RIETI Discussion Paper Series 24-E-021

The Fundraising of AI Startups: Evidence from web data

ZHU, Chen
University of Tokyo

MOTOHASHI, Kazuyuki
RIETI

RIETI
Research Institute of Economy, Trade & Industry, IAA

The Research Institute of Economy, Trade and Industry
https://www.rieti.go.jp/en/

RIETI Discussion Paper Series 24-E-021
February 2024

The Fundraising of AI Startups: Evidence from web data*

Chen Zhu (The University of Tokyo) Kazuyuki Motohashi (RIETI)

Abstract

Startups have emerged as pivotal innovators in the commercialization of AI technology. Nonetheless, these nascent enterprises often require substantial capital infusion to realize the economic returns from their innovations. This study examines the role of prototypes in facilitating their fundraising process. We utilized historical web content to identify the presence of prototypes and employed web traffic data to monitor their customer growth. Our findings indicate that prototyping positively affects the potential customer attraction process, signaling the feasibility and profitability of their business hypotheses to potential investors. In addition, as a technologically intensive industry, most AI startups begin with a technology-centric approach. While a technology-led starting point underscores competitiveness, it also inherently introduces uncertainty. We offer quantitative evidence demonstrating how prototyping acts as a moderating factor, reducing the impact of such uncertainty by expediting investor decision-making.

Keywords: AI startups; prototyping; entrepreneurial strategy; fundraising process

JEL codes: L26, M13, O32

* This study is conducted as a part of the project "Research on Digital Innovation Models" undertaken at the Research Institute of Economy, Trade and Industry (RIETI). The draft of this paper was presented at a DP seminar of the Research Institute of Economy, Trade and Industry (RIETI). I would like to thank participants of the RIETI DP Seminar for their helpful comments.

1. Introduction

The widespread use of AI technology has demonstrated its ability to enhance human productivity and economic growth, and transform many industrial fields (Furman and Seamans, 2019; Brynjolfsson et al., 2021). Big tech firms such as Google, Amazon, Apple, Meta, and Microsoft (Big Five) have invested tens of billions of dollars into AI and machine learning research (Kossuth and Seamans, 2018). The huge amount of investment brings large incumbents several advantages over small businesses: they can attract the best AI talents from academia (Woolston, 2022); have well-established complementary assets (e.g., large amounts of training data and computational resources), and have access to their own marketing channels. In this situation, how can artificial intelligence (AI) startups survive among established tech firms? The answer resides in a simple tenet: innovation is the currency, and agility is the key to success. Indeed, many startups in the field have emerged as trailblazers, advancing the boundaries of what is possible through machine learning and AI applications. Their agility allows them to quickly identify and fill demands in niche markets with innovative solutions (Weiblen and Chesbrough, 2015).

However, although startups have been recognized as one of the pivotal contributors to the development of AI solutions addressing practical issues, these endeavors typically require a substantial capital infusion to enter the market. Startups often raise funds from venture capital (VC) firms to extend their minimum viable products (MVPs) and scale their businesses and user bases (Davila et al., 2003). According to the AI Index 2018 and 2022 Annual Report, VC funding for AI startups increased by 350% from 2013 to 2017, and the trend continued to double until 2021 (Shoham et al., 2018; Zhang et al., 2022). In this regard, VCs have become one of the main channels through which AI startups can raise money.

This study examines the factors that facilitate the efficient acquisition of funding for AI entrepreneurs, namely, a shorter waiting time to attain financial support. As a prelude, we discuss why the timing of acquiring financial resources is important for early-stage ventures from two perspectives: capital-constrained entrepreneurs and the value-added roles of VC. Unlike established enterprises, fledgling firms usually lack adequate credit lines and have an advantageous cash position (Schwienbacher, 2007). At this stage, entrepreneurs play the role of a jack-of-all-trade, managing a wide range of operations, including product development, human resource management, and fundraising. Consequently, there is often a zero-sum issue with these responsibilities, in which the time devoted to fundraising may detract from product development (Hsu, 2007). In the early stages of their life cycle, most ventures burn more money than they earn. Financial distress arises when cash flows fall short of meeting liabilities. Hence, understanding the factors that contribute to a faster funding timeline can help startups allocate resources more effectively, thereby shaping informed fundraising strategies. Besides, scholars have long recognized the value-added role of venture capitalists (VCs) beyond being mere financial intermediaries. They serve as instrumental facilitators for emerging ventures to acquire the essential resources for successful growth, such as potential customers and strategic alliances (Gorman and Sahlman, 1989; Stuart et al., 1999; Hsu, 2004). Therefore, the timely acquisition of funding becomes pivotal, as it expedites the integration of resources and enhances the company's competitive position.

Prior research has identified several factors that may contribute to the fundraising of AI startups. Bessen et al. (2022) survey hundreds of AI startups and find that startups with proprietary data are more likely to secure future VC funding than those using publicly available data sources. Weber et al. (2022) manually examined the websites of a sample of 100 AI startups and created a taxonomy of different AI business models. Their work also indicates that distinct business models may be related to the different levels of difficulty in commercializing AI-enabled products or services. Motivated by our examination of AI startups' homepages, we posit that AI startups may use prototypes, often in the form of software or platform applications, to both attract potential customers and signal the feasibility and profitability of their business hypotheses.

Principal-agent conflicts persist until the interests or incentives of entrepreneurs and potential investors align (Leach and Melicher, 2010). The asymmetry in information accessibility between entrepreneurs and investors, with entrepreneurs holding private knowledge about their innovations, contributes to the opacity of the due diligence process. This opacity is further exacerbated by the 'black-box' nature of AI technology, which adds an additional layer of complexity to investors in uncovering hidden information. The fundamental logic of machine learning and AI lies in developing algorithms without explicit rule-based programming (i.e., intelligence), which learn patterns from data (i.e., training) and exhibit robust generalization for decision-making in unseen situations (i.e., prediction). In this sense, a prominent challenge for investors is to evaluate the model's real-world applicability and economic values, given their limited insight into the specific features of the developed AI models and the representativeness of the training data. Consequently, entrepreneurs are strategically positioned to selectively emphasize the merits of their innovation, while withholding their drawbacks[1].

Overarching common interests center on increasing the value of ventures, or at least enhancing the potential for it. Therefore, entrepreneurs can use experimentation to explore alternatives for transforming ideas into value-generating tools. A growing body of literature has identified prototyping as an experimental tool (Thomke, 1998; Terwiesch and Loch, 2004). Because prototyping is not a one-time task, entrepreneurs can use it as a strategy to learn about customer responses and iteratively optimize their products. Indeed, prototyping is strongly related to the nature of AI technology. As AI products mainly rely on machine learning models, training on initial datasets does not guarantee that they can effectively solve real-world challenges. In this light, prototyping may serve as an entrepreneurial strategy, enabling entrepreneurs to not only assess market alignment but also gather secondary data from potential customers, thereby enhancing the accuracy of their models. Additionally, an analysis of startups' historical web content revealed that approximately 48% of AI startups had a developed prototype, a percentage significantly exceeding the 23% ratio reported by Tyebjee and Bruno (1984), with a focus on various industries. This can be attributed to the established cloud computing infrastructure and the abundance of open-source tools and libraries[2]. Consequently, unlike traditional industries that rely heavily on external suppliers for physical components or materials, AI startups can develop prototypes internally, leveraging in-house talent and expertise.

[1] The 'black-box' nature of AI renders the underlying algorithm less interpretable. Nonetheless, the development team typically possesses a certain intuition about the model's capabilities and limitations (i.e., a form of tacit knowledge tied to the innovation).
[2] A pivotal milestone in the history of cloud computing was the launch of Amazon Web Services (AWS) in 2006.

Existing literature suggests that prototyping can assist nascent enterprises in signaling the feasibility of their technology development to potential investors, thereby facilitating the process of securing external financing (Audretsch et al., 2016). We expand this logic by examining how prototyping attracts potential customers, which then forms a profitability signal for investors. Our line of argument begins with the theoretical development of a prototype that may serve as a boundary object between entrepreneurs and potential customers. A prototype mitigates buyer-seller information asymmetry by providing a shared and comprehensible space for development teams and early adopters to exchange values. The presence form of a prototype may vary, depending on the combination of distinct business models and customer segments. For example, when catering to customers with high technical proficiency, prototyping an application programming interface (API) allows for more direct interaction and integration of AI services into their own undertakings. However, for more general customers, entrepreneurs may opt to 'black-box' their underlying algorithmic models and systems into low-code or no-code tools, effectively reducing the learning costs for adoption. Hence, the presence of prototypes enhances the transparency of the proposed business values, expediting the decision-making time of early adopters. The establishment of a potential user base then indicates the successful alignment of market expectations, signaling the profitability of the proposed business hypotheses to external investors.

Historically, it has been challenging to evaluate the proposed mechanism empirically because of limited data availability on the innovation activities and customer disclosures of young firms. Firms are not required to disclose information until they go public; therefore, early research in entrepreneurial finance often relies on surveys[3]. We overcome these limitations by combining five distinct data sources: web scraping data; two commercial databases, Crunchbase and SEMrush; a patent database provided by the United States Patent and Trademark Office (USPTO)[4]; and the USPTO Trademark Case Files dataset[5] by Graham et al. (2013). We collected data on 396 AI startups in the United States from Crunchbase. We then applied the Internet Archive's Wayback Machine API to scrape their historical web contents. Li et al. (2018) demonstrated that web content could be an alternative source for monitoring SMEs' business activities because they often disclose their products and technologies through their homepages to attract potential customers and investors. Web scraping data were used to examine whether there was a prototype before receiving early-stage funding (i.e., Series A in particular). In addition, historical web traffic data collected from SEMrush are used as a proxy for monitoring companies' potential customer attraction processes. Finally, we create a rich set of control variables from both the startup and VC sides. Figure 1 illustrates the research framework.

[3] According to a survey on entrepreneurial finance studies (Cumming and Johan, 2017), it is not mandatory for new ventures to formally disclose their business activities in most countries, with exceptions including Scandinavian countries and certain continental European nations like France and Belgium.

[4] See https://patentsview.org/download/data-download-tables

[5] See https://www.uspto.gov/ip-policy/economic-research/research-datasets/trademark-case-files-dataset

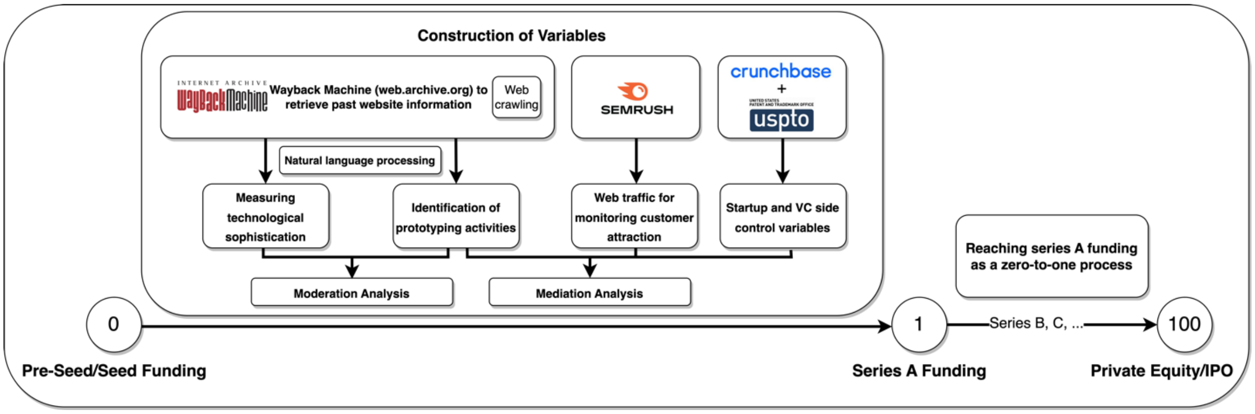

Figure 1. Research Framework

2. Theoretical Framework

2.1. Prototyping as an Entrepreneurial Strategy

Technological innovation creates opportunities but also introduces uncertainty to fledgling ventures. On the one hand, ventures with advanced technological expertise may aim to evolve into technology-intensive enterprises, and this competence can be a sustained competitive edge (Colombo and Grilli, 2005; Aggarwal et al., 2015), resulting in significant economic returns and consequently enhancing their bargaining power for more favorable valuations from VC (Hsu, 2004). On the other hand, high levels of technological innovation also yield pronounced uncertainty among decision-makers (Zaltman et al., 1973). Nelson and Winter (1977) noted that research and development (R&D) inherently involves uncertainty because the optimal path to commercial success is rarely clear.

In most studies, uncertainty pertains to the future payoffs from VC investments. Technology-intensive companies often require a lengthy time frame and a substantial allocation of resources for exploratory development. Greater knowledge complexity in the early development stage extends the lead time to the first product shipment, making the anticipated financial gains uncertainly placed in the distant future (Schoonhoven et al., 1990). Such uncertainty in future returns stems mainly from two distinct types of risks: technical risk and market risk. Technical risk concerns the feasibility of innovation, specifically, whether the innovation is technically completable (Pindyck 1993). The second type of risk pertains to profitability, concerning whether the innovation aligns with market expectations. Although VC investors always buy with a certain uncertainty, they seek agents who can mitigate these risks.

Prior research theorizes that organizational learning via trial-and-error is a promising strategy for entrepreneurs facing uncertainty (Levinthal, 2017; Gans et al., 2019). Technical risk can be resolved by creating prototypes (i.e., a ‘just-do-it’ strategy), which directly demonstrates the feasibility of the proposed idea. However, the value distribution of an idea is contingent not only on its technical development but also on its alignment with market demand (Rosenberg, 1994). Stated differently, even when conditional on a technically feasible idea, predicting economic returns remains challenging without market testing. In this regard, prototyping is not a one-time task but an experimentation tool for learning market responses iteratively. Through market tests using prototypes, entrepreneurs can gain insights into the alignment of their innovation with customer demands. Feedback from early adopters then reshapes product innovation and improvements (von Hippel, 1986).

In summary, the logic presented above suggests that although startups with a high level of technological sophistication may attain more advantageous valuations from VC, the inherent uncertainty and risks could cause investors to hesitate for an extended period. In addition, limited performance track records also result in notable information asymmetry for investors when assessing the quality of these emerging companies (Shane and Stuart, 2002). However, startups can adopt prototyping as a strategy to demonstrate the feasibility of the proposed business hypothesis and engage in the market, thereby initiating the customer discovery phase. A promising product-market fit can then mitigate market risk and demonstrate the potential to generate profits.

H1. Highly sophisticated startups may gain better valuations, but the associated high risk can delay investors' decisions moderated by the adoption of prototypes.

2.2. Prototypes as a Boundary Object

It is no secret that enterprises (customers) understand that they have to adapt AI in some form, but are struggling to figure out how to do so. The theory of Rogers (2003) on the innovation decision process delineates the stages in which decision-making units (e.g., individuals or organizations) traverse the course of accepting and adopting innovations. The process essentially unfolds the mechanism geared toward diminishing information asymmetry, wherein an individual must be endowed with knowledge to reduce uncertainty about the merits and demerits associated with innovation. It initiates an individual's endeavor to amass a sufficient level of knowledge to understand the nature of the innovation and its operational functionality. When it comes to innovations characterized by a significantly higher degree of technical complexity, the requisite knowledge for adoption is considerably greater compared to those with lower complexity. In this case, by drawing on the lean startup theory, AI startups can embed their foundational algorithmic models into a prototype (i.e., minimum viable products [MVPs]), which usually manifests in the form of software or platform applications (Shepherd and Gruber, 2021; Blank and Eckhardt, 2023). The prototype functions as a boundary object, establishing a shared knowledge base between entrepreneurs and customers (Nonaka, 1994; Carlile, 2002). A boundary object is an artifact that acts as a bridge in the flow of knowledge among various agents (Star, 1989; Carlile, 2004), ultimately facilitating the sense-making process (Grichnik et al., 2016). The prototype concretizes the intangible underlying knowledge and technology, enhancing its accessibility to potential customers to grasp and apply. Before reaching the decision stage, the inherent uncertainty regarding innovation persists until individuals either experiment with it or are convinced by trusted entities (i.e., the persuasion stage). Trials in which customers are allowed to experience the partial functionality of an innovation are the most straightforward approach to expediting the decision-making process (Rogers, 2003). In this context, AI startups that have already crafted prototypes can present and deliver a demo to potential customers, showcasing the capabilities and advantages of their AI solutions. If innovation aligns with customer expectations or exhibits a specific level of relative advantage, the majority of potential customers are likely to proceed with an adoption decision. Meanwhile, experimenting with a prototype can just as logically result in a rejection decision, as in the case of adoption. Rejection feedback helps entrepreneurs by providing insights into refining existing prototypes or even reconfiguring their initial business hypotheses (Markman et al., 2008).

In summary, prototypes can serve as boundary objects that alleviate information asymmetry between entrepreneurs and potential customers, thereby accelerating the innovation decision process. Using the organic traffic data as a proxy for monitoring customer attraction process, we argue that AI startups with a high level of product readiness (i.e., prototypes that are ready for market) will significantly reduce the time required to accumulate sizeable potential customers.

H2. AI startups with developed prototypes can attract potential customers more rapidly. (Prototype → Customer growth).

2.3. Informational Asymmetry and Profitability Signals

Entrepreneurship is intrinsically tied to experimentation, as entrepreneurs seek to fulfill existing or create new demands for economic gain by experimenting with novel ideas (Rosenberg, 1994; Kerr et al., 2014). Founders address uncertainties by formulating business hypotheses, validating them in the market, and analyzing the results (Blank, 2013). For early-stage companies, obtaining external resources is both a critical issue for sustaining their experimentation financially and a difficult challenge in the sense that their assets are mostly intangible, such as knowledge and technology (Hsu, 2004). As it is impossible to evaluate the economic returns from these experiments, the problem of information asymmetry is particularly pronounced at this stage, which induces severe agency costs for VC investors (Gompers, 1995). To mitigate risks, VC firms often employ stage financing, which allows them to retain the right to suspend future capital if the company fails to meet predetermined milestones (Admati and Pfleiderer, 1994; Gompers, 1995).

Signaling is a means of reducing information asymmetry across parties (Spence, 1973). Drawing heavily on patents, Hsu and Ziedonis (2008) showed that patents can assist investors in estimating the value of early-stage startups. From a legal perspective, patents secure exclusive rights for appropriating innovations and obtaining future benefits, enabling entrepreneurs to employ them as signals to attract potential investors (Long, 2002). Nevertheless, there is a disparity in the conclusions regarding the impact of patents. Gompers and Lerner (2001) note that patents and trademarks, although more tangible than ideas, remain insufficient on their own to secure financing from most lenders until they are complemented by specialized assets. In line with this argument, Audretsch et al. (2016) suggest that innovative ventures could develop a prototype as a complementary means of signaling the feasibility of their knowledge-based business hypotheses. As Duc and Abrahamsson (2016) stated in an interview with software startup founders, “It is important to show investors that you are committed, and past the idea stage. Without a prototype, most professional investors will not take you seriously.” Although investors specializing in certain industries possess domain-specific knowledge, it is evident that a prototype always supersedes a lengthy hundred-page document when conveying business plans.

Prototypes and business models are also interrelated in the context of new ventures. The business model encapsulates the underlying logic of how companies create, deliver, and capture value when commercializing their ideas and technologies (Teece, 2010; Chesbrough, 2010). As per Blank’s (2013) conceptualization, a business model can be viewed as a theory, with the startup serving as the vehicle for testing. In this regard, a prototype with sufficient critical features aligned with the entrepreneur's initial vision can accelerate the validation of their business hypotheses. The quick response of early adopters can streamline the monitoring and evaluation processes of VC investors in their decision-making (Tian, 2011). According to existing literature, AI startups with developed prototypes are categorized as AI tool and platform providers (Weber et al., 2022). With predefined products, startups of this type can employ either a subscription- or a usage-based pricing model. Market-ready products can be easily understood and evaluated by customers, resulting in shorter sales cycles. However, AI development facilitators often help customers build customized AI solutions. Given that startups in this category need to provide more tailored and integrated work to meet the specific needs of individual customers, this translates into a much longer sales cycle and scaling challenges, particularly in the early stages of their operations. Therefore, we hypothesize that AI startups with developed prototypes will secure early-stage funding more rapidly.

H3. AI startups with developed prototypes will spend less time securing early-stage funding (Prototype $\rightarrow$ Fundraising process).

As discussed earlier, the presence of a prototype serves as an instrument for mitigating information asymmetry by signaling the feasibility of a business hypothesis. However, the quality of such signals is unexamined. For early-stage financing, our primary focus in this study is Series A funding, a highly competitive round where investors are not simply looking for great concepts but for revenue-generating business models and an established user base. The potential to build a stable user base is an indicator of high cue validity for prospective investors, indicating that a startup has the capacity to scale up its business and generate sustainable profits. Cue validity is a probabilistic metric that quantifies the likelihood of a piece of information (i.e., a particular cue) accurately predicting a specific outcome (Rosch, 1975). In this context, the process of customer validation entails an examination of the extent to which the proposed prototype aligns with market expectations, signaling whether there is a scalable and profitable business model on the horizon (Blank, 2003; Ries, 2011). For the purpose of venture screening, the presence of a substantial base of potential customers within a startup should exhibit higher cue validity than its absence. Groening et al. (2016) suggested that investors may incorporate a firm's treatment of customers as a pivotal factor in assessing its value.

The concept of boundary objects argues that prototypes can accelerate decision-making process by establishing a sharable and understandable space between entrepreneurs and potential customers. Accelerated customer acquisition can result from either a direct decision to adopt or rejection decisions that prompt founders to realign with the market needs[6]. The growing number of potential customers then signals the profitability of the envisioned prototype to investors. Therefore, we hypothesize that potential customer growth plays a mediating role in accelerating the fundraising process of AI startups that have already prototyped their innovations.

[6] In this study, we utilize organic web traffic data to monitor customer growth. It is essential to clarify that web traffic includes both actual customers and potential customers, some of whom may be at the persuasion or confirmation stage. Prior research has recognized that (actual) customer relationships can be an important firm resource (Gouthier and Schmid, 2003; Park et al., 2017). Since the conversion rate of monthly visits is unknown, we primarily consider it as a proxy for potential customers and argue that a higher growth rate of potential customers can signal a promising outlook for startups' profitability.

H4. AI startups with developed prototypes can expedite the establishment of a potential customer base, which, in turn, hastens the process of securing early-stage funding (Prototype → Customer growth → Fundraising process).

3. Data Collection and Variables

3.1. Crunchbase Data

To test the proposed hypotheses, we developed a novel dataset comprising 396 U.S. AI startups founded between 2010 and 2017. Crunchbase is a crowd-sourced database that provides a rich set of business information on companies, founders, and investors. It has been widely used in management and economics research in general (Dalle et al., 2017) and entrepreneurship strategies in particular (Koning et al., 2022; Guzman and Li, 2023). Initially, we extracted companies with AI industry labels from Crunchbase. For each startup, we collected its detailed profile data, including the official website, founding members, and fundraising history. Specifically, this study targets round A, a stage where a startup often already has its business model and demonstrates the potential to generate more profits and scale up, representing a pivotal 'zero-to-one' phase. Thus, when examining each startup's financial records, we retain all startups that have received Series A funding; however, companies supported by corporate venture capital (CVC) are excluded. This is to rule out the potential bias induced by startups backed by large businesses. The timing of obtaining Series A funding, measured monthly, is the one of the primary dependent variables in the following regression analyses. Besides, we also collected the amount of money received during Series A as another dependent variable. To address the skewed distribution of the valuation data, the founding amounts were specified as natural logs in the multivariate regression. Other profile data were used to construct several control variables, which are discussed later.

3.2. Prototype Identification with Wayback Machine

Prior literature has relied on survey data to track startups' prototyping activities (Tyebjee and Bruno, 1984; Audretsch et al., 2012). In this study, we propose a new approach to identify whether a startup has developed prototypes prior to Series A funding by examining the occurrence of a set of AI product keywords in its historical web content, which also benefits from the comprehensive Internet coverage of AI startups[7]. To achieve this, we utilized the Wayback Machine APIs to scrape the web records preceding the specified timestamp[8]. Since 2001, the Internet Archive Wayback Machine has stored more than billions of web pages worldwide. After that, we follow Arora et al.'s (2015) approach to clean and integrate the scraped web pages, which are then used to construct the prototype dummy.

3.3. Technological Sophistication with Wayback Machine

Existing literature measures the level and sophistication of technological innovation in new ventures through the utilization of survey data, such as interviews with their founders (Eisenhardt and Schoonhoven, 1990; Sapienza and Gupta, 1994). This study proposes a web-based indicator

[7] The keywords used to identify the existence of prototypes prior to Series A are shown in Appendix A.
[8] The code for retrieving historical web records within a predetermined time window can be found here: https://github.com/zhujohn9604/WaybackMachineAPI.

to quantify each startup's technological sophistication and efforts. Startups that dive deeply into R&D tend to emphasize their technological strengths by sharing detailed information about their technologies on their websites. These technical efforts were captured by counting the occurrence of a list of AI-related technical keywords provided by the WIPO[9]. These keywords are selected by examining AI functional applications and techniques. In addition, the Wayback Machine APIs may return multiple records for the same URL, and we only considered the record with the highest occurrence value for each URL. Finally, we calculated the average based on the startup's unique records.

3.4. Potential Customer Tracking with Web Traffic Data

Obtaining sufficient data to monitor venture growth has proven to be challenging. Although emerging ventures may disclose their customer relationships in some financial announcements, privacy policies often limit such disclosures, leaving only a few mentions. Kerr et al. (2014) suggests that web traffic data could be a potential proxy for tracking potential customers. Inspired by this work, we collected organic web traffic records using the SEMrush Magic tool. For each startup, we downloaded historical records of monthly visits landed on the homepage. Then, the modified relative strength index (MRSI) was adopted to measure the speed and change in customer growth. Specifically, we have

$$MRSI = \frac{1}{T}\left(100 - \frac{100}{1+RS}\right),$$

$$RS = \frac{Average\ Gain}{Average\ Loss}$$

where the average gain refers to the average increase in the number of monthly visitors over the selected period T, and the average loss indicates the average decrease over the same period. For each startup, we estimated the value of MRSI from its founding date to the date of receiving Series A funding. This variable is denoted as *Customer* in the formal analysis.

3.5. Variable Construction

Our primary independent variables are *Prototype*, *Technology*, and *Customer*, as introduced above. Besides, we also include both startup-side and VC-side control variables to alleviate the potential spurious relationships between the dependent variables and our key variables. On the startup side, we first controlled for each startup's basic information, including age, geography, and application fields. Crunchbase provides the headquarters' regions and industry classifications for each company. Our sample includes six regions (e.g., Greater Boston, Greater New York, San Francisco Bay Area, etc.) and seven application fields (e.g., biotechnology, finance, entertainment, etc.). We controlled for application fields because startups specializing in different industries may deal with different transaction types (e.g., individual customers versus organizational customers), which, in turn, leads to heterogeneity in the potential customer attraction process. Besides, human capital and knowledge are the two major inputs in the AI industry. Thus, a solid relationship with universities will provide startups with continuous access to AI talents and cutting-edge scientific

[9] The keywords used to measure a startup's technical sophistication are shown in Appendix A.

output (Lutchen, 2018; Hannigan et al., 2022). Indeed, since many startups are willing to disclose their university linkages on their websites, we build a variable called *University* to capture this[10]. In addition, we surveyed several other factors that influence a new firm's performance and chances of obtaining VC funding. Eisenhardt and Schoonhoven (1990) suggest that a larger founding team size is positively associated with a new firm's future growth. In addition, entrepreneurship literature shows that prior entrepreneurial experience can act as a positive signal for investors (Colombo et al., 2004). Entrepreneurs with more funding experience are, therefore, more likely to receive funding from the prospective VC (Kirsch et al., 2009). We measured founders' experience based on the number of companies founded before the establishment of the target startup. Founding team experience was calculated using the average founding experience of each founder. Finally, intellectual property such as patents and trademarks was also considered. Mann and Sager (2007) found a robust positive relationship between patenting activity and VC fundraising for software startups. Krasnikov et al. (2009) suggest that trademarks measure a firm's brand-building efforts, which are positively associated with financial value. Based on this, we include the number of patents and trademarks filed by startups before receiving Series A funding.

It is a common practice for investors to form a syndication to provide funding resources to startups. Syndication allows investors to share risks and evaluate their assessments against other knowledgeable entities (Gompers, 1995; Lerner, 1994). Hence, we first control for the number of investors in each transaction for VC-side consideration. Crunchbase provides investor-type information for each investment firm. We include different VC types to consider the risk preferences of different entities (e.g., government and family offices). Finally, Tian (2011) suggests that geographic distance affects VC investors' decision-making. A shorter distance between entrepreneurial firms and investors implies lower monitoring costs, which increases the likelihood of funding. Appendix B describes the detailed procedure used to construct the distance measure.

3.6. Summary Statistics

Table 1 presents the summary statistics of our data. The meaning of each variable is associated with the interpretations in the above sections. Tyebjee and Bruno (1984) examined startups across diverse industries and reported that nearly 23 percent of startups possess a fully developed prototype before securing VC financing. Our key independent variable *Prototype* has a mean of 0.48, indicating that 48% of the startups in our sample had developed prototypes before receiving Series A funding. This higher ratio can be attributed to readily accessible complementary assets, such as open-source development tools and cloud computing resources, which facilitate easier and faster AI prototype development. In addition, three categorical variables are absent from the table: *Geography*, *Application Fields*, and *VC Types*. According to Crunchbase, the startups in our sample are headquartered in six major regions: San Francisco Bay Area, Silicon Valley (21%), Greater New York Area, East Coast (14%), San Francisco Bay Area, West Coast (26%), Greater Boston Area, East Coast (9%), Greater Los Angeles Area (6%), and others (24%). Moreover, in terms of Crunchbase industry group lists, their application fields can be classified into information technology (20%), biotechnology (16%), finance (20%), traditional industries (20%), entertainment (7%), business intelligence (23%), and green technology (4%). Finally, some VC

[10] The university linkage is quantified based on the frequency of keyword occurrence in a company's historical web content, with the list of keywords provided in Appendix A.

investors are multi-labeled; thus, the *VC Type* specifies the potential variations, such as VC only (77%), VC and PE (13%), family office VC (2%), government-backed VC (1%), and others (7%).

	Mean	Std. dev.	Minimum	Maximum
Dependence variables				
Time to Series A funding	50.00	18.20	11	123
Log(Series A valuation)	15.94	0.73	11.29	17.53
Key independent variables				
Prototype	0.48	0.50	0	1
Technology	3.24	3.37	0	41.31
Customer	1.43	0.99	0.03	8.25
Startup side control variables				
Age	7.29	1.89	5	12
University	2.74	4.63	0	48
Founders	2.48	1.15	1	7
Experience	0.30	0.55	0	4.5
Patents	0.62	1.79	0	16
Trademarks	0.91	2.55	0	31
VC side control variables				
Syndication	4.23	3.01	1	19
Distance	2.45	1.62	0	5

Table 1. Summary Statistics

4. Empirical Analysis

4.1. Prototype as a tool to Mitigate Technological Uncertainty

This section examines the moderating effect of prototyping on technological uncertainty. We perform ordinary least squares (OLS) estimations, wherein we first regress the level of technology sophistication against the timing of Series A funding and the associated funding amount. When regressing against the valuation amount in Series A, we included an important control variable, pre-money valuation, which is the cumulative amount of capital raised by entrepreneurs before the Series A round (also in logarithmic form).

In addition to the OLS regression, we tested the proposed hypotheses using survival models. Survival time measures the duration for which a startup remains in operation before receiving Series A funding. We consider an accelerated failure time (AFT) model to estimate the impact of the proposed factors on startup survival time:

$$S(T) = S_0(\phi T) \tag{1}$$

where ϕ is an acceleration factor and can be written mathematically as

$$\phi = \exp\,(\beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n). \tag{2}$$

This implies that Equation (2) reduces to $S_0(T)$, which is referred to as the baseline survival function, when all covariates are equal to zero. The AFT model is commonly represented in a log-linear form as follows:

$$\log(T) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n + \varepsilon \quad (3)$$

where β_0 is the baseline survivor value and ε measures the residual variability in the survival times. We adopt a Weibull distribution to model the survival times since it achieves the best performance in terms of Akaike's Information Criterion (AIC) among others (e.g., Log-Normal, Generalized Gamma). The AFT model is equivalent to the proportional hazards (PH) model when the survival times are assumed to follow a Weibull distribution, but the interpretation of estimated coefficients differs in sign (Bradburn et al., 2003).

Before discussing the empirical results, Table 2 shows the pairwise correlations among the covariables of interest. No critical multicollinearity was observed as the absolute values of the correlations were less than 0.3. We estimate the proposed specification using OLS and AFT models, and present the results in Table 3. As can be seen in column (2), with all variable controlled, our findings indicate that AI startups with a greater level of technological sophistication tend to secure higher valuations during the Series A round ($\hat{\beta} = 0.014, t = 1.57, p < 0.10$). Regarding the timing of attaining financial support, the positive and statistically significant coefficients for *Technology* in columns (4) and (6) indicate that higher technological uncertainty is associated with a longer waiting time prior to Series A. In the last specification, column (7) of Table 3, the interaction effect between technological sophistication and prototype is also included. This interaction term tests the prediction that intensive development of technology introduces uncertainty, which may extend investors' time spent on deliberation, whereas the presence of prototypes can moderate the impact of this uncertainty on investor deliberation time. The coefficient of the interaction term is negative and statistically significant ($\hat{\beta} = -0.014, p < 0.05$). Figure 2 shows that high technical sophistication is associated with an extended waiting time (the 95% confidence interval is above zero), and prototyping can significantly reduce this waiting time, as reflected by a shift in the 95% confidence interval towards negative values. In summary, the results support the first hypothesis that highly sophisticated startups may secure more favorable valuations. However, the associated high risk often leads to extended investor deliberation, which can be moderated by adopting prototypes.

	Prototype	*Technology*	*Customer*	*Age*	*University*	*Founders*	*Experience*	*Patents*	*Trademarks*	*Syndication*	*Distance*
Prototype	1										
Technology	0.146	1									
Customer	0.201	0.025	1								
Age	-0.118	-0.083	-0.261	1							
University	0.046	0.101	-0.060	0.074	1						
Founders	-0.001	-0.045	0.013	-0.067	-0.030	1					
Experience	0.031	-0.006	0.079	0.034	0.033	-0.082	1				
Patents	-0.033	-0.070	-0.016	0.019	-0.001	0.074	0.053	1			
Trademarks	0.024	0.025	0.094	-0.008	0.007	-0.010	-0.016	0.246	1		
Syndication	0.006	0.110	0.110	-0.086	-0.022	-0.082	0.043	0.038	0.010	1	
Distance	-0.069	-0.012	-0.038	-0.071	0.038	0.053	-0.003	-0.033	0.083	0.002	1

Table 2. Correlations

	Dependent variable: Log (Series A valuation)		Dependent variable: Time to Series A				
	OLS		OLS		Weibull AFT		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Technology	0.023**	0.014*	0.262	0.429*	0.007	0.005*	0.016***
	(0.010)	(0.009)	(0.255)	(0.246)	(0.004)	(0.003)	(0.006)
Prototype	-0.014	0.010	-6.382***	-4.469**	-0.027	-0.036**	0.006
	(0.083)	(0.067)	(1.850)	(1.835)	(0.023)	(0.018)	(0.026)
Technology x Prototype							-0.014**
							(0.006)
Startup-side control variables							
Pre-money valuation		0.263***					
		(0.050)					
Age		-0.063***		2.705***		-0.068***	-0.068***
		(0.020)		(0.550)		(0.005)	(0.005)
University		0.003		0.068		0.001	0.001
		(0.007)		(0.174)		(0.002)	(0.002)
Founders		-0.023		-0.499		-0.010	-0.011
		(0.027)		(0.773)		(0.007)	(0.007)
Experience		-0.034		-2.688*		-0.040**	-0.042**
		(0.076)		(1.453)		(0.017)	(0.017)
Patents		0.063***		0.329		-0.001	-0.001
		(0.018)		(0.370)		(0.005)	(0.005)
Trademarks		0.013		-0.370		-0.003	-0.003
		(0.012)		(0.286)		(0.004)	(0.004)
Application field	N	Y	N	Y	N	Y	Y
Geography	N	Y	N	Y	N	Y	Y
VC-side control variables							
Syndication		0.040***		0.187		0.004	0.006
		(0.011)		(0.344)		(0.006)	(0.006)
Distance		0.021		0.802		0.003	0.003
		(0.023)		(0.567)		(0.003)	(0.003)
VC Type	N	Y	N	Y	N	Y	Y
Observations	309	309	396	396	396	396	396
AIC	609	565	3400	3400	-100	-220	-220

Notes: The number of observations for columns (1) and (2) is 309 because some startups' funding amount information is missing. Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3. Regression Results on Technology Sophistication

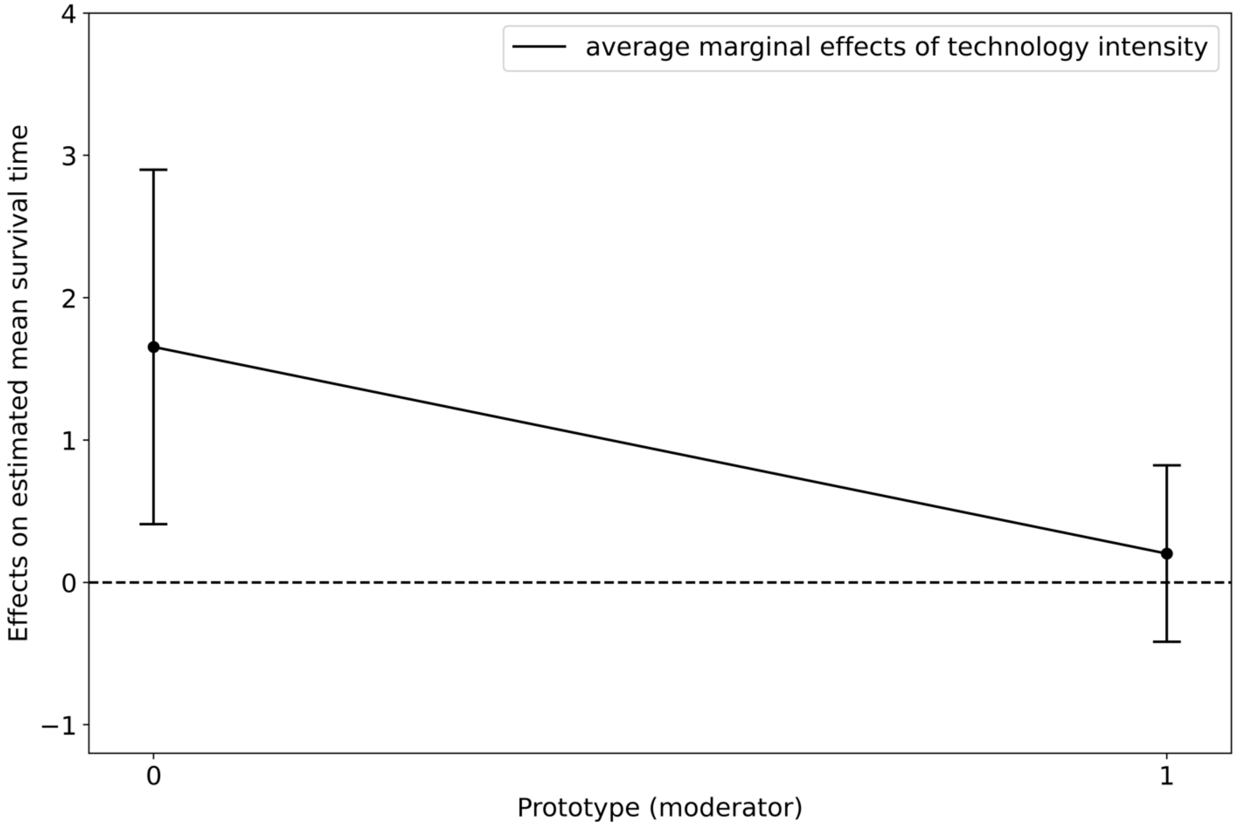

Figure 2. The Moderating Effects of Prototyping

4.2. Econometric Models of Prototyping, Potential Customer Attraction, and Fundraising Efficiency

We then examined the impact of prototyping on attracting potential customers, which, in turn, mediates more efficient fundraising. Following the mediation framework proposed by Baron and Kenny (1986), our analysis consisted of two parts. The main effect analysis specifies the direct impact of prototyping on fundraising efficiency, and the mediation effect analysis examines how the independent variable (*Prototype*) affects the dependent variable (*Time to Series A*) through the mediator (*Customer*). Specifically, we test the direct impact by estimating the OLS model as follows:

$$T_i = \beta_{10} + \beta_{11}Prototype_i + \sum_j \beta_{1j}X_j + \varepsilon_{1i} \tag{4}$$

where T represents the duration until Series A funding is obtained, and our interest is β_{11}, the direct impact of prototyping on fundraising time. The mediation effect analysis included a two-stage least squares (2SLS) estimation:

$$Customer_i = \beta_{20} + \beta_{21}Prototype_i + \sum_j \beta_{2j}X_j + \varepsilon_{2i} \tag{5}$$

$$T_i = \beta_{30} + \beta_{31} Prototype_i + \beta_{32} Customer_i + \sum_j \beta_{3j} X_j + \varepsilon_{3i} \tag{6}$$

In this specification, equations (5) and (6) examine the mediation path for potential customer attraction. The estimated coefficient $\widehat{\beta_{21}}$ is the estimate of the effect of prototyping on attracting potential customers, and $\widehat{\beta_{32}}$ is the estimate of potential customer attraction on fundraising efficiency when controlling for prototyping.

Table 4 shows the results of the analysis. As observed in column (2), a positive and statistically significant association exists between prototyping and customer growth ($\hat{\beta} = 0.399, t = 3.99, p < 0.01$), suggesting that prototyping may facilitate the customer acquisition process. In a test of main effect, we find that prototyping significantly decreases the time of securing Series A funding ($\hat{\beta} = -6.128, t = -3.40, p < 0.01$). When controlling for customer growth, the effect of prototyping becomes insignificance ($\hat{\beta} = -1.840, t = -1.27, p = 0.20$), and customer growth significantly reduces the waiting time ($\hat{\beta} = -10.755, t = -10.30, p < 0.01$).

To mitigate potential bias induced by confounding variables, we estimate the coefficients of the same specification by incorporating a set of control variables from both the startup and VC sides. The positive relationship between prototyping and customer growth remains significant ($\hat{\beta} = 0.333, t = 3.36, p < 0.01$). One may notice that only startup-side control variables are included in this estimation. Besides, the main effect of prototyping on waiting time stays consistent ($\hat{\beta} = -4.469, t = -2.44, p < 0.05$). When controlling for customer growth, prototyping loses significance ($\hat{\beta} = -1.209, t = -0.78, p = 0.44$), and customer growth significantly improves the fundraising efficiency ($\hat{\beta} = -9.646, t = -9.14, p < 0.01$).

We also conducted the Sobel test (Sobel, 1982) with a 2SLS estimation, and all variables were controlled. The results of the Sobel test confirm that the mediation effect of customer acquisition is statistically significant ($Z = -3.351, p < 0.01$), and the proportion of the total effect being mediated is 0.73. The Sobel test assumes that the product of $\widehat{\beta_{21}}$ and $\widehat{\beta_{32}}$ should be normally distributed, which might be too strong. Hence, we further validated the estimated indirect effects using bootstrapping (Hayes, 2009), which does not rely on the assumption of normality. Bootstrapping with 500 replicates was applied to improve sample representativeness and re-estimate the mediation effect. The results show a statistically significant indirect effect ($\hat{\beta} = -3.32, p < 0.01$), with zero excluded from the confidence interval ($C.I. = [-5.097, -1.347]$). Thus, we found strong evidence of mediating effects.

Taken together, the empirical results demonstrate that young enterprises can leverage prototypes to expedite the customer attraction process, which in turn serves as a profitability signal to potential investors. Finally, as supplementary evidence, Figure 3 illustrates the monthly visit dynamics of AI startups with prototypes compared with those without, where the x-axis represents the months preceding Series A. Startups with prototypes show significantly higher monthly visits, and the trend becomes sharp when it is close to the date of Series A.

	OLS			OLS		
	(1)	(2)	(3)	(4)	(5)	(6)
	$P \rightarrow T$	$P \rightarrow C$	$P, C \rightarrow T$	$P \rightarrow T$	$P \rightarrow C$	$P, C \rightarrow T$
Prototype	-6.128***	0.399***	-1.840	-4.469**	0.333***	-1.209
	(1.801)	(0.100)	(1.444)	(1.835)	(0.099)	(1.557)
Customer			-10.755***			-9.646***
			(1.044)			(1.055)
Startup-side control variables						
Technology				0.429*	-0.006	0.324*
				(0.246)	(0.013)	(0.193)
Age				2.705***	-0.135***	1.382***
				(0.550)	(0.025)	(0.477)
University				0.068	-0.010	-0.020
				(0.174)	(0.007)	(0.160)
Founders				-0.499	0.008	-0.358
				(0.773)	(0.038)	(0.675)
Experience				-2.668*	0.137	-1.497
				(1.453)	(0.083)	(1.198)
Patents				0.328	-0.020	0.089
				(0.370)	(0.024)	(0.380)
Trademarks				-0.370	0.039	0.055
				(0.286)	(0.027)	(0.263)
Application field	N	N	N	Y	Y	Y
Geography	N	N	N	Y	Y	Y
VC-side control variables						
Syndication				0.187		0.389
				(0.344)		(0.254)
Distance				0.802		0.409
				(0.567)		(0.489)
VC Type	N	N	N	Y	N	Y
Observations	396	396	396	396	396	396
R^2	0.03	0.04	0.36	0.19	0.16	0.42

Notes: The pivotal variables are denoted by three abbreviations: *P (Prototype)*, *C (Customer)*, and *T (Timing)*. Robust standard errors are reported in parentheses. * p < 0.10, ** p < 0.05, *** p < 0.01.

Table 4. Mediation Analysis

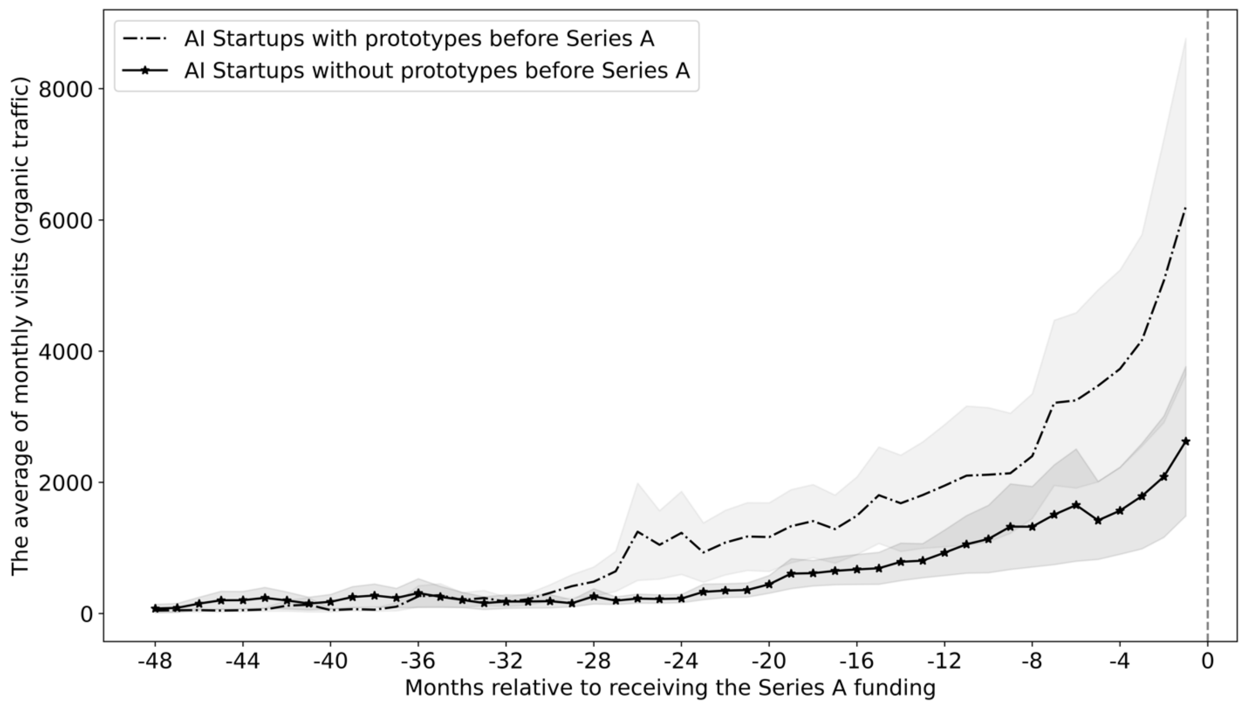

Figure 3. Monthly Visit Dynamics

4.3. Alternative Specifications and Robustness Checks
4.3.1. Survival Models

To assess the robustness of our results further, we tested the mediation effects using survival models. We estimate the Weibull AFT models, including both startup and VC-side control variables (Table 5). As can be seen in column (2) shows, the negative and statistically significant coefficient for *Prototype* indicates that developing prototypes accelerates the process of securing early stage funding for startups. Specifically, prototyping resulted in a 3.5% reduction in the waiting time[11]. When controlling for customer growth, the effect of prototyping becomes insignificance ($\hat{\beta} = -0.017, p = 0.28$), and customer growth significantly reduces the waiting time ($\hat{\beta} = -0.094, p < 0.01$). Numerically, a unit increase in *Customer* is estimated to reduce duration time by approximately 9%. In summary, Table 5 confirms the robustness of the indirect effects introduced by the customer mediators.

4.3.2. Organic Traffic versus Paid Traffic

Up to this point, the customer acquisition process has been tracked by analyzing the organic web traffic data provided by SEMrush. Although we controlled for trademark registration, which can

[11] This result is because the dependent variable is in logarithmic form while the explanatory variable is not. Comparing startups with prototypes to those without, it yields a difference of $1-e^{-0.036}$, which equates to a 3.5% decline.

serve as a proxy for a startup's marketing efforts, such as brand building (Krasnikov et al., 2009), the number of trademarks is still not directly linked to the company's customer acquisition activity. Young enterprises may use online advertising to enhance their website exposure (Goldfarb and Tucker, 2011). Considering this, the empirical results may suffer from an endogeneity issue, in which high monthly visits are driven by advertising rather than prototyping. To address this issue, we extend our dataset by incorporating paid traffic data obtained from SEMrush. Paid traffic refers to website visitors secured through online advertising, distinct from organic traffic, which counts visitors based on unpaid search engine results.

The variable $Customer_{paid}$ represents a momentum oscillator that measures the speed and change in monthly paid visits, akin to the approach employed for organic traffic data. We then estimate the proposed 2SLS specification by including this new variable, and present the results in Table 6. As can be seen in column (2), when controlling for paid traffic, we still find a positive and statistically significant association between prototyping and customer growth ($\hat{\beta} = 0.340, t = 3.39, p < 0.01$). In column (3), the mediation conclusions remain consistent, with customer growth still significantly decreasing the waiting time ($\hat{\beta} = -9.550, t = -9.16, p < 0.01$). We then perform a bootstrapping test with 500 replications to evaluate the robustness of the mediation effect $\widehat{\beta_{21}} \times \widehat{\beta_{32}}$, with paid traffic controlled. The results show a statistically significant mediation effect ($\hat{\beta} = -3.35, p < 0.01$), with zero excluded from the confidence interval ($C.I. = [-5.303, -1.440]$). In the specification of survival models, when controlling for customer growth, the effect of prototyping loses significance ($\hat{\beta} = -0.019, p = 0.24$), and customer growth demonstrates a reduction in waiting time ($\hat{\beta} = -0.094, p < 0.01$). Finally, paid traffic has a negative effect on fundraising time in both panels and a significant effect in the OLS panel, indicating that advertising efforts can also be a potential factor in enhancing fundraising efficiency.

	OLS	AFT Model	
	(1)	(2)	(3)
	$P \rightarrow C$	$P \rightarrow T$	$P, C \rightarrow T$
Prototype	0.333***	-0.036*	-0.017
	(0.099)	(0.018)	(0.016)
Customer			-0.094***
			(0.007)
Startup-side control variables			
Technology	-0.006	0.005*	0.004
	(0.013)	(0.005)	(0.003)
Age	-0.135***	-0.068***	-0.082***
	(0.025)	(0.005)	(0.004)
University	-0.010	0.001	0.000
	(0.007)	(0.002)	(0.002)
Founders	0.008	-0.010	-0.007
	(0.038)	(0.007)	(0.007)
Experience	0.137	-0.040**	-0.023
	(0.083)	(0.017)	(0.016)
Patents	-0.020	-0.001	-0.004
	(0.024)	(0.005)	(0.005)
Trademarks	0.039	-0.003	0.001
	(0.027)	(0.004)	(0.004)
Application field	Y	Y	Y
Geography	Y	Y	Y
VC-side control variables			
Syndication		0.003	0.002
		(0.003)	(0.003)
Distance		0.004	-0.003
		(0.006)	(0.005)
VC Type	N	Y	Y
Observations	396	396	396
AIC	1100	-220	-330

Notes: The pivotal variables are denoted by three abbreviations: *P (Prototype)*, *C (Customer)*, and *T (Timing)*. Robust standard errors are reported in parentheses. * p < 0.10, ** p < 0.05, *** p < 0.01.

Table 5. Survival Analysis

	OLS			AFT Model	
	(1)	(2)	(3)	(4)	(5)
	$P \rightarrow T$	$P \rightarrow C$	$P, C \rightarrow T$	$P \rightarrow T$	$P, C \rightarrow T$
Prototype	-4.753**	0.340***	-1.449	-0.039**	-0.019
	(1.828)	(0.100)	(1.555)	(0.018)	(0.016)
Customer			-9.550***		-0.094***
			(1.042)		(0.007)
$Customer_{paid}$	-2.954**	0.075	-2.164**	-0.019	-0.015
	(1.202)	(0.079)	(1.010)	(0.014)	(0.012)
Startup-side control variables	Y	Y	Y	Y	Y
VC-side control variables	Y	N	Y	Y	Y
Observations	396	396	396	396	396
AIC	3390	1088	3263	-220	-327

Notes: The pivotal variables are denoted by three abbreviations: *P (Prototype)*, *C (Customer)*, and *T (Timing)*. Robust standard errors are reported in parentheses. * p < 0.10, ** p < 0.05, *** p < 0.01.

Table 6. Robustness Check by Controlling Paid Traffic

5. Conclusion and Discussion

What is the right strategy for AI new ventures? Although startups have been one of the most important pillars for commercializing AI technology, the capital-constrained positions may impede their ability to fully realize their visions and subsequently appropriate the returns from their innovations. We explore the effects of prototyping on the potential customer acquisition process, which signals the feasibility and profitability of the investors' business hypotheses. To test the proposed mechanism, we collected a unique dataset of 396 AI startups in the U.S., their prototyping activities, and customer track records. We introduce a new approach for identifying prototypes using web scraping data and employ web traffic as a proxy for monitoring potential customer growth. We then conduct OLS regression and survival analysis to examine how prototyping benefits firms. The results suggest a positive and statistically significant association between prototyping and customer growth, which, in turn, mediates shorter fundraising times.

In addition, as a technologically intensive industry, AI startups often launch products and services with a strong focus on technology. Although a technology-led starting point demonstrates competitiveness, it inherently introduces uncertainty, especially for fledgling ventures. We propose a web-based indicator to measure a startup's technical sophistication before Series A. Startups that develop deep technology often face high levels of uncertainty, which can lead to extended fundraising periods. We provide quantitative evidence of how prototyping moderates the impact of such uncertainty on investor deliberation time.

The debate between intentional (Porter, 1980) and emergent strategy (Mintzberg, 1990) has long been central to the field of strategic management. However, our results, in line with Levinthal (2017), fall within a middle ground between a meticulously planned, goal-driven strategy and a more adaptive, learning-oriented approach. The findings imply that entrepreneurs can intentionally allocate resources to support the prototyping process, thereby aligning their vision with the market's genuine needs. Nevertheless, it transcends a single top-down strategy to shape the direction of development. Prototyping is not a one-time task, and does not always follow a linear trajectory. As AI products are fundamentally based on machine learning models, prototypes offer an effective way to validate the accuracy of the underlying models. Most startups retain their rights to the secondary use of customer data, enabling them to iteratively gather and analyze customer feedback (Rahnama and Pentland, 2022). This iterative learning process facilitates continuous product and service improvements, ultimately leading to competitive advantage.

Besides, scholars of strategy and innovation have long been interested in the isolating mechanisms that immobilize the involuntary knowledge transfer of a focal innovation beyond firm boundaries (Rumelt, 1984; Teece, 1986; Liebeskind, 1996; Jonsson and Regner, 2009), with major mechanisms encompassing patents, secrecy, lead time, and complementary assets (James et al., 2013). Sharapov and Macaulay (2022) advance this research stream by conceptualizing how design, as a specific form of knowledge manifestation, inhibits potential imitators' behaviors. In this context, prototyping emerges as an innovation strategy for AI startups to selectively disclose valuable knowledge to the market and showcase functionality, while maintaining a level of confidentiality regarding the underlying algorithms. The iterative nature of prototyping further increases the imitation costs for the counterparty.

Most innovation and economic studies on AI have been focused on its social impact (e.g., job replacement) and the nature of the AI technology per se (e.g., AI as a general-purpose technology [GPT] or the invention of a method of the invention [IMI]) (Huang and Rust, 2018; Agrawal et al., 2019; Cockburn et al., 2019), without shedding much light on its commercialization, which is a critical part of the innovation generation process (Utterback, 1971; Rogers, 1995). In this study, we combine multiple web data sources to investigate the commercialization of AI from the lens of startup innovation and entrepreneurial activities. In this light, our work also informs research agendas regarding the use of web data to investigate firms' innovation activities. In innovation studies, patent and publication data have been used for decades to measure science and technological advancement (Trajtenberg et al., 1995; Nagaoka et al., 2010). When analyzing the commercialization stage of technology, while the claims of a patent may reflect the commercial use of the protected inventions (Merges, 1988), it is worth noting that patent data primarily tend to be accessible to established entities. In this sense, this study complements existing literature by demonstrating the potential of web data to study the innovation activities and commercialization of new technology in small enterprises.

This study has some limitations that may inspire future research. First, we include several control variables from both the startup and VC sides to reduce the bias of the estimated results. However, other factors, such as accelerator programs, may also influence the development of startups. Besides, this study relies on longitudinal data and does not capture the dynamics of startup activities. However, startups' business hypotheses and capabilities may change over time. To this end, we encourage future studies to use panel data to capture these dynamics.

Appendix

Appendix A

Variables	Operationalization (case-insensitive)
Prototype	platform, software, SaaS, interface, API, demo, bot
University	university, college, institute, academ*
Technology sophistication	artificial intelligen*, computational intelligen*, neural network, bayesian network, data mining, decision model, deep learning, genetic algorithm*, inductive logic programm*, machine learning, natural language generation, natural language processing, reinforcement learning, supervised learning, supervised training, swarm intelligen*, unsupervised learning, unsupervised training, semi supervised learning, semi supervised, training, connectionis*, expert system, fuzzy logic, transfer learning, learning algorithm, learning model, support vector machine, random forest, decision tree, gradient tree boosting, xgboost, adaboost, rankboost, logistic regression, stochastic gradient descent, multi-layer perceptron, latent semantic analysis, latent dirichlet allocation, multi-agent system, hidden markov model, clustering, computational creativity, descriptive model, inductive reasoning, overfitting, predictive analytics, predictive model, target function, backpropagation, self learning, objective function, feature selection, active learning, regression model, stochastic approach, stochastic technique, stochastic method, stochastic algorithm, probabilistic approach, probabilistic technique, probabilistic method, probabilistic algorithm*, recommendation system*, text analy*, text recognition, speech analy*, speech recognition, hand writing analy*, handwriting recognition, facial analy*, facial recognition, face analy*, face recognition, character analy*, character recognition

Notes: The keywords for capturing university linkage are provided by Li et al. (2018), who utilizes web data to examine the effects of the Triple Helix model (university-government-industry) on enterprise growth. The keywords used for measuring a startup's technological sophistication are derived from the lists of AI-related technical keywords presented in the WIPO AI Technology Report (Methodology Part: https://www.wipo.int/tech_trends/en/artificial_intelligence/). The variables *University* and *Technology sophistication* are constructed as frequency-based indicators, whereas the variable *Prototype* is based on the occurrence of listed keywords on historical websites.

Appendix B

Crunchbase provides location information for different entities, which is represented in the format of "city, state, country". The distance between an entrepreneurial firm and the lead investor is determined by the following formula:

$$Distance = w_1 I_{city} + w_2 I_{state} * + w_3 I_{country}$$

where I_{city} is a city-level indicator function. I_{city} to 0 if two entities located in the same city, and 1 otherwise. The weights for w_1, w_2, w_3 are set to 1, 2, and 2, respectively. Two examples are provided in Table A.1.

Entity	City	State	Country	Distance
Startup A	Austin	Texas	United States	3
VC G	Cambridge	Massachusetts	United States	
Startup M	Los Angeles	California	United States	0
VC T	Los Angeles	California	United States	

Table A.1. Measuring the distance between startups and investors

References

Admati A, Pfleiderer S (1994) Robust financial contracting and the role of venture capitalists. Journal of Finance. 49(2):371–402.

Aggarwal R, Kryscynski D, Singh H (2015) Evaluating Venture Technical Competence in Venture Capitalist Investment Decisions. Management Sci. 61(11): 2549-2824.

Agrawal A, Gans J, Goldfarb A (2019) Economic policy for artificial intelligence. Innov. Policy Econ. 19(1):139–159.

Audretsch, DB, Bönte W, Mahagaonkar P (2012) Financial signaling by innovative nascent ventures: The relevance of patents and prototypes. Res. Policy 41, 1407–21.

Baron RM, Kenny DA (1986) The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. J. Personality Soc. Psych.51(6):1173–1182.

Bessen J, Stephen MI, Lydia R, Robert S (2022) The role of data for ai startup growth. Res. Policy 51, 104513.

Blank S (2013) Why the lean startup changes everything. Har.Bus. Rev.

Blank S, Eckhardt JT (2023) The Lean Startup as an Actionable Theory of Entrepreneurship. Journal of Management.

Bradburn MJ, Clark TG, Love SB, Altman DG (2003) Survival analysis part II: Multivariate data analysis—an introduction to concepts and methods. British J. Cancer. 89:431–436.

Brynjolfsson E, Rock D, Syverson C (2021) The productivity J-curve: How intangibles complement general purpose technologies. American Economic Journal: Macroeconomics. 13(1):33.

Carlile PR (2002) A pragmatic view of knowledge and boundaries: Boundary objects in new product development. Organ. Sci. 13(4), 442–455.

Carlile PR (2004) Transferring, translating, and transforming: An integrative framework for managing knowledge across boundaries. Organ. Sci. 15(5), 555–568.

Chesbrough H (2010) Business model innovation: Opportunities and barriers. Long Range Planning. 43: 354-363.

Cockburn IM, Henderson R, Stern S (2018) The impact of artificial intelligence on innovation. NBER Working Paper No. 24449, National Bureau of Economic Research Cambridge, MA.

Colombo MG, Delmastro M, Grilli L (2004) Entrepreneurs' human capital and the start-up size of new technology-based firms. International Journal of Industrial Organization. 22:1183-1211.

Colombo MG, Grilli L (2005) Founders' human capital and the growth of new technology-based firms: A competence-based view. Res. Policy 34(6):795–816.

Cumming D, Johan S (2017) The Problems with and Promise of Entrepreneurial Finance. Strategic Entrepreneurship J. 11(3): 357-370.

Dalle JM, den Besten M, Menon C (2017) Using CrunchBase for economic and managerial research, Working paper, OECD Science, Technology and Industry.

Davila A, Foster G, Gupta M (2003) Venture capital financing and the growth of startup firms. J. Bus. 18(6):689–708

Duc AN, Abrahamsson P (2016) Minimum Viable Product or Multiple Facet Product? The Role of MVP in Software Startups.

Eisenhardt KM, Schoonhoven CB (1990) Organizational growth: linking founding team, strategy, environment, and growth among U.S. semiconductor ventures, 1978-1988. Admin. Sci. Quart. 35:504–529.
Furman J, Seamans R (2019) AI and the economy. Innovation Policy and the Economy. 19(1):161–191.
Gans JS, Stern S, Wu J (2019) Foundations of entrepreneurial strategy. Strategic Management J. 40(5):736–756.
Goldfarb A, Tucker CE (2011) Privacy Regulation and Online Advertising. Management Sci. 57(1)
Gompers P, Lerner J (2001) The Money of Invention: How Venture Capital Creates New Wealth. Harvard Business School Press, Boston, MA.
Gompers PA (1995) Optimal investment monitoring, and the stages of venture capital. J. Finance. 50:1461–1489.
Gorman M, Sahlman WA (1989) What do venture capitalists do? J. Bus. Venturing 4231–248.
Gouthier M, Schmid S (2003) Customers and Customer Relationships in Service Firms: The Perspective of the Resource-Based. Marketing Theory 3(1).
Graham SJ, Hancock G, Marco AC, Myers AF (2013) The USPTO trademark casefiles data set: Descriptions, lessons, and insights. J. Econom. Management Strategy 22:669–705.
Grichnik D, Dew N, Read S, Sirén, C (2016). The role of boundary objects in entrepreneurship: Moving towards an Interaction-Learning nexus. Academy of Management Proceedings, 2016(1), 13681–1368.
Groening C, Mittal V, Zhang YA (2016) Cross-Validation of Customer and Employee Signals and Firm Valuation. Journal of Marketing Research 53(1)
Guzman J, Li A (2023) Measuring Founding Strategy. Management Sci. 69(1)
Hannigan TR, Briggs AR, Valadao R, Seidel MDL, Jennings PD (2022) A new tool for policymakers: mapping cultural possibilities in an emerging AI entrepreneurial ecosystem. Res. Policy. 51:104315.
Hsu D (2004) What do entrepreneurs pay for venture capital affiliation? J. Finance 59(4):1805–1844.
Hsu D (2007) Experienced entrepreneurial founders, organizational capital, and venture capital funding. Res. Policy 36(5):722–741.
Hsu D, Ziedonis R, (2008) Patents as Quality Signals for Entrepreneurial Ventures. Academy of Management Best Paper Proceedings.
Huang M-H, Rust RT (2018) Artificial intelligence in service. J. Service Res. 21(2):155–172.
Jonsson S, Regner P (2009) Normative barriers to imitation: Social complexity of core competences in a mutual fund industry. Strategic Management J. 30: 517–536.
James SD, Leiblein MJ, Lu S (2013) How firms capture value fromtheir innovations. Journal of Management, 39: 1123–1155.
Kerr WR, Lerner J, Schoar A (2014) The Consequences of Entrepreneurial Finance: Evidence from Angel Financings. The Review of Financial Studies 27(1):20-55.
Kerr WR, Nanda R, Rhodes M (2014) Entrepreneurship as Experimentation. Journal of Economic Perspectives 28(3): 25-48.
Kirsch D, Goldfarb B, Gera A (2009) Form or substance: The role of business plans in venture capital decision making. Strategic Management J. 30(5):487–515.
Koning R, Hasan S, Chatterji A (2022) Experimentation and Start-up Performance: Evidence from A/B Testing. Management Sci. 68(9)
Kossuth J, Seamans R (2018) The competitive landscape of AI startups. Harvard Bus. Rev.

Krasnikov A, Mishra S, Orozco D (2009) Evaluating the Financial Impact of Branding Using Trademarks: A Framework and Empirical Evidence. Journal of Marketing 73(6)

Liebeskind, JP (1996) Knowledge, strategy, and the theory of the firm. Strategic Management J. 17: 93–107.

Leach JS, Melicher RS (2010) Entrepreneurial Finance (Fourth Edition). USA: South-Western centage learning.

Lerner J (1994) Venture capitalists and the decision to go public. Journal of Financial Economics 35(3):293-316.

Levinthal DA (2017) Mendel in the C-suite: Design and the evolution of strategies. Strategy Sci. 2(4):282–287.

Li Y, Arora S, Youtie J, Shapira P (2018) Using web mining to explore Triple Helix influences on growth in small and mid-size firms. Technovation 76-77:3-14.

Long C (2002) Patent signals. The University of Chicago Law Review 69 (2), 625–679.

Lutchen KR (2018) Why companies and universities should forge long-term collaborations. Harvard Bus. Rev.

Mann RJ, Sager TW. Patents, venture capital, and software start-ups. Res. Policy (2007) 36(2):193–208.

Markman GD, Siegel DS, Wright M, (2008) Research and technology commercialization. Journal of Management Studies 45, 1401–23.

Merges RP (1988) Commercial success and patent standards: Economic perspectives on innovation. California Law Review 76:803–876.

Mintzberg H (1990) The design school: Reconsidering the basic premises of strategic management. Strategic Management J. 11(3):171–195.

Nagaoka S, Motohashi K, Goto A (2010) Patents statistics as an innovation indicator. In: Hall, B.H., Rosemberg, N. (Eds.), Handbook of Economics of Innovation. 2:1083–1127.

Nelson RR, Winter SG (1977) In search of useful theory of innovation. Res. Policy 6(1): 36-76.

Nonaka I (1994) A dynamic theory of organizational knowledge creation. Organ. Sci. 5 14-37.

Park SH, Chen R, Gallagher S (2017) Firm Resources as Moderators of the Relationship Between Market Growth and Strategic Alliances in Semiconductor Start-UPS. Academy of Management Journal 45(3).

Pindyck RS (1993) Investments of uncertain cost. Journal of Financial Economics 34(1): 53-76.

Porter ME (1980) Competitive Strategy: Techniques for Analyzing Industries and Competitors (Free Press, New York).

Rahnama H, Pentland AS (2022) The new rules of data privacy. Harvard Bus. Rev.

Ries E (2011). The lean startup: How today's entrepreneurs use continuous innovation to create radically successful businesses. Crown Books.

Rogers EM (1983) The Diffusion of Innovation. FreePress, New York.

Rumelt, RP (1984) Towards a strategic theory of the firm. In R. B. Lamb (Ed.), Competitive strategic management: 556–570. Englewood Cliffs, NJ: Prentice-Hall.

Rosch E (1975) Cognitive representation of semantic categories. Journal of Experimental Psychology. 104(3): 192 – 233.

Rosenberg N (1994) Economic Experiments. In Chap. 5 in Exploring the Black Box: Technology, Economics, and History. Cambridge University Press.

Rosenberg N (1994) Exploring the black box: Technology, economics, and history. Cambridge, England: Cambridge University Press.

Sapienza HJ, Gupta AK (1994) Impact of Agency Risks and Task Uncertainty on Venture Capitalist-CEO Interaction. Academy of Management Journal 37(6):1618-1632.

Schoonhoven CB, Eisenhardt KM, Lyman K (1990). Speeding products to market: waiting time to first product introduction in new firms. Admin. Sci. Quart. 35(1): 177-207.

Schwienbacher A (2007). A theoretical analysis of optimal financing strategies for different types of capital-constrained entrepreneurs. J. Bus. Venturing 22: 753-781.

Shane S, Stuart T (2002) Organizational Endowments and the Performance of University Start-ups. Management Sci. 48(1)

Shepherd DA, Gruber M (2021). The Lean Startup Framework: Closing the Academic–Practitioner Divide. Entrepreneurship Theory and Practice 45(5)

Shoham Y, Perrault R, Brynjolfsson E, Clark J, Manyika J, Niebles JC, Bauer Z (2018). The AI index 2018 annual report. AI Index Steering Committee, Human-Centered AI Initiative. Stanford University, Stanford, CA.

Sobel ME (1982) Asymptotic confidence intervals for indirect effects in structural equation models. Sociol. Methodology 13(1):290–312.

Spence M (1973) Job market signaling. Quarterly Journal of Economics.

Star SL (1989) The structure of ill-structured solutions: Boundary objects and heterogeneous distributed problem solving. M. Huhns, L. Gasser, eds. Readings in Distributed Artificial Intelligence. Morgan Kaufman, Menlo Park, CA.

Stuart T, Huang H, Hybels R (1999) Interorganizational endorse-ments and the performance of entrepreneurial ventures. Admin. Sci. Quart. 44:315–349.

Sharapov D, Macaulay SC (2020) Design as an isolating mechanism for capturing value from innovation: from cloaks and traps to sabotage. Acad. Manag. Rev.

Teece DJ (1986) Profiting from technological innovation: Implications for integration, collaboration, licensing and public policy. Research Policy, 15: 285–305.

Teece DJ (2010) Business models, business strategy, and innovation. Long Range Planning 43(2–3):172–194.

Terwiesch C, Loch CH (2004) Collaborative Prototyping and the Pricing of Custom-Designed Products. Management Sci. 50(2): 133-279.

Thomke SH (1998) Managing Experimentation in the Design of New Products. Management Sci. 44(6): 743-877.

Tian X (2011) The causes and consequences of venture capital stage financing. Journal of Financial Economics 101(1):132-159.

Tratjenberg M, Henderson R, Jaffe A (1997) University versus corporate patents: A window on the basicness of invention. Econom. Innovation New Tech. 5:19–50

Tyebjee TT, Bruno AV (1984) A model of venture capitalist investment activity. Management Sci. 30, 1031–1042.

Utterback JM (1971) The process of technological innovation within the firm. Academy of Management Journal. 10:75-88.

von Hippel E (1986) Lead users: A source of novel product concepts. Management Sci. 32: 791-805.

Weber M, Beutter M, Weking, J, Bohm M, Krcmar H (2022) AI Startup Business Models. Bus. Inf. Syst. Eng. 64(1):91–109.

Weiblen T, Chesbrough HW (2015) Engaging with start-ups to enhance corporate innovation. Calif. Manage. Rev. 57(2):66–90.

Woolston C (2022) Is big tech draining AI talent from academia? Nature 610, S26-S27.

Zaltman G, Duncan R, Holbeck J (1973) Innovations and organizations. New York: Wiley

Zhang D, Maslej N, Brynjolfsson E, Etchemendy J, Lyons T, Manyika J, Ngo H, Niebles JC, Sellitto M, Sakhaee E, Shoham Y, Clark J, Perrault R (2022). The AI index 2022 annual report. AI Index Steering Committee, Human-Centered AI Initiative. Stanford University, Stanford, CA.