [14, 15]. The important function of digital technology for startups lies in at least three different roles [16]: Digital startups may use digital technologies as a context, an enabler, or an outcome of their business. As a context, digital technologies help a startup coordinate, communicate, lead, organize, plan, and control. As an enabler, they facilitate better tangible and intangible methods of decision-making. As an outcome, digital technologies are the product or service the company produces, such as in delivering software to customers or producing hardware. Overall, digital startups make situated use of these three possible roles for IT to scale their venture.

Regarding such digital technologies, it is still widely discussed which occasions create the affordances for faster growth [17]. Fast growing digital startups often show significant traction enrolling customers, a validated business model and higher total funding than slower growing startups [18]. However, what makes digital business models easier to scale is that the marginal cost of serving an additional user on many digital networks is, for all purposes, zero, apart from the small incremental cost of (cloud) computing capacity [1]. At the same time, where low entry costs and plentiful capital is often available to digital startups, the entry barriers are thus quite low, resulting in an very large number of entrants into this specific market of startups [19]. Thus, Kenney and Zysman [19] argue that the competition ignites an equity-consuming race to build a market.

An important aspect of these growth processes is that different business sectors show different setup times, adoption speeds, sales cycles, and market opportunities.

There has already been some research on growth mechanisms of specific types of non-AI digital startups, such as *platform* startups (e.g., [20–23]) and *professional service* startups (e.g, [24–26]). Huang et al. [20] summed up the mechanisms of rapid growth of an innovative platform startup as (i) being data-driven, (ii) being able to release and launch modifications instantly and (iii) being able to swiftly transform novel value-in-use—and therefore able to reduce marginal costs for every new customer to close to zero.

In contrast, professional service startups are characterized by high knowledge intensity, low capital intensity, and professionalized workforces [27]. Such workforces, as well as the increasing capital intensity in the context of ongoing digitization, often still prevent them from rapid growth in the early years of their startup [26]. At the same time, the need for professional service providers is growing given an increasing knowledge demand in a technological society, putting ever more pressure on the need for an expertise in how to manage this knowledge [26].

2.2 Growth of AI Startups

The focus of this paper is on AI startups. We define an AI startup as *a digital startup having AI as a core component of its business model*. This means that while we consider startups where AI is an enabler or outcome of the business model, we do not consider digital startups as AI startups where AI technology merely acts as the context to improve its work processes [16]. Examples of AI startups include companies in categories such as machine learning [28–30], intelligent systems [31–33], natural language processing [34–38], and predictive analytics [39–42].

When considering the growth of AI startups, prior literature has pointed out two vital aspects. First, AI has been characterized as the next general purpose technology to be attributed with the property of enabling significant complementary investments, which include business process redesign, co-invention of new business models, and human capital [3]. To be even more specific, potent AI models allow for a high prediction quality and can therefore consistently identify and meet customer requirements [43]. In addition, it has been pointed out that once a mature AI model is set up, it has the potential for being transferred to other business use cases within the enterprise [1]. Second, Gregory et al. [8] stress the importance of data network effects for creating user value. This is a pertinent value to complement those well-documented direct and indirect network effects. Data network effects occur when the more that the AI platform learns from the data it collects on users, the more valuable the AI platform becomes to

each user. In turn, just as Andrew Ng [44] observed, talent and data seem to be the most scarce, yet crucial resources for a flourishing AI startup.

Both IS literature and industry reports indicate common issues for AI startups: As Berente et al. pointed out, there is significant uncertainty for businesses regarding how to manage AI [6]. It adds a level of complexity that surpasses traditional, less data-intense IT applications [45]. Alsheibani, Cheung, and Messom [5] highlight barriers for creating and capturing value using AI technologies, such as unclear business cases for AI implementations, lack of leadership support, and limited technological capabilities. These aspects not only point to the need for a solid backbone of skillful developers, an adequate toolset and a model that is trained on unique data but also to the fact that there seems to be a lack of clarity about how to extract business value from the use of unfolding AI technologies. Furthermore, current industry reports underline practical problems stemming from the mentioned issues [7]: First, expensive cloud infrastructure usage that requires ongoing human support is often required. Second, a great number of edge cases pose a problem for the initial model setup—it has been surmised that *AI lives in the long tail* [46]. Lastly, defensive moats are weaker due to the commoditization of AI models and algorithms. Considering these uncertainties, we present in the following subsection a way to quantitatively measure and predict the properties of the scaling behavior of AI startups.

2.3 Quantifying the Scaling Characteristics of AI Startups

In this subsection, we aim to unearth certain universal principles regarding growth and scaling of startups that could, in turn, provide a basis for a quantitative predictive framework. For starters, we define **scaling** as how size-related concepts relate to each other during the growth process, such as the relation of the number of employees and the operating profits of startups. Scaling, therefore, refers to how the individual components of a system respond when its size changes [47]. Scaling arguments can lead to a deep understanding of the dynamics of the system of our interest, especially if it is a continuously evolving complex adaptive system, such as the organizational structure of a startup.

According to Bettencourt et al. [48], there are three types of scaling dynamics. Sublinear scaling characterizes sigmoidal growth that eventually converges to the carrying capacity N. The driving forces put in more economic terms are *efficiency*, *savings in size* or *economies of scale* [47]. The slope β of the

corresponding log-transformed regression line describing the relationship of the scaling variables is smaller than one. Linear scaling characterizes growth that is exponential since the relative or percentage growth rate is constant. The slope of the corresponding log-transformed regression line equals one. Superlinear scaling characterizes growth that diverges within a finite time t. In economic discourse, superlinear growth often describes *increasing returns to scale*, as well as *self-reinforcing growth mechanisms* that lead to unbounded growth—or a boom. Of course, if resources get sparse, superlinear dynamics are necessarily followed by a collapse. The slope of the corresponding log-transformed regression is greater than one.

The scaling perspective has been put to use in the field of biology in order to characterize the growth properties of human and animal ecosystems [49–51] and in the field of urban growth [48, 52]. While scaling theory has not been widely used in the management and IS literature, there are a few notable works. Axtell [53] deployed an agent-based model in which the theory serves as an input for assumption formation. The simulation model analyzes endogenous firm dynamics and labor flows via heterogeneous agents. Another study that deployed a scaling perspective to analyze the dependence of growth on company size, and derived from all US traded manufacturing companies from 1975 to 1991 a model wherein the probability of a company's growth depends on its past and present sales accounts [54].

Furthermore, the scaling theory has been used by West [47] to measure the scaling behavior of publicly traded companies using the Compustat dataset. West discovered sublinear scaling of employees and net income. This sublinear scaling mechanism hints at a pattern of bureaucratic control, which is typically needed to administer the execution of the company's operational business model, within an ever growing organizational structure. When applying the scaling perspective to digital startups one is advised to bear in mind that entrepreneurial ecosystems are a much more turbulent and less predictable object of consideration than bigger companies [55]. Yet, losing some of its predictive quality, the theory can give insights into the different scaling behaviors of varying types of startups.

When comparing the growth characteristics of AI startups with the two non-AI startups types (*platform* and *service startups*), we find an indication that AI startups might scale faster than service startups, as their marginal costs for more sales units is linked with employing more consultants[3]. However, in congruence,

AI startups may scale slower than platform startups as the marginal costs for one sales unit are close to zero for the latter.

3. The Data

We use Crunchbase database, which is an open-source directory containing community-generated data on global technology startups and investors. We used the Crunchbase 'business group' categories to gather three classes of startups (AI, platform, and service startups). In addition, we would remove startups from one group if they also appeared in one of the other two groups in order to make the classes mutually exclusive. Together these comprised, respectively, 12,373, 11,839 and 24,401 individual startups with a maximum age of 10 years.

We considered three variables as relevant for our scaling analysis: the number of employees, the estimated revenue range, and the total amount of funding. The table in Appendix 1 shows a summary of the descriptive statistics for those variables, as considered in our model for each of the examined startup types. The differences in the number of observations stem from missing values in the Crunchbase dataset regarding the three mentioned variables.

The average number of employees is larger by a factor of roughly two for those service startups in comparison to AI and platform startups. Note that the startups of all three groups differ in their average age (4.2, 6.9, and 5.9 years for AI, platform, and service startups). We made use of an age-weight in an additional analysis, in order to consider this temporal dimension. Regarding the total amount of funding, the median seems to give a better picture in terms of not letting outliers distort the average amount: The median funding amount for AI startups is more than three times that of platform startups and 1⅔ times that of service startups.

Regarding reliability of the Crunchbase dataset, we found that this platform tightly monitors their data. In particular, Crunchbase takes three means to ensure data curation [56]: First, the editors are part of the business to control for face value validity of the data. Second, Crunchbase uses machine-learning algorithms to compare data against publicly available information. Finally, data analysts recruited by Crunchbase take manual care of data validation. Being able to give basic trust to the data sources, we will subsequently present the methodology used to analyze the data in this paper.

---

3 Or as Anne Marie Neatham, COO of Ocado Technology put it:

"Human beings can do everything that AI can do. They just can't do it to scale.", found in [1]

Table 1. Scaling exponents for business metrics vs. measurements of startup sizes

Startup size X	Business Metric Y	Group	β	Adj-R²	N
Number of employees	Estimated size of revenue	AI	0.32	0.12	1,989
		Platform	0.20	0.05	5,047
		Service	0.43	0.23	7,016
Age-discounted number of employees	Estimated size of revenue	AI	0.30	0.12	1,912
		Platform	0.24	0.07	3,838
		Service	0.31	0.13	5,115
Number of employees	Total sum of funding	AI	1.30	0.36	3,081
		Platform	1.06	0.24	1,204
		Service	1.06	0.28	2,389
Age-discounted number of employees	Total sum of funding	AI	1.31	0.35	2.942
		Platform	1.06	0.24	1,194
		Service	1.08	0.28	2,371

In the following, we apply a scaling analysis as sketched in subsection 2.3. Note that we are treating ordinal data as continuous to perform a log-transformed regression analysis. The argument to do so is following the assumption that the numerical distance between each set of subsequent categories is close enough to each other in our analysis. Consequently, the results will be rendered as close to reality as possible.

4. Results

The scaling analysis in this section is comprised of two main steps: comparing the scaling behavior of (i) the number of employees vs. the estimated revenue, and (ii) comparing the number of employees vs. the total sum of funding. For both steps, we first consider the size in total and, second, we examine the age-discounted firm size. Summary results for selected exponents are presented in Table 1, and scaling relationships are visualized in Figure 2.

The figures show the regression line and confidence margins for the log-transformed variables as well as the slope β of the log-transformed regression line. Using a measure N of startup size at time t, power law scaling takes the form $Y(t) = Y_0 N(t)^{\beta}$. Y denotes a business metric, which is either the estimated size of revenue or total sum of funding; Y_0 is a normalization constant. The exponent β reflects dynamic rules at play across the startups: It can be understood as an increase in the size N of the startup with a factor of one, which will lead to an average increase in the corresponding business metric by a factor of β. We will elaborate on the individual results in the following paragraphs.

The first analysis compares the startup's number of employees with its estimated revenue. We find sublinear scaling for all startup groups. The differences in the respective scaling exponents is still striking. We found the slope of our regression line for AI startups (0.31) to be higher by a factor of roughly 1.5 than that of platform startups (0.20). The slope of service startups in turn is

higher by 0.12 than for AI startups. Taking into account the higher baseline value for AI startups with small team sizes (below 10), the result suggests that marginal revenue is more tightly coupled with an increase in the number of employees for service startups than for AI startups.

We included the same analysis but with an age-discounted employee count to account for the temporal aspect of the startups' age. The age-discounted employee function $f(x,y) \triangleq x * e^{\alpha*y}$ with x being the number of employees, y being the startup age, and α being a weighting factor. The weighting factor α was set to 0.05 after testing the robustness of different models. The results show that among service startups, the slope drops noticeably to 0.31 while the slope of AI startups does not change nearly as much. This may suggest that age plays a more drastic role in increasing revenue for service startups than it would for AI startups.

The next analysis depicts the scaling of the total amount of funding compared to the number of employees. The results show superlinear scaling dynamics for all three startup types. As expected, AI startups have the steepest increase of funding with employment growth ($\beta = 1.30$), followed by platform and service startups with an equal slope ($\beta = 1.06$). In addition, the median of the total funding for AI is about three times higher than for platform start-ups and about 1.5 times higher than for service start-ups. The same analysis with an age-discounted employee count produces nearly the same results—this suggests the minor role of the temporal aspect concerning startup size in the scaling mechanism of the acquired funding. The results seem to stress the ability to acquire substantially higher funding for AI startups.

It is striking that the explained variance for the funding is significantly higher than for revenue. This is reasonable since funding can be put to use immediately in order to gather more resources whereas revenue may be bound to certain commitments [19]. However, as the variable revenue still explains a reasonable amount of
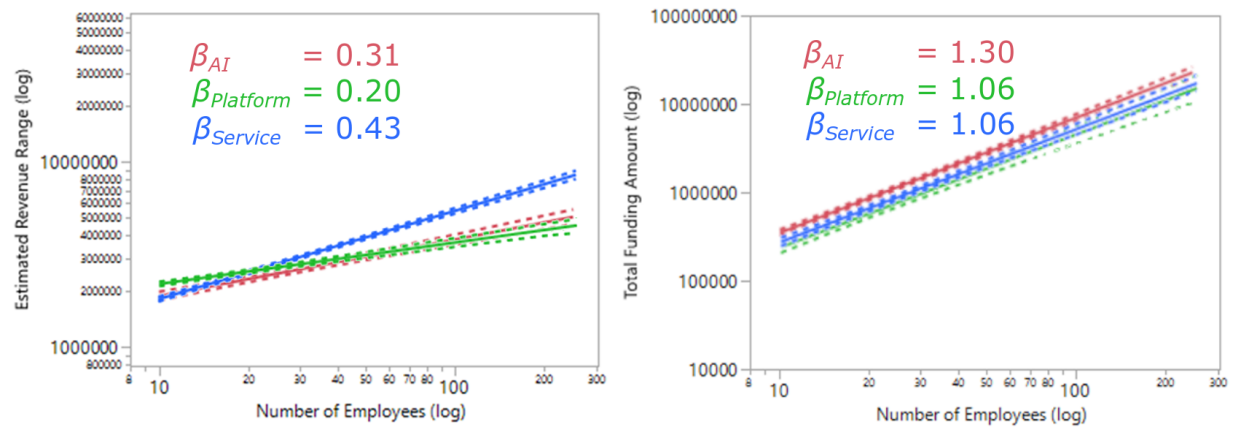
Figure 2. Scaling relationships for AI (red), platform (green), and service startups (blue): (a) the estimated amount of revenue vs. the number of employees, which scales sublinearly and (b) the total amount of funding vs. the number of employees, which scales superlinearly

the overall variance and because the differences in the effect strengths are so vastly different for the startup groups, we believe it is an insightful variable in terms of group comparison.

We included two further analyses by (i) splitting up the data geographically into *Europe, Northern America* and *Asian-Pacific* and (ii) examining different business categories of AI startups and included the results in Appendix 2. The regional analysis regarding revenue (i) reveals close to no differences for service startups but does demonstrate larger differences for platform and AI startups, namely: The beta coefficient is smallest for Northern America (AI: 0.23; platform: 0.18) and highest for Asian-Pacific (AI: 0.42; platform: 0.44) with Europe in between (AI: 0.29; platform: 0.25). Since the average value of the target variable is close to similar for the three categories, this may hint at a connection between employees and revenue that is much smaller in the US. In terms of funding, the regional differences are much smaller. The business category analysis for AI startups (ii) reveals that in the FinTech and the Health Care sector employees and revenue are connected most closely. Since these sectors are heavily regulated, market entry and data sharing prove to be especially difficult [9, 57]. Regarding funding amount, the more traditional sectors of Analytics and E-Commerce are less likely to gain more funding when having more employees.

We ran several robustness checks to rule out other explanations. First, we cutoff revenue outliers at different top percentages to control for possible distortions by the most performant startups. With a 10% cutoff, the slope for every startup group dropped similarly by around 3% (AI startups) to 5% (service startups). Second, grouping the startups by age and running the same analyses shows that both the

predictive quality and the slope *β* rise with a higher startup age, which is in line with the theory that bigger companies more consistently need more personnel in order to scale [47].
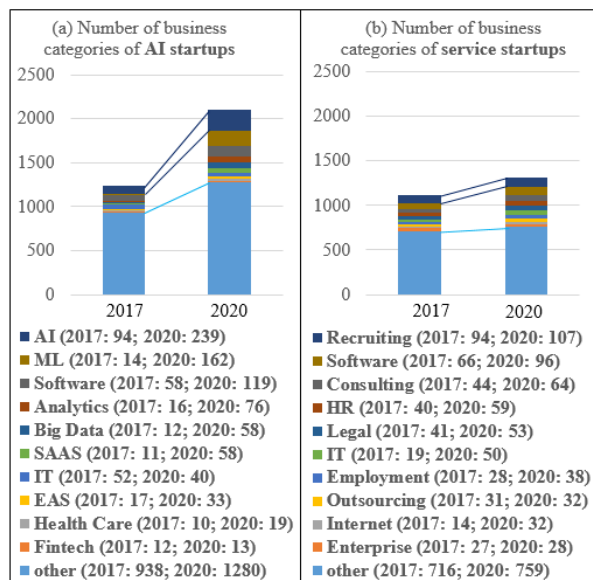
5. Discussion

The previous section revealed sublinear scaling dynamics for the estimated sum of revenue in relation to the startup size for all three startup types. This result is in line with that of West's [47] analysis regarding the scaling behavior of publicly traded companies we have referred to in subsection 2.3. Note that young startups will, of course, find it easier than larger companies to grow their revenues at higher percentage rates since it is the case that a small number is easier to double than a large one.

Regarding the group comparison between AI startups with platform and service startups, we find that marginal increase of revenue is linked less closely to the number of employees for AI startups than for the service startups, but more closely linked than that found for the platform startups. However, when incorporating the temporal aspect of a startup's age, the effect size for service startups is reduced to roughly the same amount as that of AI startups, while the effect size of platform startups is considerably lower. This hints to the argument that the necessity for humans-in-the-loop is much less present for platform startups than for the other two startup types. In particular, AI startups seem to need nearly as much personnel to scale as service startups appear to.

Furthermore, we find a superlinear growth pattern of acquired funding in relation to the startup size for all startups in our sample. The effect is biggest for the AI startup type hinting to the ever-rising popularity of AI

(a) Number of business categories of **AI startups**

(b) Number of business categories of **service startups**

- AI (2017: 94; 2020: 239)
- ML (2017: 14; 2020: 162)
- Software (2017: 58; 2020: 119)
- Analytics (2017: 16; 2020: 76)
- Big Data (2017: 12; 2020: 58)
- SAAS (2017: 11; 2020: 58)
- IT (2017: 52; 2020: 40)
- EAS (2017: 17; 2020: 33)
- Health Care (2017: 10; 2020: 19)
- Fintech (2017: 12; 2020: 13)
- other (2017: 938; 2020: 1280)

- Recruiting (2017: 94; 2020: 107)
- Software (2017: 66; 2020: 96)
- Consulting (2017: 44; 2020: 64)
- HR (2017: 40; 2020: 59)
- Legal (2017: 41; 2020: 53)
- IT (2017: 19; 2020: 50)
- Employment (2017: 28; 2020: 38)
- Outsourcing (2017: 31; 2020: 32)
- Internet (2017: 14; 2020: 32)
- Enterprise (2017: 27; 2020: 28)
- other (2017: 716; 2020: 759)

Figure 3. Increase of Crunchbase business categories from 2017 to 2019 for a sample of 300 (a) AI startups and (b) service startups matched with the AI startups according to their company size and revenue

an experience reflected in the number of investors in recent years [58].

Our findings pose interesting questions regarding new economies of scale and scope [1] in the growth process of AI startups. **Scale effects** occur when a firm's operating model is designed to cope with greater volume, complexity, and number of customers [1]. **Scope effects** are defined by the range of activities provided by a firm as measured by the number of products or services provided. A firm's operating model supports scope through having centralized functions or using advanced technologies [1]. IS literature has documented the important role of 'scaling the user base' for stimulating successful venture creation [20, 59], while being less concerned with scaling through other means such as data (see for exceptions [60, 61]). Thus, AI has been associated with new economies of scale [1, 43], but beyond user scaling recent accounts have emphasized the role of data at scale [43, 62]. While it is not entirely clear whether the business value of large data has unlimited positive returns [43, 63], access to enough relevant data is unquestionably a bottleneck factor for many AI startups to scale.

Some IS and innovation management accounts, which focus on platform firms have emphasized the role of scope effects in configuring diverse resources and knowledge into a coherent ecosystem [23, 64]. Certain researchers have argued that the increasing variety in different use cases makes AI (or specific forms, such as deep learning) a general purpose technology [65–67].

This means that AI is not limited to particular use cases or application domains, but transformative to many industries and domains [1]. This points to new scope effects where AI technologies remove bottlenecks, thus enabling startups to be less loyal to industry boundaries and transfer potent models to other sectors.

We find evidence for this in the form of a massive increase in business categories of AI startups compared to service startups. Figure 3 provides an overview of the increase of business categories from 2017 to 2019 for a sample of 300 (a) AI startups and (b) service startups that were matched to them according to their size and revenue. An alternative explanation for this phenomenon could be the hype surrounding such AI technology [68], which has resulted in many startups deliberately choosing to integrate either AI technologies or, at minimum, the AI technology labels.

Our results from the scaling analysis provide first empirical evidence that - on a large scale - the scaling behavior of AI startups does not differ substantially from service startups in terms of the need for human resources. Although both kinds of startups scale significantly differently than platform startups tend to, this is mainly due to having bigger teams not being as significantly connected with higher revenue. This points to AI startups having indeed a higher need for investing in human resource. The literature has consistently identified the two most common features of AI and digital platform startups regarding scaling dynamics: an operating model that allows for near to zero marginal costs for acquiring new customers [1, 69] and the much discussed network effects [8, 70]. Since there are measurable differences in the scaling behavior, an interesting research opportunity would be to further tease out those differences in the underlying mechanisms.

6. Conclusion

We aimed to better understand growth and scaling dynamics of AI startups. Using a large data set from Crunchbase, we analyzed the scaling behavior of AI startups and compared it to platform and professional service startups—making use of a quantitative, predictive framework. We found sublinear growth of revenue and superlinear growth of the total sum of funding both in a relation to the startup size as measured by the number of employees. Regarding the group comparison, we found that the marginal increase of revenue is as closely connected to employee count for AI startups as it is for service startups. Platform startups, on the other hand, seem to require less human resources for them to scale. Gregory et al [8] highlight the importance of data in the first place and respective data

network effects. In extension, we show that limited data availability and efforts for configuring the AI model may cause AI startups to need more people in order to grow during their first years. For management and investors, it is important to understand that data and AI model setup are crucial considerations for the first years of oversight and that these points potentially limit growth posing a risk to survival.

Our paper is not without limitations. First, the 'label' AI has been used to identify AI startups and to distinguish them from other types of startups. While data is well curated in Crunchbase further research should cross-validate whether the label and reality fit. Second, we focused on companies included in Crunchbase, which might have introduced a survival bias since being present in this data source may already be an indicator of a successful venture. Nevertheless, since our focus lies on *digital* entrepreneurs, we believe the large sample from Crunchbase can provide a comprehensive picture of the regarded startup types.

7. References

[1]   M. Iansiti and K. R. Lakhani, *Competing in the Age of AI: Strategy and Leadership When Algorithms and Networks Run the World*. Harvard Business Review Press, 2020.

[2]   C. C. Aggarwal, *Neural Networks and Deep Learning*. Springer International Publishing, 2018.

[3]   E. Brynjolfsson, D. Rock, and C. Syverson, "Artificial Intelligence and the Modern Productivity Paradox: A Clash of Expectations and Statistics," in *The Economics of Artificial Intelligence: An Agenda*, National Bureau of Economic Research, Inc, 2018, pp. 23–57.

[4]   J. Du and Y. Temouri, "High-growth firms and productivity: evidence from the United Kingdom," *Small Bus. Econ.*, vol. 44, no. 1, pp. 123–143, Jan. 2015.

[5]   S. Alsheibani, D. Y. Cheung, and D. C. Messom, "Factors Inhibiting the Adoption of Artificial Intelligence at organizational-level: A Preliminary Investigation," *AMCIS 2019 Proc.*, Jul. 2019.

[6]   N. Berente and N. Dame, "Call for Papers MISQ Special Issue on Special Issue Editors :," pp. 1–5, 2019.

[7]   M. Casado and M. Bornstein, "The New Business of AI (and How It's Different From Traditional Software)," 2020.

[8]   R. W. Gregory, O. Henfridsson, E. Kaganer, and H. Kyriakou, "The Role of Artificial Intelligence and Data Network Effects for Creating User Value," *Acad. Manag. Rev.*, Mar. 2020.

[9]   A.-K. Witte, D. Fuerstenau, and R. Zarnekow, "Digital Health Ecosystems for Sensor Technology Integration - A Qualitative Study on the Paradox of Data Openness," in *ICIS 2020 Proceedings*, 2020.

[10]   A. Desantola and R. Gulati, "Scaling: Organizing and growth in entrepreneurial ventures," *Acad. Manag. Ann.*, vol. 11, no. 2, pp. 640–668, Jun. 2017.

[11]   J. W. Carland, F. Hoy, W. R. Boulton, and J. A. C. Carland, "Differentiating Entrepreneurs from Small Business Owners: A Conceptualization," *Acad. Manag. Rev.*, vol. 9, no. 2, p. 354, Apr. 1984.

[12]   F. Zhao and A. Collier, "Digital Entrepreneurship: Research and Practice," in *9th Annual Conference of the EuroMed Academy of Business*, 2016.

[13]   E. Davidson and E. Vaast, "Digital Entrepreneurship and its Sociomaterial Enactment," in *43rd Hawaii International Conference on System Sciences*, 2010, pp. 1–10.

[14]   R. Beck, M. Avital, M. Rossi, and J. B. Thatcher, "Blockchain technology in business and information systems research," *Bus. Inf. Syst. Eng.*, vol. 59, no. 6, pp. 381–384, 2017.

[15]   P. Rippa and G. Secundo, "Digital academic entrepreneurship: The potential of digital technologies on academic entrepreneurship," *Technol. Forecast. Soc. Change*, vol. 146, pp. 900–911, Sep. 2019.

[16]   J. Recker and F. von Briel, "The Future of Digital Entrepreneurship Research: Existing and Emerging Opportunities," *ICIS 2019 Proc.*, Nov. 2019.

[17]   E. Autio, S. Nambisan, L. D. W. Thomas, and M. Wright, "Digital affordances, spatial affordances, and the genesis of entrepreneurial ecosystems," *Strateg. Entrep. J.*, vol. 12, no. 1, pp. 72–95, Mar. 2018.

[18]   E. Autio, "Entrepreneurship support in Europe : trends and challenges for EU Policy," no. February, p. 3, 2016.

[19]   M. Kenney and J. Zysman, "Unicorns, Cheshire cats, and the new dilemmas of entrepreneurial finance," *Ventur. Cap.*, vol. 21, no. 1, pp. 35–50, 2019.

[20]   J. Huang, O. Henfridsson, M. J. Liu, and S. Newell, "Growing on Steroids: Rapidly Scaling the User Base of Digital Ventures Through Digital Innovation," *MIS Q.*, vol. 41, no. 1, pp. 301–314, 2017.

[21]   J. Kim and J. Yoo, "Platform Growth Model: The Four Stages of Growth Model," *Sustainability*, vol. 11, no. 20, pp. 1–16, 2019.

[22]   M. Jocevski, A. Ghezzi, and N. Arvidsson, "Exploring the growth challenge of mobile payment platforms: A business model perspective," *Electron. Commer. Res. Appl.*, vol. 40, p. 100908, Mar. 2020.

[23]   M. G. Jacobides, C. Cennamo, and A. Gawer, "Towards a theory of ecosystems," *Strateg. Manag. J.*, vol. 39, no. 8, pp. 2255–2276, 2018.

[24]   R. Susskind and D. Susskind, *The Future of the Professions: How Technology Will Transform the Work of Human Experts*. Oxford University Press, 2017.

[25]   C. M. Christensen, D. S. Wang, and D. van Bever, "Consulting on the Cusp of Disruption," *Harv. Bus. Rev.*, 2013.

[26]   C. Kronblad, "How Digitalization Changes our Understanding of Professional Service Firms," *Acad. Manag. Discov.*, Apr. 2020.

[27]   A. Von Nordenflycht, "What is a professional service firm? Toward a theory and taxonomy of knowledge-intensive firms," *Acad. Manag. Rev.*, vol. 35, no. 1, pp. 155–174, Jan. 2010.

[28]   Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553. Nature Publishing Group, pp. 436–444, 27-May-2015.

[29]   G. Meyer *et al.*, "A machine learning approach to improving dynamic decision making," *Inf. Syst. Res.*, vol. 25, no. 2, pp. 239–263, Mar. 2014.

[30]   A. Roy *et al.*, "Performance comparison of machine learning platforms," *INFORMS J. Comput.*, vol. 31, no. 2, pp. 207–225, Jan. 2019.

[31]   S. Gregor and I. Benbasat, "Explanations from intelligent systems: Theoretical foundations and implications for practice," *MIS Q.*, vol. 23, no. 4, pp. 497–530, 1999.

[32]   M. Negnevitsky, *Artificial intelligence : a guide to intelligent systems*. Addison-Wesley, 2005.

[33]   J. Pearl, *Probabilistic reasoning in intelligent systems : networks of plausible inference*. 1988.

[34]   A. L. Berger, V. J. Della Pietra, and S. A. Della Pietra, "A Maximum Entropy Approach to Natural Language Processing," *Comput. Linguist.*, vol. 22, no. 1, pp. 39–71, 1996.

[35]   R. Collobert, J. Weston, J. Com, M. Karlen, K. Kavukcuoglu, and P. Kuksa, "Natural Language Processing (Almost) from Scratch," 2011.

[36]   C. D. Manning and H. Schütze, *Foundations of Statistical Natural Language Processing*. MIT Press, 1999.

[37]   Z. M. Shi, G. M. Lee, and A. B. Whinston, "Toward a Better Measure of Business Proximity: Topic Modeling for Industry Intelligence," *MIS Q.*, vol. 40, no. 4, pp. 1035–1056, 2016.

[38]   D. Shin, S. He, G. M. Lee, A. B. Whinston, C. Suleyman, and K.-C. Lee, "Enhancing Social Media Analysis with Visual Data Analytics: A Deep Learning Approach (Fourthcoming)," *MISQ*.

[39]   Y.-K. Lin, H. Chen, R. A. Brown, S.-H. Li, and H.-J. Yang, "Healthcare Predictive Analytics for Risk Profiling in CHronic Care: A Bayesian Multitask Approach," *MISQ*, vol. 41, no. 2, pp. 473–495, 2017.

[40]   R. B. Parikh, Z. Obermeyer, and A. S. Navathe, "Regulation of predictive analytics in medicine," *Science (80-.).*, vol. 363, no. 6429, pp. 810 LP – 812, Feb. 2019.

[41]   G. Shmueli and O. R. Koppius, "Predictive analytics in information systems research," *MIS Quarterly: Management Information Systems*, vol. 35, no. 3. pp. 553–572, Sep-2011.

[42]   G. M. Lee, S. He, J. Lee, and A. B. Whinston, "Matching Mobile Applications for Cross Promotion (Forthcoming)," *ISR*.

[43]   A. Agrawal, J. Gans, and A. Goldfarb, *Prediction Machines: The Simple Economics of Artificial Intelligence*. Harvard Business Review Press, 2018.

[44]   A. Ng, "What artificial intelligence can and can't do right now," *Harv. Bus. Rev.*, 2016.

[45]   A. Rai, P. Constantinides, and S. Sarker, "Editor's Comments: Next-Generation Digital Platforms: Toward Human–AI Hybrids," *Manag. Inf. Syst. Q.*, vol. 43, no. 1, Mar. 2019.

[46]   M. M. Al-Razouki, "The Long Tail of Artificial Intelligence," 2018.

[47]   G. West, *Scale: The Universal Laws of Life and Death in Organisms, Cities and Companies*. Orion Publishing Group, Limited, 2018.

[48]   L. M. A. Bettencourt, J. Lobo, D. Helbing, C. Kühnert, and G. B. West, "Growth, innovation, scaling, and the pace of life in cities," *Proc. Natl. Acad. Sci. U. S. A.*, vol.

104, no. 17, pp. 7301–7306, 2007.

[49]   G. West, J. Brown, and B. Enquist, "The Fourth Dimension of Life: Fractal Geometry and Allometric Scaling of Organisms," *Science*, vol. 284, pp. 1677–1679, Jul. 1999.

[50]   G. West, J. Brown, and B. Enquist, "A general model for the structure and allometry of plant vascular systems," *Nature*, vol. 400, pp. 664–667, Aug. 1999.

[51]   G. B. West, J. H. Brown, and B. J. Enquist, "A general model for ontogenetic growth," *Nature*, vol. 413, no. 6856, pp. 628–631, Oct. 2001.

[52]   D. Helbing *et al.*, "Power Laws in Urban Supply Networks, Social Systems, and Dense Pedestrian Crowds," in *Complexity Perspectives in Innovation and Social Change*, Springer Netherlands, 2009, pp. 433–450.

[53]   R. Axtell, "Endogenous Firm Dynamics and Labor Flows via Heterogeneous Agents," *Handb. Comput. Econ.*, vol. 4, pp. 157–213, Jan. 2018.

[54]   M. H. R. Stanley *et al.*, "Scaling behaviour in the growth of companies," *Nature*, vol. 379, no. 6568, pp. 804–806, Feb. 1996.

[55]   D. M. Steininger, "Linking information systems and entrepreneurship: A review and agenda for IT-associated and digital entrepreneurship research," *Inf. Syst. J.*, 2018.

[56]   R. Basole, H. Park, and R. O. Chao, "Visual Analysis of Venture Similarity in Entrepreneurial Ecosystems," *IEEE Trans. Eng. Manag.*, vol. PP, pp. 1–15, 2018.

[57]   K. N. A. Arthur and R. Owen, "A Micro-ethnographic Study of Big Data-Based Innovation in the Financial Services Sector: Governance, Ethics and Organisational Practices," *J. Bus. Ethics*, vol. 160, no. 2, pp. 363–375, 2019.

[58]   X. Mou, "Artificial Intelligence: Investment Trends and Selected Industry Uses," *Int. Financ. Corp.*, pp. 1–8, 2019.

[59]   O. Henfridsson and B. Bygstad, "The Generative Mechanisms of Digital Infrastructure Evolution," *MIS Q.*, vol. 37, no. 3, pp. 907–931, 2013.

[60]   S. Jarvenpaa and M. L. Markus, "Data Perspective in Digital Platforms: Three Tales of Genetic Platforms," *Proc. 51st Hawaii Int. Conf. Syst. Sci.*, pp. 4574–4583, 2018.

[61]   H. Rothe, S. Jarvenpaa, and A. Penninger, "How do Entrepreneurial Firms Appropriate Value in Bio Data Infrastructures: An Exploratory Qualitative Study," in *Proceedings of the 27th European Conference on Information Systems (ECIS)*, 2019.

[62]   R. M. Bond *et al.*, "A 61-million-person experiment in social influence and political mobilization," *Nature*, vol. 489, no. 7415, pp. 295–298, Sep. 2012.

[63]   D. Lazer, R. Kennedy, G. King, and A. Vespignani, "The Parable of Google Flu: Traps in Big Data Analysis," vol. 343, no. March, pp. 1203–1205, 2014.

[64]   A. Gawer, "Bridging Differing Perspectives on Technological Platforms: Toward an Integrative Framework," *Res. Policy*, vol. 43, no. 7, pp. 1239–1249, 2014.

[65]   T. G. Dietterich and E. J. Horvitz, "Rise of concerns about AI," *Commun. ACM*, vol. 58, no. 10, pp. 38–40, Sep. 2015.

[66]   M. R. Frank, D. Wang, M. Cebrian, and I. Rahwan, "The evolution of citation graphs in artificial intelligence,"

Nat. Mach. Intell., vol. 1, no. February, pp. 79–85, 2019.

[67]   J. Klinger, K. Stathoulopoulos, V. Embankment, and U. Kingdom, "Deep learning, deep change? Mapping the development of the Artificial Intelligence General Purpose Technology," *arXiv*, 2018.

[68]   S. Raisch and S. Krakowski, "Artificial Intelligence and Management: The Automation-Augmentation Paradox," *Acad. Manag. Rev.*, Feb. 2020.

[69]   P. Hartmann and J. Henkel, "The Rise of Corporate Science in AI: Data as a Strategic Resource," *Acad. Manag. Discov.*, Apr. 2020.

[70]   H. Varian, "Artificial Intelligence, Economics, and Industrial Organization," National Bureau of Economic Research, Inc, 2018.

Appendix

Appendix 1: Variable Type, Range and Distribution

Variable	Type and range	Group	Rank distribution (number of startups in parenthesis)	Descriptive statistics	
Number of employees	Grouped; range represented by average: **5**: (0,10] empl. **25**: (10, 50] empl. **75**: (50, 100] empl. **175**: (100, 250] empl. **250**: >250 empl.	AI	**5** (4,773); **25** (2,740); **75** (272); **175** (175); **250** (51); N = 7,986	*Avg*: 18.00; *SD*: 32.16; $CI_{0.05}$: [18.30, 19.71]	$Q_{0.25}$: 5; *Mdn*: 5 $Q_{0.75}$: 25
		Platform	**5** (5,873); **25** (2,327); **75** (320); **175** (142); **250** (113); N = 8,775	*Avg*: 18.76; *SD*: 36.55 $CI_{0.05}$: [18.00, 19.53]	$Q_{0.25}$: 5; *Mdn*: 5 $Q_{0.75}$: 25
		Service	**5** (9,487); **25** (4,956); **75** (858); **175** (1028); **250** (650); N = 16,979	*Avg*: 34.05; *SD*: 59.74 $CI_{0.05}$: [33.15, 34.95]	$Q_{0.25}$: 5; *Mdn*: 5 $Q_{0.75}$: 25
Estimated revenue range	Grouped; range represented by group average **$1M:** 1 M$ (or less) **$5M:** (1,10] M$ **$30M:** (10,50] M$ **$75M:** (50,100] M$ **$250M:** (100,500] M$	AI	**$1M** (931); **$5M** (1,138); **$30M** (52); **$75M** (6); N = 2,127	*Avg*: 3,865,068.2; *SD*: 4,606,137 $CI_{0.05}$: [3,669,206; 4,060,929]	$Q_{0.25}$: 1M; *Mdn*: 5M $Q_{0.75}$: 5M
		Platform	**$1M** (2,409); **$5M** (2,776); **$30M** (180); **$75M** (26); **$250M** (17); N = 5,408	*Avg*: 4,990,569.5; *SD*: 15,255,475 $CI_{0.05}$: [4,583,889; 5,397,249]	$Q_{0.25}$: 1M; *Mdn*: 5M $Q_{0.75}$: 5M
		Service	**$1M** (3,721); **$5M** (3,279); **$30M** (785); **$75M** (96); **$250M** (80); N = 7,961	*Avg*: 8,408,617; *SD*: 26,413,412 $CI_{0.05}$: [7,828,314; 8,988,920]	$Q_{0.25}$: 1M; *Mdn*: 5M $Q_{0.75}$: 5M
Total amount of funding	Continuous Values	AI	N = 3,549	*Avg*: 18,375,499; *SD*: 163,799,459 $CI_{0.05}$: [12,984,668; 23,766,330]	$Q_{0.25}$: 333,666; *Mdn*:1.6M $Q_{0.75}$: 5.845.000
		Platform	N = 1,387	*Avg*: 21,336,39; *SD*: 246,667,547 $CI_{0.05}$: [8,343,636; 34,329,149]	$Q_{0.25}$: 100,000; *Mdn*: 500,000 $Q_{0.75}$: 250,000
		Service	N = 2,852	*Avg*: 33,505,429; *SD*: 416,917,899 $CI_{0.05}$: [18,197,788; 48,813,071]	$Q_{0.25}$: 143,875 *Mdn*: 993,500 $Q_{0.75}$: 5,000,000

Appendix 2: Startups by Different Locations and AI Startups by Different Business Categories

Startup size X	Business Metric Y	Group	Location	β	Adj-R²	N
No. of employees	Estimated size of revenue	AI	North America	0.23	0.08	963
			Europe	0.29	0.12	467
			Asia-Pacific	0.42	0.29	200
		Platform	North America	0.18	0.03	1,911
			Europe	0.25	0.06	593
			Asia-Pacific	0.44	0.14	573
		Service	North America	0.37	0.17	3,222
			Europe	0.38	0.20	808
			Asia-Pacific	0.37	0.16	728
Number of employees	Total sum of funding	AI	North America	1.20	0.39	1,509
			Europe	1.15	0.36	852
			Asia-Pacific	1.34	0.39	350
		Platform	North America	1.03	0.26	632
			Europe	1.12	0.27	210
			Asia-Pacific	0.89	0.12	139
		Service	North America	1.09	0.33	1,113
			Europe	0.98	0.25	458
			Asia-Pacific	0.88	0.16	349

Startup size X	Business Metric Y	Business Category	β	Adj-R²	N
Number of employees	Estimated size of revenue	Machine Learning	0.26	0.10	887
		Big Data	0.26	0.08	362
		Analytics	0.29	0.11	333
		SaaS	0.27	0.11	255
		FinTech	0.43	0.23	128
		Robotics	0.22	0.07	100
		E-Commerce	0.23	0.07	82
		Marketing	0.20	0.04	71
		Medical	0.34	0.16	48
		Autonomous Vehicles	0.29	0.15	37
Number of employees	Total sum of funding	Machine Learning	1.23	0.37	1,506
		Big Data	1.27	0.40	451
		Analytics	1.09	0.34	413
		SaaS	1.13	0.38	417
		FinTech	1.28	0.44	211
		Robotics	1.31	0.44	152
		E-Commerce	1.09	0.23	136
		Marketing	1.19	0.44	89
		Medical	1.32	0.42	85
		Autonomous Vehicles	1.37	0.50	69