The effect of Artificial Intelligence experts on Venture Capital funding

DSMA MSc Thesis
July 9th, 2023
J, Kokkelkorn (599905)

Avagyan, Vardan – Thesis Coach

Preface

Abstract

Research on new ventures has shown that AI is becoming increasingly crucial for their development. However, implementing AI in companies is challenging, and resource shortages are frequently cited as reasons for startup failures. Drawing on AI expert characteristics and industry legitimacy theory, this study explores the role of AI experts in securing venture capital for startups. Employing a two-stage selection analysis to account for potential selection bias and utilizing an XGBoosting predictive model, a comprehensive dataset of 412 new ventures is analyzed. Bearing in mind that this research is an initial exploratory attempt to illuminate the role of AI in new venture funding using secondary data, the following insights are revealed: The first Heckman model, focusing on funding amounts, indicates a negative impact of pre-seed firms and a positive influence of AI expert education on funding. The second Heckman model, exploring the days until funding, demonstrates that earlier years in the sample experienced longer durations for securing first-round funding compared to 2023, and AI expert startup experience prolongs the funding process. The XGBoosting models showcase strong predictive capabilities for both total funding amounts and days until funding. Additionally, these predictive models identify the most crucial variables contributing to predicting the outcome variables, thereby offering valuable insights into VC decision-making processes.

Table of Contents

1. Introduction

New enterprises frequently turn to venture capital (VC) companies for funding because they typically lack the funds necessary to launch and grow their companies. In the early stages of the organizational life cycle, professionally managed capital that has been pooled in funds and invested in privately held enterprises is referred to as venture capital (Gorman & Sahlman, 1989). From $347 billion spread over 31,623 deals in 2020 to a record $671 billion spread across 38,644 deals in 2021, global VC investment increased. However, only a small percentage of new businesses are successful in completing this critical phase of their organizational development (KPMG, 2021). Around 1 out of every 100 new ventures receives finance, according to Kirsch et al. (2009).

Several factors that venture capitalists consider when selecting whether to invest in a new venture have been uncovered by research in entrepreneurship, finance, and marketing. In addition to these systemized funding criteria, academic researchers have found evidence that venture capitalists also consider artificial intelligence (AI) to be important for new venture success (Mou, 2019). The competition to buy AI startups and technologies is getting more intense, according to Mou (2019). Large businesses and VCs are attempting to incorporate machine learning into their product lines and investment portfolios. Perhaps investments in AI follow machine learning. Unfortunately, it is quite challenging to implement AI in startups efficiently (Hulme, 2022). The latter subsequently causes a high rate of failure for new businesses. Startup failures have been heavily linked to a lack of human and financial resources. However, AI research is quite surprisingly silent on how AI-related aspects may increase new ventures' funding and timing of funding.

The current study fills this knowledge gap by examining the function of AI experts in securing venture capital. If venture capitalists recognize the significance of AI-related factors in their financing decisions, the question then arises as to whether AI expert factors, who serve as the face of AI in new ventures, have an impact on funding in addition to traditional funding criteria. To develop the conceptual model, I draw an organizational legitimacy theory based on that from Rao et al. (2008) and Homburg et al. (2014). I will argue that specific AI expert traits may improve the new venture's AI credibility, increasing funding amount and decreasing time until funding. More specifically, I aim to predict that a new business with an AI expert who has more work experience, and a valuable educational background will be able to meet the cognitive expectations of investors for the right artificial intelligence skills. Furthermore, I contend that variables that indicate the level of industry legitimacy of the venture and the work environment of the AI expert attenuate this association. Therefore, the research question is as follows, *'How do AI experts' characteristics affect VC funding?'*

To test the hypotheses, a data set from multiple sources that contain longitudinal data on 412 new ventures was compiled. I analyze the data using a two-step Heckman continuous regression model and eXtreme Gradient Boosting. The two-step Heckman selection model enables us to analyze the relationship between AI expert characteristics, the timing of funding and the amount of funding while accounting for self-selection bias that may result from the decision to establish an AI expert in the startup team. eXtreme Gradient Boosting allows us to predict the days until funding and how much every startup gets funded and enables us to gain insights into variable importance. By examining the scores of variable importance, we can understand which features have the most significant impact on predicting the funding amount and the days until funding. This knowledge is valuable for feature selection, understanding the underlying dynamics, and making informed decisions.

The first Heckman model, focusing on funding amounts, reveals a negative impact of pre-seed firms and a positive influence of AI expert education on funding. The second Heckman model, exploring the days until funding, demonstrates that earlier years in the sample experienced longer durations for securing first-round funding compared to 2023, and that AI expert startup experience prolongs the funding process.

Eventually, this study can be regarded as the first study of the AI expert's function in new venture investment. The results of this study therefore warrant additional research, especially considering the underlying mechanisms. Given this restriction, the results offer three primary contributions to the body of literature. First, previous research has mostly concentrated on the opinions of investors about newly launched businesses that provide public stock at launch or are already listed on a stock market. The focus of this study, however, is on much more recent endeavors and draws on the work of Homburg et al. (2014). I build on their research by looking into venture capitalists, a group of investors that is not currently on the research agenda for the intersection of artificial intelligence, entrepreneurship and finance. A fundamental difference between the investment decisions of venture capitalists and IPO stage investors is that the latter can observe prior VC investment as “a reliable measure of the success the firm has had in the past in the securing financial capital, and so is and indicator of the firm’s potential for growth as well” according to Higgins and Gulati (2006). So, venture capitalists face a decision-making position that offers very limited information about a new venture's quality, in contrast to IPO-stage investors. The findings of this study offer guidance on how entrepreneurs should set up their startup team to meet the expectations of venture capitalists more closely, hence improving their chances of receiving more and faster funding. Second, this study theoretically contributes to the expanding body of work on AI for business (Lee et al., 2019; Gungor, 2020). The research has looked at alliances, absorptive capacity, and flow signaling in new businesses. To offer a theoretical foundation for why the AI expert might also be important to potential investors, the current study draws on an organizational legitimacy theory. As a result, this research offers new insights into the complex relationships between a new venture's AI strategy and its chances of receiving more and faster finance. Finally, this research is the first to examine AI experts for new businesses. This area of study expands on earlier studies that looked at how CMOs and CFOs affect new venture funding (Homburg et al., 2014). High managerial discretion, which refers to the level of influence of experts on a firm's activities and success, is a characteristic of young enterprises. As a result, new businesses offer a distinctive setting for considering AI expert-related activities. This provides a suitable opportunity to evaluate AI experts' impact, since this has not been studied before.

Management could benefit greatly from research on whether adding an AI expert to a startup might enhance venture capital investment. Managers may find it useful in a number of ways to look into how having an AI expert affects a startup's ability to obtain more and faster funding from venture capitalists. The first benefit is that it can shed light on the priorities, tastes, and perspectives of venture capitalists, as well as how they see the role of artificial intelligence in a firm. Second, it can help the management decide whether to bring on an AI expert at an early stage of the startup's growth. Finally, it can assist managers in learning how to persuade potential investors of the value of artificial intelligence and machine learning during fundraising.

In addition, there are numerous academic reasons to investigate if hiring an AI expert leads to successful startups. First off, it can add to the growing body of knowledge regarding the function of artificial intelligence and machine learning in startups and their effects on many business facets. Second, it can shed light on how venture investors' priorities and preferences are changing as well as how they see AI's place in startups. Thirdly, it can shed light on how entrepreneurs can persuasively explain the value of AI to potential investors during fundraising and how investors perceive AI.

The rest of this study is structured as follows: Section 2 will provide an extensive literature review of related studies. Section 3 will present this study's methodology and research design. Sections 4 will summarize the results. Lastly, section 5 will conclude this study and discuss the research contribution and limitations.

2. Literature Review and Hypothesis Development

Enterprises often seek funding from venture capital companies, which pool professionally managed capital for investment in privately held enterprises. Even though there has been an increase in global VC investment, few startup companies actually acquire funding. While deciding whether to invest, VC firms take into account a variety of elements, including AI, which is becoming more crucial for the development of new ventures. Yet, implementing AI in companies is difficult, and resource shortages are frequently cited as reasons why startups fail. This study aims to determine how AI-related aspects may raise funding and acquire funding faster. The next section looks at the empirical relationship on how AI experts could increase funds and decrease time until funding as a member of the startup team.

Prior entrepreneurship research has discovered a number of variables that venture capitalists use to determine whether to invest in a certain new firm. These factors are divided by Zacharakis, McMullen, and Sheperd (2007) into those that pertain to human capital and market characteristics. For instance, whereas human capital relates to experience in a particular position, experience in the relevant industry, and startup experience, market parameters pertain to the demand for the product, the capacity to protect intellectual property, and the competitive nature of the business. Using the aforementioned criteria, earlier research appears to be more precise. The criteria were divided into team characteristics, product characteristics, market characteristics, and financial aspects by Zacharakis and Meyer (1998). According to their study, team characteristics include experience, education, and personality; product characteristics include concrete attributes and the availability of prototypes; market characteristics include product demand, whether the new venture creates a new market, and competitive industry; and, finally, financial characteristics include liquidity and return on investment. The overall objective of this study is to determine whether certain AI expert characteristics have an impact on VC investment. Therefore, all previously established funding criteria that are applicable and readily available in this specific study environment are included in this study.

2.1 Legitimacy of New Ventures

According to Suchman (1995), the concept of legitimacy is a broad perception or presumption that an entity's acts are preferable, right, or appropriate within some socially formed system of norms, values, beliefs, and definitions. Cognitive and sociopolitical legitimacy are two different types of organizational legitimacy, according to Aldrich (1999). I will concentrate on the first one, which is the widespread belief that a new entity is appropriate. Because they are brand-new and unfamiliar entities, new ventures by definition lack cognitive legitimacy. Yet, as legitimacy is a requirement for obtaining funding, business owners must persuade investors of the appropriateness of their organizations (Zimmerman and Zeitz, 2002).

According to earlier articles, VCs are presently very interested in startups using AI. According to Forbes (2023), since 2020, there has been a 425% rise in overall investment in AI-related businesses. This demonstrates that because they view artificial intelligence as a key predictor of new venture success, venture capitalists pay particular attention to AI-related issues. The article goes on to say that understanding artificial intelligence may help prevent failed new business ventures. In addition to improving overall products, AI may help companies make better business decisions by helping them analyze market trends and client preferences. This is easier said than done, though. Therefore, establishing artificial intelligence legitimacy in the eyes of potential investors thus constitutes a central challenge for new ventures.

According to Zimmerman and Zeitz (2002), the seminal works on legitimacy in new venture contexts view it as a quality signal because other indications, particularly economic ones, are not available in a new venture context. As new businesses follow socially built norms of appropriate organizational features, this legitimacy will gradually grow (Bruton et al., 2010). Also, adhering to established standards and guidelines will eventually improve startup performance. For instance, Zimmerman and Zeitz (2002) contend that legitimacy aids in determining the caliber of a new enterprise in the absence of any preceding market-based performance benchmarks.

In the next section the current study focusses on the centrality of the AI expert for establishing artificial intelligence legitimacy by discussing his or her main roles. This shows that VC's consider AI as very promising and exciting.

2.2 Contribution of AI experts to New Venture Legitimacy

AI expert are professionals with expertise in artificial intelligence (AI) and related fields. They possess specialized knowledge and skills in developing and implementing AI technologies and solutions. In a startup setting, AI experts can play various roles depending on their specific expertise and the needs of the company. The most common roles are: Chief AI Officers (CAIOs), Data Scientists, AI Researchers, Machine Learning Engineers and Natural Language Processing experts. To narrow this research down, the role of AI experts will be specified further.

By investigating AI experts, I uncover valuable insights about AI strategy, decision-making processes, industry-specific challenges, and the interplay between AI and business outcomes in startup settings. This research can contribute to a deeper understanding of the role of AI leadership and inform best practices for AI implementation and management in startups.

AI Experts are highly skilled professionals who possess specialized knowledge and expertise in the field of artificial intelligence (Galanos, 2018). They play a crucial role in the development, implementation, and advancement of AI technologies within organizations. AI experts are typically responsible for a range of tasks that contribute to the successful application of AI in various domains.

First and foremost, AI experts are involved in research and development activities (Galanos, 2018). They stay up-to-date with the latest advancements and emerging trends in AI and actively contribute to the development of cutting-edge AI algorithms, models, and techniques. Through their deep understanding of machine learning, natural language processing, computer vision, and other AI disciplines, they strive to push the boundaries of what AI can achieve.

AI professionals also play a key role in the data analysis and preprocessing. In order to obtain, clean, and prepare datasets for training AI models, they collaborate closely with data engineers. This entails determining pertinent aspects, addressing missing data, and guaranteeing the data is unbiased and representative. AI experts have the knowledge and abilities to use the proper feature engineering techniques and data preparation procedures, which are essential for developing reliable and accurate AI models. Another important responsibility of AI experts is model development and optimization. They are skilled in a variety of AI frameworks and libraries and have knowledge in creating, developing, and optimizing AI models. AI experts are knowledgeable about the nuances of many algorithms and architectures and may choose the ones that are best suited for a certain task. To attain high accuracy and generalization, they enhance model performance using strategies including hyperparameter tuning, regularization, and ensembling. Ethics and responsibility in AI are also areas where AI experts play a significant role. They are conscious of the moral issues and potential biases related to AI systems. AI experts strive to create and put into use models and systems that are just, transparent, and responsible. They actively participate in debates on AI ethics, privacy, and security and work to reduce any dangers and negative societal effects. Furthermore, AI experts frequently contribute to collaboration and knowledge sharing inside their firms and the larger AI community. To promote innovation and the interchange of ideas, they can publish research papers, take part in conferences, and work on open-source projects. Additionally, other team members receive assistance and mentoring from AI experts, which helps to develop talent and promote continual learning and progress. Overall, AI experts are instrumental in advancing the field of artificial intelligence and driving its practical application in various industries. Their knowledge, research efforts, and problem-solving abilities help to create cutting-edge AI technology and give businesses access to the power of AI for better innovation, automation, and decision-making (Galanos, 2018).

The decisional role within the field of AI is accounted for by the mere presence of an AI expert. This is a quality that venture capitalists can easily spot because their position in the top management team serves as a measure of both the corporate status of artificial intelligence and the corporate adoption of the AI concept. For the informational and relational roles, however, the AI experts of a new venture does not have a considerable track record in that venture that may indicate her or his capability of fulfilling these roles. Hence, sufficient amounts of previously acquired human and social resources that are considered advantageous for CAIOs' role-specific activities may be used to legitimize AI.

Knowledge, reputation, and business contacts are the most important social and human resources. These materials are inaccessible to outside observers. Information on the formal education and work experience of the top management team, which are the main ways to obtain the relevant resources, is indicative of the level of such resources. As a result, we emphasize AI expert education and experience as reliable and obvious signs of a new venture's AI legitimacy. In particular, we outline how education and experience help AI experts with their relational and informational jobs by providing them with resources like contacts, knowledge, and reputation.

2.3 AI Expert Education

AI expert education is the term used to describe the level of the AI expert's education, which may be determined by whether they attended a prestigious university. According to Homburg et al. (2014), formal education increases experts' capacity to carry out informational and relational tasks since it gives them access to information, contacts, and reputational resources.

First, education might help AI experts with their informative work by supplying them with contacts and knowledge. According to Scott (1994), the expert acquires specialized, explicit, and codified information through formal education. According to Zimmerman and Zeitz (2002), a diploma from a prestigious business school proves that the company is aware about the best management practices. As a result, formal education may suggest that the AI expert is knowledgeable about cutting-edge information gathering technologies and is able to codify the information acquired to make it relevant for all of the aforementioned AI expert roles. Second, education may facilitate AI experts' relational tasks by endowing them with reputation and contracts. Finkelstein (1992) asserts that education generates reputational resources since enrolment at particular institutions is associated with importance among the corporate elite. Having this accreditation makes it easier for AI experts to endorse partnerships with outside parties because their functional credentials are more widely accepted. Moreover, education improves relational duties by fostering social connections through networks or school contacts that are beneficial for fostering partnerships (Brush et al., 2001). This results in the first hypothesis:

***H1**: AI experts' education increases VC funding and accelerates time until funding.*

2.4 AI Expert Experience

According to Lam (2000), experience, as opposed to education, produces tacit knowledge, which is action-oriented, challenging to define, and focused on routines and operational abilities. As a result, AI experts' experience refers to the depth of their professional experience and endows them with social and human resources, such as contacts and knowledge (Hitt et al., 2001).

When assessing someone's experience, prior research has concentrated on two categories of work experience: role experience and work experience. Role experience is experience related to a particular function, such as an AI expert, whereas firm experience is the quantity of work experience within that firm. Although company experience is largely useless in a newly founded firm, only role experience appears helpful in this study. In addition, this research expands on Boyd, Chandy, and Cunha's (2010) work by further classifying job experience into artificial intelligence, industrial, and startup experience. Homburg et al. (2014) makes the assumption that contacts and knowledge gained through these various sorts of experience may signify an AI expert's capacity to carry out relational and informational duties. These encounters will be described in detail in the following sections.

2.4.1 AI Expert Artificial Intelligence Experience

AI experts' artificial intelligence experience is the depth of professional experience in AI-related tasks that has led to increased expertise in the field of AI. In addition to these knowledge resources, AI experts with AI experience have connections to the AI community that make it possible to quickly identify new trends and techniques. These connections would help AI experts do relational tasks more effectively since they may include useful service providers, such firms that develop AI. The following will be the second hypotheses:

***H2**: AI experts' AI experience increases VC funding and accelerates time until funding.*

2.4.2 AI Expert Industry Experience

AI expert industry experience refers to the depth of professional experience in the particular industry environment of the new enterprise. For instance, Colombo and Grilli (2005) note that the new company can benefit from managers' familiarity with technologies, customer needs, competitor strengths, and weaknesses, as well as from their contacts with possible clients and suppliers from their prior employment. So, for AI experts' tasks, tacit knowledge gained via industry experience can be crucial. Moreover, AI experts' relational tasks with AI researchers, vendors, and suppliers can be facilitated by already-existing contacts to lead customers and suppliers in the sector. Hence, the third theory will be:

***H3**: AI experts' industry experience increases VC funding and accelerates time until funding.*

2.4.3 AI Expert Startup Experience

AI experts' startup experience refers to the level of work experience in new venture contexts. More precisely, startup experience demonstrates an understanding of the particular difficulties and limitations that entrepreneurs face as well as the capacity to deal with high uncertainty and make quick decisions in small and young firms (Delmar and Shane, 2004). Experience with startups may be especially crucial for AI experts since those with more startup knowledge may be better able to identify the information that is most relevant for completing AI-related activities in a business context. In addition, AI experts with startup expertise may have a better understanding on how to face the common startup issues. These difficulties may be related to crucial contacts for expanding the startup, AI service suppliers, and possible investors. Hence, the fourth hypothesis will be:

H4*: AI experts' startup experience increases VC funding and accelerates time until funding.*

2.5 Environmental Moderator

In addition to the previously mentioned factors, research has suggested that the degree of uncertainty in the organizational environment affects how the top management team affects the future of the company (Carpenter and Fredrickson, 2001). Zimmerman and Zeitz's (2002) industry legitimacy theory emphasizes uncertainty as a significant moderator variable. The relationship between AI experts' education, experience, and funding is therefore predicated on the degree of environmental uncertainty, according to this article. Homburg et al. (2014) takes into account the venture's institutional environment. The part that follows will provide more information on this idea.

2.5.1 Industry Legitimacy

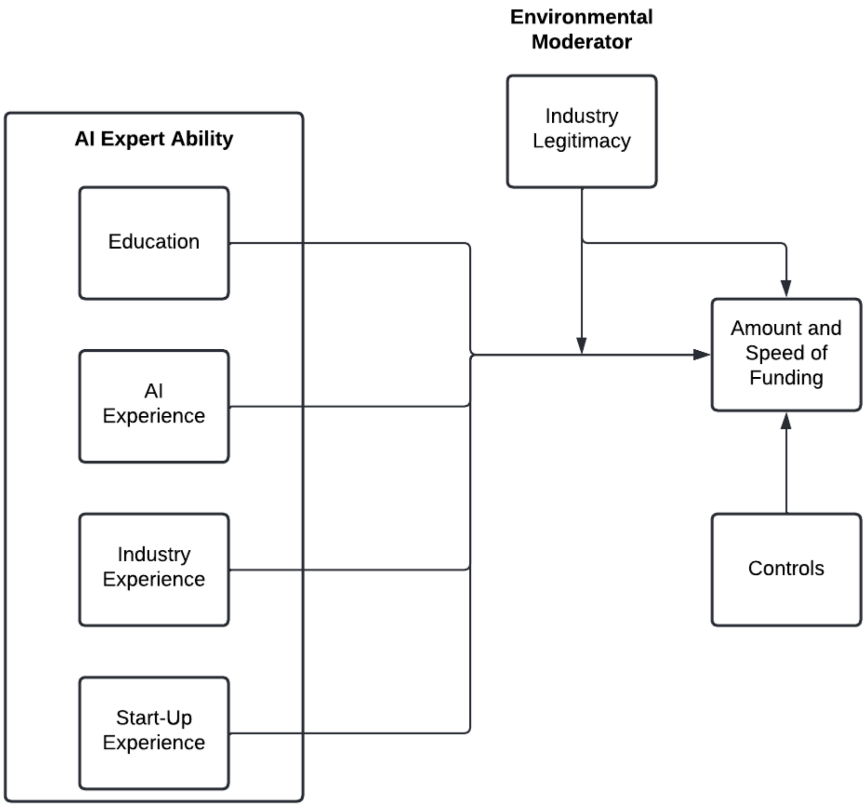
Homburg et al. (2014) address industry legitimacy as a potential source of uncertainty affecting AI experts impacts when discussing the institutional environment. Industry legitimacy, according to Aldrich and Fiol (1994), refers to accepted organizational practices, standards, ideas, models, and processes in the industry. They say that potential investors have a thorough understanding of how successful businesses in the specific market should run when industry legitimacy is high. This suggests that a new business has a higher need to build legitimacy while operating in recently founded industry with low legitimacy. Because VCs are likely to rely even more on signals of venture quality in an emerging industry than in more established industry scenarios, this raises the AI expert's importance for establishing legitimacy (Hsu, 2007). Consequently, the roles of education and experience as indicators of AI experts' ability to complete informational and relational tasks may be even more relevant than in mature industries. Hsu (2007) adds that low industry legitimacy strengthens the role of education as an indicator of relevant contacts.

In addition, the suppliers, vendors, and purchasers in this emerging, low-legitimacy market have not yet been confirmed (Macdonald, 1985). Existing relationships and other experiences from prior industry experience may therefore be even more crucial than in more established, mature sectors.

Overall, industry legitimacy acts as a moderator variable because it influences the strength and direction of the relationship between the education and experience of AI experts and their likelihood of securing faster and more VC funding. This brings us to the fifth assumption:

***H5**: The positive effects of AI expert education, AI expert AI experience, AI expert industry experience, and AI expert startup experience on the amount and speed of acquiring VC increase as industry legitimacy decreases.*

Figure 1 – Conceptual Framework



2.6 Most Relevant Predictors Using XGBoosting

In recent years, researchers have shown a great interest in understanding the variables that influence the amount of funding from venture capital firms, and the duration until first-round funding (Ross et al., 2021). The identification of these variables is crucial for startups seeking funding as well as VCs making investment decisions. Researchers can get important insights into the dynamics of the entrepreneurial financing process by examining the factors that have a major impact on the amount of VC capital and the pace at which funding is received (Ross et al, 2021).

To this end, academics have used a variety of statistical and machine learning methodologies to model and forecast the results of VC investing (Ünal et al., 2019). The use of eXtreme Gradient Boosting (XGBoost) algorithms is one powerful approach that is gaining popularity in the industry. XGBoost is an ensemble learning technique that enhances prediction performance and feature importance analysis by fusing the benefits of gradient boosting and decision trees. According to Chen & Benesty (2016), XGBoost is an effective, scalable implementation of Friedman (2001)'s gradient boosting framework. Although the principles of gradient boosting are the same in both XGBoost and gradient boosting methodology (GBM), XGBoosting uses a more regularized model formalization to control overfitting, providing superior performance. The GBM framework is comparable to XGBoost, however XGBoost is more efficient. Data scientists frequently utilize it to respond to machine learning difficulties with state-of-the-art solutions (Chen & Guestrin, 2016). The fact that XGBoost is employed in more than half of the winning responses to machine learning challenges run by Kaggle is evidence of its accuracy (He, 2016). According to Elith et al. (2008), the main disadvantage of single-tree models is their relatively poor predictive accuracy, which is mitigated by XGBoosting's ability to fit multiple trees. A sequence of brief decision trees can be generated sequentially using the boosting strategy, with each tree being constructed after the one before it. Every decision tree is individually modified to the model residuals and added to the fitted function to update the residuals. James et al. (2017) states that this improves the model, especially because the trees can be relatively small, with just a few terminal nodes. Eventually, the final model consists of a linear combination of hundreds or even thousands of trees. Elith et al. (2008) even thinks of it as a regression model where each term is a separate tree.

In the context of VC financing, XGBoost offers a useful tool to identify the most important factors that affect how much capital startups secure and how quickly they receive first-round funding. Researchers can quantify the relative importance of each predictor in influencing funding outcomes by training an XGBoost model on a dataset that includes important predictor variables, such as founder qualities, firm-specific attributes, market conditions, and industry factors (Dellermann et al., 2017).

For both researchers and practitioners, the ability to pinpoint the most crucial predictors through XGBoost analysis has substantial implications. First, from a research perspective, identifying the key variables that affect VC financing might improve our understanding of the underlying processes influencing investment choices. These findings can help establish the theoretical foundations of entrepreneurial finance and provide insight into the elements that influence funding decisions in the fiercely competitive and highly dynamic startup ecosystem. Second, from a practical perspective, understanding the most relevant indicators can offer business owners and startup founders valuable guidance. Entrepreneurs can better position their ventures to attract VC investors by customizing their pitches, improving their business plans, and identifying the factors that have the most impact on investment quantities and duration. Investors can use these insights to minimize risks and maximize returns by using them to make educated judgments about investment possibilities.

By employing XGBoost analysis in the context of VC financing, I am able to uncover the key determinants that drive the amount and pace of VC funding, enabling a more comprehensive understanding of the factors influencing the amount of VC capital and the speed at which it is obtained. The insights gained from such analyses have the potential to inform both academic research and entrepreneurial practice, ultimately contributing to the growth and success of the startup ecosystem.

3. Methodology

The methodology section of the current study will be explained in the parts that follow by first going into greater detail about the dataset and the measurements used to obtain the values before explaining the methods used.

3.1 Sample

This research builds upon the data gathering techniques from the aforementioned paper Homburg et al. (2014). Data scarcity is a definite challenge for empirical research on new ventures, according to Srinivasan et al. (2008). Therefore, a unique multisource dataset was compiled. This study obtained a sample from

Crunchbase, which is a comprehensive online platform that serves as a centralized database of information on companies, startups, investors, and other professionals in the business ecosystem. Moreover, Crunchbase offers an academic research program that provides researchers with the ability to receive either fully free or discounted access to the Crunchbase dataset on a case-by-case basis. After a few months of applying and waiting, during which Crunchbase thoroughly reviewed my academic intentions, I received an API key that granted me access to the Crunchbase database for a limited period of time.

Crunchbase is widely used by entrepreneurs, investors, researchers, and analysts to access valuable insights and data for market research, investment analysis, and networking purposes. Although this is the first AI-related study that makes use of Crunchbase, the platform has been used earlier for studies related to entrepreneurship, finance, and marketing.

To have the possibility to collect enough information on startups that have only received VC funds once, while also being recently founded, I set the funding date between January 1, 2020, to April 24, 2023. Additionally, I only included firms founded from 2020 onwards.

Initially, the dataset contained 1,725 firms. Dissolute firms, firms that underwent an IPO, acquired firms, firms with relevant missing data, and firms with more than one funding round were then eliminated. In addition, the focus of the study is on seed and early-stage new ventures, so firms older than 2020 and with more than 250 employees were also eliminated (Homburg et al., 2014). These eliminations resulted in 412 remaining firms.

Then, team compositions were analyzed for every firm to see whether an AI expert was present. After careful consideration and analysis of the AI-related jobs that exist within the startup environment, I have narrowed down the AI expert roles to the following: AI Engineer, AI Developer, Machine Learning Engineer, Data Scientist, AI Researcher, and Natural Language Processing (NLP) Engineer. Out of the 412 startups, 57 startups had an AI expert in their startup team. For every AI expert, biographical data was collected. Specifically, I coded the variable "AI expert presence" as 1 if there is a team member with an AI-related job present. Company websites and LinkedIn were used to collect the aforementioned data manually for every single company. These websites also allowed this study to include a variable named "Number of Employees" to add to the sample. So, for each startup identified, detailed data was manually collected. This involved meticulously examining company websites and LinkedIn profiles to gather the necessary information. Moreover, these manual efforts enabled the study to include an additional variable, 'Number of Employees,' enhancing the overall sample and enriching the analysis. Overall, this was a very educational but tough experience.

3.2 Measurement

Consistent with the two-step Heckman regressions, I use total *amount of funding* and *days until funding* as dependent variables, which is similar to Homburg et al. (2014)'s first and second analysis. Total amount of funding refers to the amount of USD that were collected by the startups. Days until funding refers to the number of days between the startup's founding and first funding date. All 412 startups include a funding date which means that all cases will be useful for the analysis part.

To measure AI expert education, the Homburg et al. (2014) approach will be adopted. Binary variables are coded as 1 if the respective employee studied at one of the prestigious schools mentioned in Palmer and Barber (2001). These prestigious schools are Columbia University, Dartmouth College, Harvard University, Massachusetts Institute of Technology, Northwestern University, Stanford University, University of California (Berkeley and Los Angeles), University of Chicago, University of Michigan, and University of Pennsylvania. In addition, the variables industry experience, job experience and startup experience were all coded as dummy variables for AI experts and CAIOs whenever the respective AI expert had experience with the industry, job or a startup.

Eventually, combining previous findings would lead to dummy variables for AI expert characteristics: AI expert education, AI expert experience, AI industry experience and AI expert startup experience.

To add an industry-level moderator, Crunchbase industry description variable was used to create an industry description with assigned four-digit SIC levels. This has been realized by using text-mining techniques which could rewrite industry descriptions based on certain words to the 12 industry descriptions found in table 1. Eventually, legitimacy scores are given for each industry based on Homburg et al. (2014). In their paper, five academic raters assessed the industries in line with the following specifications: 1 = very low cognitive legitimacy, 10 = very high cognitive legitimacy (See table 1). Every score allocation was manually checked by me.

Table 1 – Sample composition for Industry Legitimacy

SIC	*Industry Description*	*Frequency*	*Percentage*	*Industry Legitimacy*
2836	Biological products, Except diagnostic	22	0.05	4.8
3570	Computer and office equipment	18	0.04	8.4
3944	Games and video	13	0.03	8.4
5961	Catalog and Mail-Order Houses	27	0.07	6.8
7300	Enterprise	21	0.05	9.0
7311	Advertising	4	0.01	9.0
7372	Prepackaged software	260	0.63	7.8
7374	Data processing and preparation	15	0.04	5.2
7379	Computer related services	7	0.02	5.3
8200	Education	13	0.03	8.8
8700, 8741, 8742	Business and management services	12	0.03	8.6
	Total		100	

Besides the aforementioned variables, the study also controls for location, human capital, and type of funding. Firstly, Crunchbase provides information on the location where the startup was founded. In this study, this location has been transformed into dummy continent variables: Europe, Americas, Asia, Oceania, and Africa. The dummy value of 1 indicates that the startup was founded in a specific continent, while 0 indicates it was not. Controlling for location is important as it can significantly impact the startup's access to resources, funding networks, and investor ecosystems. Different regions may have varying levels of entrepreneurial activity, investment climate, and support infrastructure. By considering location as a control variable, researchers can account for these regional differences and isolate the effects of other variables on funding speed.

Additionally, numeric variables such as the number of founders and number of employees are included in the study. Controlling for these variables is crucial when investigating the factors that influence the speed of venture capital (VC) funding for startups. The number of founders directly affects the startup's ability to develop and execute its business plan. A higher number of founders may indicate a wider range of skills, expertise, and resources, which could positively influence its attractiveness to investors. Similarly, the number of employees reflects the startup's capacity to scale and carry out its operations effectively. A larger workforce may signify a higher level of organizational readiness and capability to handle increased funding. By controlling for these variables, researchers can isolate the effects of other factors, such as industry sector, innovation potential, or team composition, and determine their specific impact on the speed of VC funding.

Lastly, the study also controls for the funding status of the firm, specifically seed or pre-seed, as well as the last funding type received from VC, which could be pre-seed, seed, or series A funds. Pre-seed funding occurs at an earlier stage when startups are refining their ideas and transforming them into viable business concepts. Seed funding follows the pre-seed stage and supports early operations and product development. Series A funding represents a significant funding round that occurs after a startup has progressed through its early stages, secured seed funding, and demonstrated growth potential and market traction.

In summary, controlling for location, the number of founders, the number of employees, and funding status allows researchers to examine the specific effects of other variables on the speed of VC funding for startups while accounting for regional differences, human capital factors, and the stage of development.

3.3 Methods

3.3.1 Heckman Selection Model

To model the relationships between AI expert characteristics, amount of funding and pace of funding, a regression with continuous variable Y_i has been specified. Furthermore, I drew inspiration from the methodology proposed by Wooldridge (2003) and Homburg et al. (2014) and applied it to my dataset. Specifically, I employ a two-step sample selection approach known as the Heckman selection model. Heckman's idea was to treat selection bias in the model. Selection bias arises when the process of selecting a sample for analysis is not random or representative of the underlying population. This can lead to biased estimates of the relationships between variables of interest. In Probit models, selection bias often occurs when there is non-random sample selection, such as when certain observations are more likely to be included in the sample based on their characteristics (Wooldridge, 2003). According to Morrissey et al. (2016), choosing whether it is appropriate to use Heckman type models to investigate sample selection bias, the data under analysis must meet a number of criteria. Firstly, there must be a full set of observations for each variable for both participants and non-participants. Secondly, there must be a dependent variable in the selection model that is an appropriate proxy for participation and non-participation. Lastly, there must be an appropriate exclusion variable in the selection model.

I employ the Probit model, which models a binary outcome variable as a function of explanatory variables, to account for selection bias in the presence of an AI expert (See Formula 1 for equation and R code). In this case, the binary outcome variable is *AI_Expert_Presence*. For both startups, with and without an AI expert, the dataset contains complete observations for each variable. However, the Probit model assumes that the sample is randomly selected from the population, which is not the case when there is selection bias associated with the presence of an AI expert (Wooldridge, 2003). More specifically, I find significant effects of the variables *CTO_Presence*, *continent_Europe*, and *Founded_Year_2022* on the selection process of startups with an AI expert at the time of their first-round. The results of this equation are utilized to construct a variable that captures the selection effect in the equations measuring the time until funding and the funding amount. Nevertheless, research evidence demonstrates that the Heckman approach can considerably inflate standard errors if there is collinearity between the correction term and the included regressors (Morrissey et al., 2016). To ensure non-collinearity between the outcome equation and the selection equation, the selection equation must incorporate an observed variable, Z_i, that affects individuals' decision to participate in the study without influencing the outcome variable. This variable is referred to as the exclusion restriction, which in this study is represented by the variable *CTO_Presence*. Overall, this section demonstrates that the model satisfies the standards for the Heckman selection model. The results of the first stage of the probit model can be found in table 3 in the appendix.

Then, the Inverse Mill's Ratio (IMR) is introduced to correct for the selection bias. To account for the non-random sample selection based on the presence of an AI expert, a new term is introduced to the Probit model, which is shown in the extension of formula 1 in the appendix. The IMR represents the ratio of the probability density function (PDF) of the observed outcome variable (the absence of an AI expert) (Wooldridge, 2003). In other words, it quantifies the relationship between the probability of selection into the sample (presence of an AI expert) and the underlying probability of the event occurring (other variables influencing the presence of an AI expert). The previously calculated IMR will also be included as an independent variable, which takes into account the presence of an AI expert as a selection variable.

In the second step of the two-step sample selection, I establish multiple linear regressions with the Log-form of *days until funding* and *amount of funding* as dependent variables, inspired by Homburg et al. (2014). The variable *total amount funded* had a mean of 3,305,459.96, a median of 1,850,853, a standard deviation of 7,870,815.24, a skewness score of 12.44, and a Shapiro-Wilk test p-value of 2.62e-37. Based on these findings, it appears that the numeric dependent variable *total amount funded* is heavily right-skewed and significantly deviates from a normal distribution. Regarding the variable *days until funding*, it had a mean of 454.28, a median of 451, a standard deviation of 271.97, a skewness of 0.42, and a Shapiro-Wilk test p-value of 8.62e-07. Although the mean, median, and skewness score suggest a relatively small right-skewness, the Shapiro-Wilk test provides strong evidence against the null hypothesis of normality. This implies that *days until funding* is highly unlikely to follow a normal distribution. In order to obtain the aforementioned descriptive statistics, I utilized basic descriptive statistics functions in R, such as *mean()*, *median()*, and *sd()*. Additionally, I employed the *hist()* function for generating histograms. For assessing skewness, I utilized the *skewness()* function from the moments package, and for conducting the Shapiro-Wilk test, I utilized the *Shapiro.test()* function from the stats package.

The appendix's formulas 2 through 7 describe the second stage of the two-step sample selection process. Here, for each formula, one will find both the mathematical equation and the corresponding R code. Formulas 2 and 3 only include control variables. The key effects based on AI characteristics are additionally included in Formulas 4 and 5, taking this a step further. Finally, the moderating relationships of industry legitimacy are included in formulas 6 and 7. Here, I want to see if the impact of a "third" variable, such as *industry legitimacy*, has an impact on the relationship between the characteristics of AI experts and the dependent variables. Moreover, A moderating variable, also known as an interaction variable, affects the strength or direction of the relationship between an independent variable and a dependent variable. Interesting to mention is that all formulas include the IMR variable retrieved from the Probit model, which is there to account for the non-random selection bias.

Although this methodology is mainly based on Homburg et al. (2014)'s analyses, one shortcoming of this study relates to the fact that my dataset only includes startups that have received funding, whereas Homburg et al. (2014) also includes startups without funding. Therefore, the main difference between our studies is that I examine the time until funding and the amount of funding conditional on startups receiving funding.

3.3.2 XG Boosting

After examining the effects of AI expert characteristics on venture capital funding, my objective was to construct a predictive model to identify the key variables to monitor when predicting the days until funding and the amount of funding. Researchers can gain valuable insights into the dynamics of the entrepreneurial financing process by exploring the factors that significantly influence the VC capital amount and the speed of funding acquisition. As a result, I developed an XGBoosting predictive model utilizing the same dataset used in the two-step sample selection model.

As reported by Chen and Benesty (2016), XGBoost is an efficient, scalable implementation of the gradient boosting framework originally proposed by Friedman (2001). Evidence of its accuracy is that XGBoost is used in more than half of the winning solutions in machine learning challenges hosted at Kaggle (He, 2016). Fitting multiple trees in boosting overcomes the biggest drawback of single-tree models: their relatively poor predictive performance (Elith et al., 2008). A sequence of brief decision trees can be generated sequentially using the boosting strategy, with each tree being constructed after the one before it. The residuals are updated by adding each individual decision tree to the fitted function and adjusting it for the model's residuals. The model can benefit from trees that are relatively small and have few terminal nodes (James et al., 2017). A linear arrangement of hundreds or perhaps thousands of trees make up the final model.

Each variable is a tree; hence it can be compared to a regression model (Elith et al., 2008).

XGBoost requires multiple parameters to be determined through learning from data. Controlling the best combination of parameters is necessary to optimize and improve the model. Tuning parameters usually regulate the model's complexity and are a key element for prediction. The most important parameters that I will use based on Carmona et al. (2019), Chen & Benesty (2016) and Ünal et al., (2019). *Number of rounds* or *maximum number* of iterations is the optimal number of rounds or trees required in an XGBoost model. It can be determined using cross-validation methods. *Maximum depth* or *size of a tree* is how many splits there are in each tree. The complexity of the boosted structure is controlled by the maximum depth. Higher depth enables the model to learn relationships that are extremely specific to a particular sample, which is used to control overfitting. After splitting the tree to the maximum depth chosen, XGBoost begins to prune the tree backwards and removes splits above which there is no benefit. The best results, which resulted in the lowest RSME for each predictive model, were achieved at a depth of 3 and 17 number of rounds for *total amount of funding* and at a depth of 3 and 12 number of rounds for *days until funding* (See formulas 8 and 9).

As mentioned before, the dependent variables *total amount funded* and *days until funding* were transformed to the logarithmic scale. The variable *total amount funded* had a mean of 3,305,459.96, a median of 1,850,853, a standard deviation of 7,870,815.24, a skewness score of 12.44, and a Shapiro-Wilk test p-value of 2.62e-37. Based on these findings, it appears that the numeric dependent variable *total amount funded* is heavily right-skewed and significantly deviates from a normal distribution. Regarding the variable *days until funding*, it had a mean of 454.28, a median of 451, a standard deviation of 271.97, a skewness of 0.42, and a Shapiro-Wilk test p-value of 8.62e-07. Although the mean, median, and skewness score suggest a relatively small right-skewness, the Shapiro-Wilk test provides strong evidence against the null hypothesis of normality. This implies that *days* *until funding* is highly unlikely to follow a normal distribution. Consequently, for both XG Boost predictive models, the dependent variables were log-transformed.

In summary, I built a predictive model to predict a quantitative response variable of startups that received funding between January 1, 2020, to April 24, 2023. The model was based on the XGBoost algorithm explained above. The two models compute the predicted *amount of first round funding* and the *days until first round funding* for startups. In successive rounds, the algorithm seeks to fit a model that maximizes its performance for the best combination of model parameters, learning from the relationship between the response and its predictors. Also, all models were computed in R and can be found under formulas 8 and 9 in the appendix. Under these formulas, I present the R code snippets that were used to compute the final models with aforementioned *number of rounds* and *tree size*. Overall, 412 instances were used.

4 Results

4.1 Main results

In this section, I will summarize the results of all analyses. First, I will provide a description of the different variables used in this analysis by presenting the summary statistics. Second, I will analyze the results of the first-stage probit model that examines the factors influencing the presence of an AI expert. Third, I will discuss the outcomes of the six two-step Heckman sample selection models. Lastly, I will elaborate on the most significant variables for predicting the total funding amount and the days until funding using the XGBoosting method.

4.1.1 Descriptive statistics

First, we will look at the fundamental descriptive statistics. The correlation matrix as well as the mean and standard deviation for each variable are displayed in Figure 2 and Table 2 in the appendices. Figure 2 shows a correlation matrix where blue dots denote highly correlated variables and red colors denote negatively correlated variables. This graphic demonstrates the strong positive correlation between AI expert characteristics. Since these characteristics are highly related, this was not a surprise. Additionally, there is a negative correlation between the factors pertaining to funding types and stages. This can be explained by the fact that we only examine one sort of funding per startup, thus if we examine a particular startup that obtains pre-seed money, it will not be able to receive Series A funding concurrently. Furthermore, the descriptive statistics table 2 provides an overview of the variables in the dataset. From the table we can observe that almost 36% of the startups in this study possess a CTO. Also, startups have an average of approximately 2 founders and around 18.7 employees. The industry legitimacy score is relatively high, with an average of 7.16. The duration between the startup founding date and funding date is approximately 449.7 days. About 13.7% of the startups have an AI expert as a team member, and they possess an average educational background in AI of around 1.9%. The AI experts have an average experience of 12.0% in AI-related roles and 11.1% in the industry. In terms of location, around 23.9% of the startups are founded in Europe, 55.7% in the Americas, 17.1% in Asia, 1.9% in Oceania, and 1.4% in Africa. The funding status is predominantly seed, with an average of 93.7%. Pre-seed funding is present in around 39.0% of the startups, seed funding in approximately 55.9%, and Series A funding in about 5.1%. Lastly, the startups are distributed across different years, with approximately 16.9% founded in 2020, 37.6% in 2021, 43.1% in 2022, and 2.4% in 2023. These descriptive statistics provide valuable insights into the characteristics of the startups in the dataset.

4.1.2 First Stage Probit Model

Table 3 presents the results of the factors determining the presence of an AI expert in a startup. Multiple variable combinations have been considered; however, Table 3 represents the final first-stage probit model, including significant variables and variables that one would expect to be significant in determining AI expert presence. Simultaneously, these results represent the outcomes of the first stage of the probit models within the two-step Heckman selection model. Among all the variables considered, these factors appear to have the most significant impact on AI expert presence. Therefore, my focus will primarily be on them. Firstly, the coefficient for *CTO_Presence* is 0.567, which is highly significant at $p < 0.001$. This suggests that for every unit increase in *CTO_Presence*, the log-odds of *AI_Specialist_Presence* increase by 0.567. Alternatively, exponentiating the coefficient ($e^{0.567}$) yields an odds ratio of approximately 1.76. This indicates that the odds of an AI expert being present are approximately 1.76 times higher when there is a Chief Technology Officer in the startup, compared to when they are absent. Secondly, the coefficient for *Europe* is 0.466, significant at $p < 0.001$. Exponentiating the coefficient ($e^{0.466}$) gives an odds ratio of approximately 1.59. This implies that the odds of AI specialist presence are roughly 1.59 times higher for startups located in Europe compared to those outside of Europe. Lastly, I found that startups founded in *the year 2022* have a negative impact on the presence of an AI expert. The coefficient associated with this variable is -0.464, significant at $p < 0.1$. Exponentiating the coefficient ($e^{-0.464}$) gives an odds ratio of approximately 0.629, indicating that the odds of an AI expert being present are approximately 0.629 times lower for companies founded in 2022 compared to companies founded in other years.

After estimating the logistic regression model above, the Inverse Mill's Ratio (IMR) is obtained for each observation in the dataset using the invMillsRatio function. The outcome, IMR, is then used in the next section to capture the relationship between the error term in the probit model and the error term in the outcome equation. This will eventually correct for the sample selection bias.

4.1.3 Model 1 – Controls only

Third, I will examine whether having an AI expert in the new venture matters before continuing to the second stage two-step Heckman sample selection models. On average, I find that startups with AI experts in their team only need 453.1 days until funding and receive $3,349,644.42 in funding, compared to startups without AI experts, 461.4 days and $3,030,276.04. In addition, Table 4 in the appendix displays two ordinary regression models with AI Experts as an independent variable, and log(*total amount of funding*) and log(*days until funding*) as dependent variables. The results indicate that AI expert presence is positively associated with the amount of funding (0.187, $p < 0.05$) and negatively associated with the number of days until funding (-0.096, $p < 0.1$). This suggests that the presence of an AI expert leads to an increase in funding by 20.6%[1], while decreasing the number of days until funding by 9.2%[2], holding everything else constant. Therefore, it is interesting and useful to examine the relationship between VC funding and AI experts further and continue with the second stage of the Heckman selection models. Hence, I create six different two-step Heckman selection models among three stages, which I will elaborate on in the following sections.

Now, let us have a look at the *controls only* model for the second stage of two step sample selection model. The outcome of model 1 is specified in the first column of tables 5 and 6 in the appendices, representing the two-step Heckman selection models with log(*total amount of funding*) and log(*days until funding*) as the dependent variables, respectively. In the following sections, I will first explain the model for total amount of funding in USD and then discuss the model for days until funding.

The first model in Table 5 represents the basic version of the two-step Heckman selection model, including only the control variables. In this model, I found that the *Funding Type Pre-Seed* variable significantly affects the total funding amount at $p < 0.001$. Specifically, the coefficient for this variable is -1.274, indicating that startups in the pre-seed stage experience a decrease of 72%[3] in the total amount of funding. This finding aligns with existing VC theory, which suggests that pre-seed funding provides the least amount of funding due to the early-stage nature of startups. At this phase, startups may have limited product development, customer traction, and revenue generation. The R^2 value of this model, which measures the proportion of the variance in the dependent variable explained by the independent variables, is 0.3968. This indicates moderate predictive power.

---

[1] $(e^{0.187}-1)*100\%$
[2] $(e^{-0.096}-1)*100\%$
[3] $(e^{-1.274}-1)*100\%$

The first model in table 6, in which the control variables are regressed against the dependent variable *days until funding*, shows that only the founded years 2020 and 2021 appear to be of a positive impact on *days until funding*. Both variables are significant at $p < 0.1$. This means that if both years' dummy variable is equal to 1, the days until funding will be extended by 336.7%[4] and 263.3%[5] respectively. The coefficients for the years indicate that startups founded in these years, experience a significant increase in days until funding, compared to the base year 2023. Specifically, startups founded in 2020 have the largest increase, followed by 2021. This suggests that more recent startups are able to secure funding at a faster pace. Compared to startups launched in 2020, companies founded in 2023 typically acquire funding more quickly. Several factors, including the COVID-19 period and potential dataset biases, can be ascribed to this. The economic recovery and rising investor confidence are two factors that contribute to the faster funding of later-year startups (Brown et al., 2020). As the pandemic's initial effects on the world economy eventually fade, investors rediscover confidence and are more willing to invest in potential projects. Startups established in 2023 are more likely to gain from this stronger economic environment, which will result in faster funding. Additionally, startups founded in later years have benefited from being able to observe and respond to market developments brought on by the pandemic. They have been successful in seeing new trends, coming up with creative solutions, and matching their business strategies to changing consumer demands. Because of their versatility, they are more appealing to investors, which speeds up the fundraising process (Brown et al., 2020). In contrast, during the early phases of the COVID-19 pandemic, startups created in 2020 had to deal with special difficulties and uncertainty. As a result of the pandemic's abrupt beginning, the investment environment was disrupted, making it harder for these startups to obtain capital. However, startups established later could benefit from what their forerunners discovered and use that expertise to better navigate the investment landscape, resulting in quicker funding outcomes. Additionally, only four months of data have been collected for the year 2023. As a result, firms founded in 2023

[4] $(e^{1.474}-1)*100\%$
[5] $(e^{1.29}-1)*100\%$

generally experience faster funding compared to other firms in the dataset. It is important to note that this limited timeframe also affects the year variables and should be considered as a potential limitation. The R^2 of this model, is only 0.003057.

4.1.4 Model 2 – Main effects

Additionally, I examine whether the main effects of education, job experience, industry experience, and startup experience affect the outcome variables *total amount of funding* and *days until funding*. The previous model, which included only control variables, was expanded to incorporate the four mentioned dummy variables. The results of both tests can be found in tables 5 and 6 in the appendices, in the second columns.

For the *total amount of funding* model, no major changes could be discovered (see Table 5). Therefore, only the variable *Funding Type Pre-Seed* remains significant at $p < 0.001$. This finding aligns with the fact that the pre-seed stage typically provides the least amount of funding as it occurs at the earliest phase of a startup's journey (Schwarzkopf, 2010). Moreover, the pre-seed funding type has a negative impact on the total amount of funding, reducing it by 71.9%[6] in this model. On the other hand, the variables AI expert education, AI expert experience, AI expert industry experience, and AI expert startup experience are found to be insignificant. The R^2 value for this model is 0.348, suggesting moderate predictive power.

For the *days until funding* model, I find an interesting new finding compared to the *controls only* model. Again, the founded years 2020 and 2021 are significantly positive at $p < 0.1$ (See table 6). This time their coefficients are 1.389 and 1.196, respectively. This implies that if the startup was either founded in 2020 or 2021, the days until funding would increase by 300.6%[7] and 230.4%[8]. Startups founded

[6] $(e^{-1.270}-1)*100\%$
[7] $(e^{1.388}-1)*100\%$
[8] $(e^{1.195}-1)*100\%$

in earlier years' experience a significant increase in the number of days until funding compared to those founded more recently. This trend can be attributed to factors such as the economic recovery and increasing investor confidence. Startups established in later years benefit from a stronger economic environment and the ability to observe and respond to market developments, making them more appealing to investors. They are also able to leverage the lessons learned from earlier startups, enabling them to navigate the investment landscape more effectively. Overall, the funding process for later-year startups is expedited due to these factors, while startups founded in 2020 faced unique challenges and uncertainties during the early phases of the COVID-19 pandemic. Also, the limitation within the dataset once again plays a role, since it only observes startups founded in 2023 for 4 months. However, in addition, *AI_Specialist_StartUp_Experience* has a significant effect on *days until funding* since $p < 0.1$. Surprisingly, its coefficient is 0.460, which implies that if the AI expert has prior startup experience, the days until funding increase by 58.4%[9]. The finding that startups with AI experts who have prior startup experience require more days to receive VC funding compared to startups with AI experts without prior startup experience may seem counterintuitive. However, there could be several explanations for this phenomenon. Firstly, startups with AI experts who have prior startup experience may have higher expectations and specific goals for their ventures. They may be more selective in choosing the right investor or funding opportunity that aligns with their vision and objectives. This selectivity and strategic approach could result in a longer search and evaluation process, leading to more days until funding. Secondly, startups with AI experts who have prior startup experience may have a higher level of ambition and aim for larger funding rounds. They may be targeting substantial investments to scale their operations or develop innovative AI technologies. Securing larger funding amounts often requires more time and effort as it involves negotiating with potential investors and meeting their specific criteria. Overall, this model's R^2 is only 0.09068.

[9] $(e^{0.46}-1)*100\%$

4.1.5 Model 3 – Final model

The last models build upon the previous two models and incorporate moderating relationships of industry legitimacy. This means that for the AI expert, its characteristics (education, experience, industry experience, and startup experience) will be interacting with the industry legitimacy score based on SIC levels. The output for both models can be found in Tables 5 and 6 in the appendices.

For the last model analyzing the *total amount of funding*, the variable *Funding Type Pre-Seed* once again emerges as significant at $p < 0.001$ (See table 5). This indicates that startups receiving pre-seed funding experience a decrease in total funding by 70.59%[10]. This finding aligns with the common understanding that the pre-seed stage typically offers less funding as it occurs at the early phase of a startup's journey (Schwarzkopf, 2010). Additionally, the variable *AI_Specialist_Education*, 7.061, is significant at $p < 0.05$ and positively influences total amount of funding. Several factors contribute to this finding. Firstly, a higher educational background in the field of AI signifies greater expertise and knowledge. Startups with AI specialists who have advanced education in AI-related fields demonstrate a deeper understanding of AI technologies, algorithms, and applications. This expertise attracts investor interest by offering innovative and cutting-edge AI solutions. Investors often support startups displaying technical proficiency and domain expertise. Secondly, a stronger educational background in AI implies a wider network and connections within the AI community. AI specialists with prestigious educational backgrounds may have studied at renowned institutions or collaborated with esteemed researchers. Such networks provide valuable resources, mentorship, and industry connections, enhancing the startup's visibility and credibility among investors. The reputation and network associated with a strong educational background positively influence investor perceptions and increase funding prospects. Additionally, a significant interaction effect is observed between the moderator *industry_legitimacy* and *AI specialist education*, indicating that the impact of industry_legitimacy on total

[10] $(e^{-1.224}-1)*100\%$

funding differs between firms with AI specialists possessing prestigious educational backgrounds and those without. Finally, the model exhibits a moderate R^2 value of 0.3691, suggesting that approximately 36.91% of the variance in the total amount of funding can be explained by the variables included in the model.

The *days until funding* model underwent slight changes compared to its previous version. In this iteration, only the founded year 2020 was found to be significant, albeit at $p < 0.1$. Its coefficient of 1.333 suggests that for every startup founded in 2020, the days until funding increased by 279.2%[11] compared to its reference year 2023. Surprisingly, the variable *Funding_Status_Seed* exhibited significance at $p < 0.05$, with a coefficient of -2.168. This leads to an 88.6%[12] decrease in days until funding holding everything else constant. None of the other control variables, moderating variables, or AI expert characteristic variables were found to be significant. Lastly, the model achieved an R^2 value of only 0.07463, implying that approximately 7.463% of the variation in the days until funding could be explained by the included variables.

Overall, based on the analysis conducted, the following conclusions can be drawn regarding the hypotheses. Firstly, H1 can be partially accepted as the significant coefficient of 7.061 indicates that including an AI expert with an education from a prestigious school increases the funding amount. However, this variable did not show sensitivity in the model predicting days until funding, leading to only partial acceptance of the hypothesis. Secondly, the second hypothesis was not supported by either of the two models. This suggests that AI experts' AI experience does not have an impact on VC funding or the time it takes for startups to secure funding. Thirdly, neither of the models provided support for the third hypothesis, which proposed that AI experts' industry experience would increase VC funding and expedite the funding timeline. Fourthly, the analysis reveals that AI experts' startup experience is positively associated with an increase in the number of days

[11] $(e^{1.333}-1)*100\%$
[12] $(e^{-2.168}-1)*100\%$

until funding, suggesting that startups with AI experts who have prior startup experience a longer duration before receiving funding. However, this finding contradicts H4 and therefore provides evidence to reject the hypothesis. Lastly, H5 is not supported by the study's findings.

4.1.6 XGBoosting – Predictive Models

In this section, I present two XGBoost predictive models that provide valuable insights into predicting the total amount of venture capital (VC) funding and the time it takes for startups to receive their initial funding. Leveraging the power of XGBoost, these models enable accurate predictions and offer a comprehensive understanding of the factors driving these outcomes. Additionally, the models shed light on the most influential variables that contribute to predicting the total funding amount and the duration until funding. By harnessing the predictive capabilities of XGBoost, we gain valuable insights into the funding dynamics of startups and uncover the key factors that shape the amount and timing of funding.

4.1.6.1 Total Amount of Funding

The first model, *total amount of funding*, with a max depth of 3 and 17 number of rounds performed outstanding on a 70-30% split. In order to evaluate its performance, I will refer to MSE, MAE, and RMSE. The MSE measures the average of the squared differences between predicted and actual values, the MAE measures the average of the absolute differences between predicted and actual values, and the RMSE is the square root of MSE and represents the standard deviation of the residuals. Lower values of MSE, MAE and RMSE indicate better model performance, as they indicate smaller prediction errors and closer assignment between predicted and actual values. The results of predicting the log *total amount of funding* were MSE 0.931, MAE 0.725, and RMSE 0.965. However, transforming these number back to normal by taking the exponential function of the outcome number gives us the following outstanding results of MSE 2.537, MAE 2.065 and RMSE 2.625. The results of the XGBoost predictive model for predicting the *Total Amount of Funding in USD* show promising performance. Considering that the dependent variable, *Total Amount of Funding in USD*, typically involves values in the range of many millions, the achieved error metrics indicate relatively small prediction errors. The MSE and RMSE values quantify the average squared and root mean squared differences between predicted and actual values, respectively, while the MAE represents the average magnitude of the errors. With the MSE, MAE, and RMSE values being low in relation to the magnitude of the predicted variable, it suggests that the XGBoost model performs well in capturing the patterns and variations in the Total Amount of Funding. These results indicate the model's ability to provide accurate predictions, bringing us closer to understanding and estimating the funding amounts for startups in the millions range.

In the journey of predicting the total amount of funding for startups, certain variables have emerged very interestingly as crucial factors (See table 7). Among them, the number of employees stands tall as the most important factor, emphasizing the significance of team size in attracting funding. It could be that startups with larger teams often have an advantage in scalability and execution capabilities, making them more attractive to investors. A second key variable is the funding status at the pre-seed stage. Presumably, this makes sense as the pre-seed stage is when startups typically require a certain amount of money, which often falls within a similar range. Therefore, this variable is of utmost importance when predicting the total amount of funding. The last variable that is important for this predictive model, is industry legitimacy. The latter finding suggests that the amount of funding received is strongly influenced by the industry in which a startup operates.

4.1.6.2 Days Until Funding

The second model, which predicts the *number of days until funding*, performed exceptionally well with a maximum depth of 3 and 12 rounds. Evaluating its performance using MSE, MAE, and RMSE, the results for predicting the log total amount of funding were MSE 0.902, MAE 0.634, and RMSE 0.949. When transforming these numbers back to their original scale by taking the exponential function, we observe outstanding results of MSE 2.465, MAE 1.885, and RMSE

2.583. These results indicate that the model accurately predicts the number of days until first-time funding, with an average deviation of less than 3 days.

The most important variables for predicting the days until funding are as follows (See table 8). Firstly, the founded year 2020 emerges as the most influential variable, aligning with the findings of the earlier two-step Heckman selection model, which indicated a significant positive impact of the year 2020 on the duration until funding. Presumably, the challenges brought about by the onset of the COVID-19 pandemic in 2020 made it harder and, consequently, longer for startups to secure funding. Secondly, the number of founders also proves to be significant in predicting the days until funding. Thirdly, the number of employees is another important factor in predicting the duration until funding.

5 Conclusion

In this section, I first summarize its main findings. Second, I elaborate on the implications for theory and directions for further research. Then, I list the implications for practice before summarizing the limitations of this research.

This study is an attempt to shed light on the role of AI experts in acquiring venture capital. Since AI experts are yet to become more popular in startup team composition, all AI-related team members were included in this study. However, to still zoom in on a more specific role, such as the CMO role in Homburg et al. (2014), this study used the AI expert role. Hence, this study tries to argue that AI experts might increase the startup's chances of obtaining more and faster funding from VCs.

Coming back to this study's research question, *'How do AI experts' characteristics affect VC funding?'*. The current study's results suggest that for AI experts, only education is positively related to the amount of funding. In addition, AI experts' startup experience has a positive effect on days until funding. Furthermore, there is no direct relationship found for AI expert AI experience and AI expert industry experience. Furthermore, no significant effects have been found when industry legitimacy increases, which contrasts with findings from prior studies. However, other interesting and relevant findings can be concluded besides prior mentions.

First of all, I have found that the presence of a CTO and Europe as the founding continent increases the chances of having an AI expert present. Conversely, the year 2022 decreases the chances of having an AI expert on board.

For the total amount of funding model, I found that the dummy variable "pre-seed" has a negative impact on the total amount of funding. However, this finding aligns with prior literature in VC theory, which suggests that the pre-seed stage provides the least amount of funding as it occurs at the earliest phase of a startup's journey (Schwarzkopf, 2010). During this stage, the business idea is still in its nascent state and may lack a solid foundation or a proven track record. Startups at this stage often have limited or no product development, customer traction, or revenue generation. Furthermore, startups with AI experts from prestigious colleges tend to receive more funding, as indicated by the positive effect of this dummy variable on the total amount of funding. AI experts with a stronger educational background have a positive impact on VC funding because it signifies a greater level of expertise and knowledge. Additionally, a stronger educational background in AI implies a wider network and connections within the AI community, providing access to valuable resources, mentorship opportunities, and industry connections. Overall, these findings emphasize the importance of the startup stage and the educational background of AI experts in shaping the total amount of funding received. Startups at the pre-seed stage face challenges due to their early development, while startups with AI experts from prestigious colleges have advantages in terms of expertise and networking opportunities.

For the days until funding model, I found that the early years in this analysis, 2020 and 2021, had a positive effect on the days until funding. This suggest that recently founded firms receive funding faster. Factors such as COVID-19 can be ascribed to this. The COVID-19 period has influenced the funding dynamics for startups. Economic recovery and increasing investor confidence have contributed to faster funding for startups established in later years (Brown et al., 2020). As the initial effects of the pandemic subside, investors regain confidence and become more open to investing in new projects. Startups founded in 2023 are poised to benefit from this improved economic environment, leading to expedited funding. Moreover, startups established in later years have capitalized on their ability to observe and adapt to market developments driven by the pandemic. Their agility in identifying emerging trends, innovating solutions, and aligning their strategies with evolving consumer demands makes them more attractive to investors, accelerating the fundraising process (Brown et al., 2020). In contrast, startups founded in 2020 faced unique challenges and uncertainties during the early phases of the pandemic. The abrupt onset of the pandemic disrupted the investment landscape, making it more challenging for these startups to secure capital. However, later-stage startups have leveraged the knowledge gained from their predecessors to navigate the investment landscape more effectively, resulting in faster funding outcomes. Besides, the dataset only spans four months into 2023, which creates the impression that firms founded in 2023 receive funding at a faster rate compared to firms founded in prior years. In addition, I found that earlier startup experience among AI experts leads to more days until funding. Overall, the combination of selective decision-making and ambitious funding goals may explain why startups with AI experts with prior startup experience face extended periods until receiving funding.

Lastly, I developed two predictive XGBoosting models to predict the total amount of funding and the days until funding. Overall, these models demonstrated a high level of accuracy, with predictions deviating by an average of only $2 and fewer than 3 days. Furthermore, the XGBoosting models allowed me to identify the most important variables in predicting these outcomes. Understanding these variables is crucial for startups as it provides insights into the key factors that influence funding amounts and the duration until funding. In the total funding amount model, the most important variables were *Number_of_Employees*, *FT_Pre_Seed*, and *Industry_Legitimacy*. In the days until funding model, the most important variables were *Founded_Year_2020*, *Number_Of_Founders*, and *Number_Of_Employees*.

5.1 Implications for theory and directions for further research

This study contributes to three research streams. Firstly, this study adds to the relatively new research stream at the AI-entrepreneurship interface by investigating the role of AI experts. Although a lot of prior research has been done on other roles within the management team, the role of AI experts has been neglected in the past. Therefore, we address the role of AI experts, and our results show that AI expert startup experience slows down the process of obtaining VC funding, but excellent education increases VC funding. AI experience, and industry experience have little to no impact. Further research might examine the role of AI experts in relation to other KPIs, such as customer satisfaction, product quality, and sales volume. Secondly, this study adds insights into AI strategy and financial performance of startups, which have not been studied before in this context. Prior research has studied it within the boundaries of marketing and focused merely on investor reactions to stock-market listed startups or IPOs (e.g., Homburg et al., 2014; Rao et al., 2008). This study takes the role of AI into account and examines VC reactions. Additionally, introducing days until funding from VCs as a dependent variable to the AI-Finance literature stream enables examination of other AI-related variables that have not been researched yet. Thirdly, this paper contributes to the existing research on AI experts by presenting a predictive model and highlighting its most important variables. The XGBoosting model demonstrates its effectiveness in accurately predicting venture capital (VC) funding amounts and timing both currently and in the near future. By leveraging advanced machine learning techniques, this model enhances our understanding of the factors that drive VC funding decisions. The identification of these crucial variables offers valuable insights for startups and investors seeking to optimize their fundraising strategies and improve decision-making. Overall, this research advances the field by providing a robust predictive framework that can support AI expert-related studies and guide practical applications in the VC funding domain.

5.2 Implications for practice

Based on the findings from the two-step Heckman analyses and the XGBoosting models, several practical implications can be drawn to help startups obtain venture capital funding efficiently.

Firstly, the outcome of the analyses showed that the years 2020 and 2021, were probably impacted by the covid-19 pandemic. This had a negative impact on the days until first time funding. This shows that raising venture capital was more difficult for startups established during these years. Entrepreneurs must be aware of the various challenges and uncertainties they may face during these times, and they must take proactive measures to overcome those challenges in their funding strategy. Their prospects of getting finance could be increased by developing strong networks, showcasing resilience, and adapting business models to the shifting market conditions brought on by the pandemic.

Secondly, the time till funding is significantly enlarged when an AI expert with startup experience is part of the startup team. This shows that hiring an AI expert with startup expertise may cause the funding process to take longer. It is crucial to keep in mind, that this adverse effect can be related to venture capitalists' heightened expectations for the team's competencies and the requirement for meticulous due diligence. Startups with AI experts should highlight the distinctive knowledge and value they offer, emphasizing their capacity to foster innovation, create cutting-edge technology, and deftly manage market trends. Additionally, utilizing the AI expert's experience to build solid alliances, prove scalability, and present a clear route to market success may assist shorten the overall time to success.

Thirdly, startups aiming to secure significant venture capital should consider recruiting AI experts who have graduated from high quality colleges or universities. The findings suggests that AI experts with a prestigious educational background are more likely to attract investor attention and receive larger funding amounts. Startups can leverage this knowledge to strategically build their team by prioritizing candidates with strong educational credentials in the field of AI. Additionally, highlighting the educational qualifications of their AI experts in pitches and presentations may enhance their credibility and increase their chances of securing VC funding.

Last but not least, it is crucial to comprehend the variables that have the most significant impact when predicting funding amounts and the duration until funding. Startups must strive to fully grasp these variables to maximize their funding potential and expedite the funding process. Regarding funding amounts, the most influential variables are the number of employees, the funding stage in which the startup operates, and the industry legitimacy score. In terms of the duration until funding, the key variables are the year of startup foundation, the number of employees, and the number of founders. By understanding and strategically addressing these variables, startups can optimize their funding prospects and streamline their path to securing financial support.

5.3 Limitations

Unfortunately, there are also a couple of limitations attached to this research that need to be carefully taken into account when adding to this research in the upcoming period.

Firstly, the results of the two-step Heckman analyses revealed that investment stages, years, education, and startup experience significantly influence both the amount and duration until funding. However, it is important to note that while these findings provide valuable insights, further research is necessary to confirm the underlying explanations for these relationships. Additional investigations should be conducted to validate the assumptions and delve deeper into the specific mechanisms through which these factors impact the outcome variables. By conducting more comprehensive research, we can enhance our understanding of the complex dynamics between these variables and funding outcomes, thereby providing a more accurate and robust foundation for decision-making and strategic planning in the startup ecosystem.

Secondly, this paper attempted to include control variables that were also used in prior papers on CMOs and CFOs. However, these controls may be considered less effective in the case of AI experts. For example, examining AI experts' personal success in obtaining financing or exploring personality-related variables might increase confidence in the robustness of our results. This is just one example, and there are many more possibilities that could be used as control variables.

Thirdly, this was an extensive research project to be executed by a single person. Both financially and timewise, it was a challenging task to accomplish successfully. Firstly, CrunchBase data is not widely available, and an expensive yearly subscription was required to obtain useful startup data. However, as a scholar, an API was provided after a long application process. This API key allowed for data collection for a short period of time. Therefore, data had to be collected as precisely as possible once the opportunity arose. This time constraint made it extremely challenging to gather all the required data accurately. As a result, I could only collect data on startups that have received funding. Additionally, researching every startup manually to gather information via LinkedIn and company websites was a very time-consuming task. LinkedIn company information could only be accessed with a monthly premium subscription, which was costly. Thus, time was limited, and it was not possible to research as many startups as preferred. If time and money were not limited, this research could be further extended, leading to more robust outcomes.

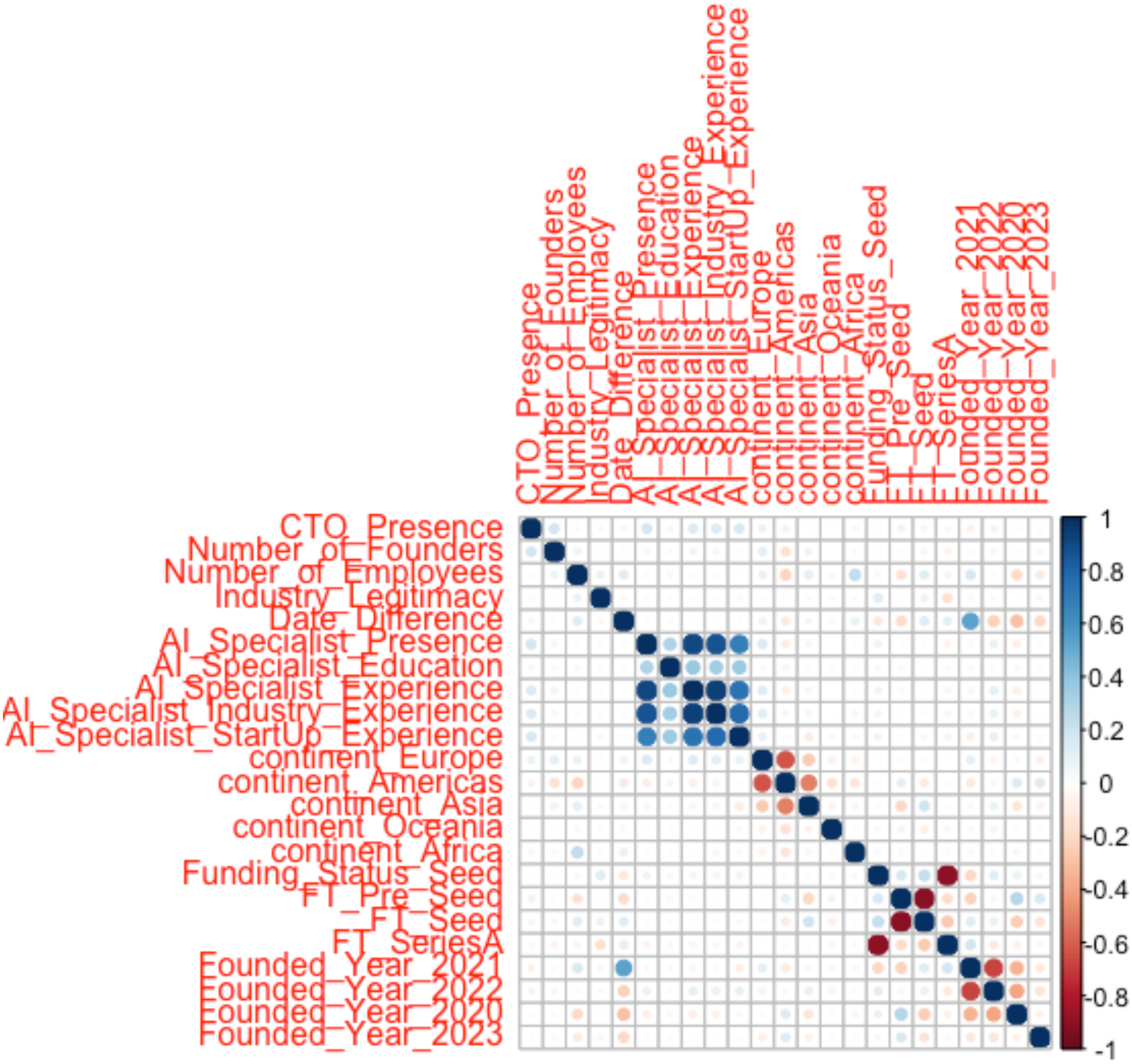
Lastly, one limitation of this study is the presence of a relatively large correlation among the AI expert characteristics, including education, job experience, industry experience, and startup experience (see Figure 2). This high correlation among the variables might introduce multicollinearity, which could potentially impact the accuracy and stability of the analyses and can also be expressed by the Variance Inflation Factor (VIF). To provide a general understanding of the VIF values, I will use model 3 as an example. Here, the VIF values are 1.2 for AI expert education, 7.7 for AI expert experience, 8.6 for AI expert industry experience, and

2.6 for AI expert startup experience. In summary, VIF values less than 5 are generally considered acceptable, indicating low multicollinearity. The variables with VIF values above 5 but below 10 (e.g., AI_Specialist_Experience and AI_Specialist_Industry_Experience) may have some moderate correlation with other predictors, but it is not a severe concern. However, if the VIF values were much higher (e.g., exceeding 10), it would indicate more problematic multicollinearity, which could potentially affect the reliability of coefficient estimates and interpretation of the regression model (O'Brien, 2007). Nevertheless, future research could explore additional techniques, such as dimension reduction methods or alternative modeling approaches, to further address the multicollinearity issue and enhance the robustness of the analyses.

6 Appendices

6.1 Figures

Figure 2 – Correlation matrix



6.2 Tables

Table 2 – Descriptive statistics

Variable Name	*Mean*	*Standard Deviation*
CTO Presence	0.3590361	0.4802968
Number of Founders	1.99518072	0.8819039
Number of Employees	18.66987952	23.1772092
Industry Legitimacy	7.15590361	1.0001486
Date Difference	449.72530120	276.5383092
AI Expert Presence	0.13734940	0.3446313
AI Expert Education	0.01927711	0.1376632
AI Expert Experience	0.12048193	0.3259172
AI Expert Industry Experience	0.11084337	0.3143170
AI Expert Start Up Experience	0.07469880	0.2632220
Continent Europe	0.23855422	0.4267140
Continent Americas	0.55662651	0.4973827
Continent Asia	0.17108434	0.3770372
Continent Oceania	0.01927711	0.1376632
Continent Africa	0.01445783	0.1195124
Funding Status Seed	0.93734940	0.2426260
Funding Type Pre-Seed	0.39036145	0.4884201
Funding Type Seed	0.55903614	0.4971018
Funding Type Series A	0.05060241	0.2194490
Found in 2020	0.16867470	0.3749163
Found in 2021	0.37590361	0.4849399
Found in 2022	0.43132530	0.4958591
Found in 2023	0.02409639	0.1535336

Table 3 – Results first stage probit model

DV: AI Specialist Presence	*First stage probit model*
Intercept	-1.358***
CTO Presence	0.567***
Europe	0.466**
Found in 2022	-0.464’
Number_of_Employees	-0.015
Number_of_Founders	-0.001
Observations	*412*

***p < 0.001 **p < 0.01 *p < 0.05 ‘p < 0.1

Table 4 – Regular regression with AI expert presence as independent variable

Controls Only	*DV: Total Amount of Funding*	*DV: Days Until Funding*
Intercept	16.106***	4.868***
AI Specialist Presence	0.187*	-0.096’
Number of Founders	-0.007	0.034
Number of Employees	0.004’	0.000
Industry Legitimacy	-0.142**	-0.022
Europe	0.062	-0.356
Americas	0.191	-0.316
Oceania	-0.278	-0.528
Asia	-0.120	-0.340
Africa	-	-
Funding Status Seed	-0.434	0.003
Funding Type Pre-Seed	-1.841***	-0.115
Funding Type Seed	-0.692	-0.160
Funding Type Series A	-	-
Founded in 2020	0.567’	1.894***
Founded in 2021	0.659*	1.295***
Founded in 2022	0.360	0.831**
Founded in 2023	-	-
Observations	*412*	*412*

***p < 0.001 **p < 0.01 *p < 0.05 ‘p < 0.1

Table 5 – Relationship between AI expert characteristics, Industry Legitimacy and amount of VC funding

DV: Total Amount of Funding	*Model 1 Controls Only*	*Model 2 Main Effects*	*Model 3 Final Model*
Industry Legitimacy			
AI expert education x industry legitimacy			-0.980*
AI expert experience x industry legitimacy			-0.644
AI expert industry experience x industry legitimacy			0.549
AI expert startup experience x industry legitimacy			0.044
Main Relationships			
AI expert education		0.205	7.061*
AI expert experience		-0.222	4.322
AI expert industry experience		0.003	-3.98
AI expert startup experience		-0.094	-0.158
Control Variables			
CTO presence	-	-	-
Number of founders	0.129	0.127	0.194
Number of employees	0.015	0.015	0.015
Industry legitimacy	-0.194	-0.175	0.257
Europe	0.430	0.514	-0.031
Americas	-0.209	-0.154	-0.723
Oceania	1.598	1.638	1.305
Asia	0.318	0.303	-0.184
Africa	NA	NA	NA
Funding status seed	0.654	0.505	-0.336
Funding type pre-seed	-1.274***	-1.270***	-1.224***
Funding type seed	NA	NA	NA
Funding type series A	NA	NA	NA
Found in 2020	-0.233	-0.373	-0.323
Found in 2021	-0.197	-0.280	-0.343
Found in 2022	-0.108	-0.272	-0.048
Found in 2023	NA	NA	NA
Intercept	12.896***	13.159***	10.808’
Inverse Mill’s Ratio	1.547*	1.556*	1.82*
Observations	*57*	*57*	*57*

***p < 0.001 **p < 0.01 *p < 0.05 ‘p < 0.1

R^2: 1) 0.3163. 2) 0.3203. 3) 0.3174.

Table 6 – Relationship between AI expert characteristics, Industry Legitimacy and Days Until VC funding

DV: Days until funding	*Model 1 Controls Only*	*Model 2 Main Effects*	*Model 3 Final Model*
Industry Legitimacy			
AI expert education x industry legitimacy			-0.436
AI expert experience x industry legitimacy			-0.291
AI expert industry experience x industry legitimacy			0.799
AI expert startup experience x industry legitimacy			-0.424
Main Relationships			
AI expert education		0.449	3.652
AI expert experience		-0.484	1.416
AI expert industry experience		-0.197	-5.859
AI expert startup experience		0.460’	3.619
Control Variables			
CTO presence	-	-	-
Number of founders	-0.005	0.044	0.070
Number of employees	0.003	0.004	0.005
Industry legitimacy	0.103	0.085	0.015
Europe	-0.639	-0.680	-1.446
Americas	-0.629	-0.896	-1.671
Oceania	-1.444	-1.245	-1.898
Asia	-0.738	-0.797	-1.611
Africa	NA	NA	NA
Funding status seed	-0.756	-1.122	-2.168*
Funding type pre-seed	-0.107	-0.084	-0.078
Funding type seed	NA	NA	NA
Funding type series A	NA	NA	NA
Found in 2020	1.474’	1.388’	1.333’
Found in 2021	1.290’	1.195’	1.150
Found in 2022	0.893	0.468	0.488
Found in 2023	NA	NA	NA
Intercept	5.489**	5.924**	7.88’
Inverse Mill’s Ratio	-0.134	0.173	0.398
Observations	*57*	*57*	*57*

***p < 0.001 **p < 0.01 *p < 0.05 ‘p < 0.1

R^2: 1) 0.3163. 2) 0.3171. 3) 0.3123.

Table 7 – Importance Matrix Total Amount of Funding

Variable	*Importance*
Number of Employees	0.322
FT_Pre_Seed	0.261
Industry_Legitimacy	0.167
Number_of_Founders	0.043
Funding_Status_Seed	0.041
Continent_Africa	0.039
FT_SeriesA	0.035
Continent_Europe	0.021
Continent_Americas	0.018
AI_Specialist_Presence	0.018
Continent_Asia	0.017
Founded_Year_2021	0.010
Continent_Oceania	0.010

Table 8 – Importance Matrix Days Until Funding

Variable	*Importance*
Founded_Year_2020	0.248
Number_of_Founders	0.186
Number_of_Employees	0.129
Industry_Legitimacy	0.093
Founded_Year_2023	0.088
Founded_Year_2022	0.085
Continent_Europe	0.071
Continent_Americas	0.036
Founded_Year_2021	0.030
Continent_Oceania	0.024
AI_Specialist_Presence	0.002

6.3 Formulas

Formula 1 – Probit Model

$$P(AISpecialistPresence = 1)$$
$$= \frac{1}{1 + e^{-(\beta_0 + \beta_1 CTOPresence + \beta_2 ContinentEurope + \beta_3 FoundedYear2022 + \beta_4 NumberOfFounders + \beta_5 NumberOfEmployees}}$$

$$myProbit = glm(AISpecialistPresence \sim CTOPresence + ContinentEurope + FoundedYear2022 + NumberOfFounders + NumberOfEmployees, family = \text{binomial}\ (link = probit), data = dfregression)$$

$$dfregression\$IMR < -invMillsRatio(myProbit)\$IMR1$$

Formula 2 – Model 1 Controls Only (Funding Amount)

$$Log(Funding\ Amount)_i = \beta_0 + \beta_1 NumberOfFounders_i + \beta_2 NumberOfEmployees_i + \beta_3 IndustryLegitimacy_i + \beta_4 ContinentEurope_i + \beta_5 ContinentAmericas_i + \beta_6 ContinentOceania_i + \beta_7 ContinentAsia_i + \beta_8 ContinentAfrica_i + \beta_9 FundingStatusSeed_i + \beta_{10} FTPreSeed_i + \beta_{11} FTSeed_i + \beta_{12} SeriesA_i + \beta_{13} FoundedYear2020_i + \beta_{14} FoundedYear2021_i + \beta_{15} FoundedYear2022_i + \beta_{16} FoundedYear2023_i + \beta_{17} IMR_i + \epsilon_i$$

$$myHeckit1 = lm(\log(FundingAmount) \sim NumberOfFounders + NumberOfEmployees + IndustryLegitimacy + ContinentEurope + ContinentAmericas + ContinentOceania + ContinentAsia + ContinentAfrica + FundingStatusSeed + FTPreSeed + FTSeed + FTSeriesA + FoundedYear2020 + FoundedYear2021 + FoundedYear2022 + FoundedYear2023 + IMR, data = dfregression[dfregression\$AISpecialistPresence == 1,])$$

Formula 3 – Model 1 Controls Only (Days Until Funding)

$$Log(Date\ Difference)_i = \beta_0 + \beta_1 NumberOfFounders_i + \beta_2 NumberOfEmployees_i + \beta_3 IndustryLegitimacy_i + \beta_4 ContinentEurope_i + \beta_5 ContinentAmericas_i + \beta_6 ContinentOceania_i + \beta_7 ContinentAsia_i + \beta_8 ContinentAfrica_i + \beta_9 FundingStatusSeed_i + \beta_{10} FTPreSeed_i + \beta_{11} FTSeed_i + \beta_{12} SeriesA_i + \beta_{13} FoundedYear2020_i + \beta_{14} FoundedYear2021_i + \beta_{15} FoundedYear2022_i + \beta_{16} FoundedYear2023_i + \beta_{17} IMR_i + \in_i$$

$$\begin{aligned} myHeckit2 = lm(&\log(DateDifference) \sim NumberOfFounders \\ &+ NumberOfEmployees + IndustryLegitimacy + ContinentEurope \\ &+ ContinentAmericas + ContinentOceania + ContinentAsia \\ &+ ContinentAfrica + FundingStatusSeed + FTPreSeed + FTSeed \\ &+ FTSeriesA + FoundedYear2020 + FoundedYear2021 \\ &+ FoundedYear2022 + FoundedYear2023 + IMR, data \\ &= dfregression[dfregression\$AISpecialistPresence == 1,]) \end{aligned}$$

Formula 4 – Model 2 Main Effects (Funding Amount)

$$Log(Funding\ Amount)_i = \beta_0 + \beta_1 NumberOfFounders_i + \beta_2 NumberOfEmployees_i + \beta_3 IndustryLegitimacy_i + \beta_4 ContinentEurope_i + \beta_5 ContinentAmericas_i + \beta_6 ContinentOceania_i + \beta_7 ContinentAsia_i + \beta_8 ContinentAfrica_i + \beta_9 FundingStatusSeed_i + \beta_{10} FTPreSeed_i + \beta_{11} FTSeed_i + \beta_{12} SeriesA_i + \beta_{13} FoundedYear2020_i + \beta_{14} FoundedYear2021_i + \beta_{15} FoundedYear2022_i + \beta_{16} FoundedYear2023_i + \beta_{17} AISpecialistEducation_i + \beta_{18} AISpecialistExperience_i + \beta_{19} AISpecialistIndustryExperience_i + \beta_{20} AISpecialistStartupExperience_i + \beta_{21} IMR_i + \in_i$$

$$myHeckit3 = lm(\log(FundingAmount) \sim NumberOfFounders \\ + NumberOfEmployees + IndustryLegitimacy + ContinentEurope \\ + ContinentAmericas + ContinentOceania + ContinentAsia \\ + ContinentAfrica + FundingStatusSeed + FTPreSeed + FTSeed \\ + FTSeriesA + FoundedYear2020 + FoundedYear2021 \\ + FoundedYear2022 + FoundedYear2023 + AISpecialistEducation \\ + AISpecialistExperience + AISpecialistIndustryExperience \\ + AISpecialistStartUpExperience + \ IMR, data \\ = dfregression[dfregression\$AISpecialistPresence == 1,])$$

Formula 5 – Model 2 Main Effects (Days Until Funding)

$$Log(Date\ Difference)_i = \ \beta_0 + \ \beta_1 NumberOfFounders_i + \\ \beta_2 NumberOfEmployees_i + \ \beta_3 IndustryLegitimacy_i + \ \beta_4 ContinentEurope_i + \\ \beta_5 ContinentAmericas_i + \ \beta_6 ContinentOceania_i + \ \beta_7 ContinentAsia_i + \\ \beta_8 ContinentAfrica_i + \ \beta_9 FundingStatusSeed_i + \ \beta_{10} FTPreSeed_i + \ \beta_{11} FTSeed_i + \\ \beta_{12} SeriesA_i + \ \beta_{13} FoundedYear2020_i + \ \beta_{14} FoundedYear2021_i \ + \\ \beta_{15} FoundedYear2022_i + \ \beta_{16} FoundedYear2023_i \ + \ \beta_{17} AISpecialistEducation_i \ + \\ \beta_{18} AISpecialistExperience_i + \ \beta_{19} AISpecialistIndustryExperience_i + \\ \beta_{20} AISpecialistStartupExperience_i + \ \beta_{21} IMR_i + \in_i$$

$$myHeckit4 = lm(\log(DateDifference) \sim NumberOfFounders \\ + NumberOfEmployees + IndustryLegitimacy + ContinentEurope \\ + ContinentAmericas + ContinentOceania + ContinentAsia \\ + ContinentAfrica + FundingStatusSeed + FTPreSeed + FTSeed \\ + FTSeriesA + FoundedYear2020 + FoundedYear2021 \\ + FoundedYear2022 + FoundedYear2023 + AISpecialistEducation \\ + AISpecialistExperience + AISpecialistIndustryExperience \\ + AISpecialistStartUpExperience + \ IMR, data \\ = dfregression[dfregression\$AISpecialistPresence == 1,])$$

Formula 6 – Model 3 Final (Funding Amount)

$$Log(Funding\ Amount)_i = \beta_0 + \beta_1 NumberOfFounders_i + \beta_2 NumberOfEmployees_i + \beta_3 IndustryLegitimacy_i + \beta_4 ContinentEurope_i + \beta_5 ContinentAmericas_i + \beta_6 ContinentOceania_i + \beta_7 ContinentAsia_i + \beta_8 ContinentAfrica_i + \beta_9 FundingStatusSeed_i + \beta_{10} FTPreSeed_i + \beta_{11} FTSeed_i + \beta_{12} SeriesA_i + \beta_{13} FoundedYear2020_i + \beta_{14} FoundedYear2021_i + \beta_{15} FoundedYear2022_i + \beta_{16} FoundedYear2023_i + \beta_{17} AISpecialistEducation_i + \beta_{18} AISpecialistExperience_i + \beta_{19} AISpecialistIndustryExperience_i + \beta_{20} AISpecialistStartupExperience_i + \beta_{21} IndustryLegitimacy_i \times AISpecialistEducation_i + \beta_{22} IndustryLegitimacy_i \times AISpecialistExperience_i + + \beta_{23} IndustryLegitimacy_i \times AISpecialistIndustryExperience_i + \beta_{24} IndustryLegitimacy_i \times AISpecialistStartupExperience_i + \beta_{25} IMR_i + \in_i$$

$$\begin{aligned} myHeckit5 = lm(\log(&FundingAmount) \sim NumberOfFounders \\ &+ NumberOfEmployees + IndustryLegitimacy + ContinentEurope \\ &+ ContinentAmericas + ContinentOceania + ContinentAsia \\ &+ ContinentAfrica + FundingStatusSeed + FTPreSeed + FTSeed \\ &+ FTSeriesA + FoundedYear2020 + FoundedYear2021 \\ &+ FoundedYear2022 + FoundedYear2023 + AISpecialistEducation \\ &+ AISpecialistExperience + AISpecialistIndustryExperience \\ &+ AISpecialistStartUpExperience + IndustryLegitimacy \\ &* AISpecialistEducation + IndustryLegitimacy \\ &* AISpecialistExperience + IndustryLegitimacy \\ &* AISpecialistIndustryExperience + IndustryLegitimacy \\ &* AIStartUpExperience + IMR, data \\ &= dfregression[dfregression\$AISpecialistPresence == 1,]) \end{aligned}$$

Formula 7 – Model 3 Final (Days Until Funding)

$$Log(Date\ Difference)_i = \beta_0 + \beta_1 NumberOfFounders_i + \beta_2 NumberOfEmployees_i + \beta_3 IndustryLegitimacy_i + \beta_4 ContinentEurope_i + \beta_5 ContinentAmericas_i + \beta_6 ContinentOceania_i + \beta_7 ContinentAsia_i + \beta_8 ContinentAfrica_i + \beta_9 FundingStatusSeed_i + \beta_{10} FTPreSeed_i + \beta_{11} FTSeed_i + \beta_{12} SeriesA_i + \beta_{13} FoundedYear2020_i + \beta_{14} FoundedYear2021_i + \beta_{15} FoundedYear2022_i + \beta_{16} FoundedYear2023_i + \beta_{17} AISpecialistEducation_i + \beta_{18} AISpecialistExperience_i + \beta_{19} AISpecialistIndustryExperience_i + \beta_{20} AISpecialistStartupExperience_i + \beta_{21} IndustryLegitimacy_i \times AISpecialistEducation_i + \beta_{22} IndustryLegitimacy_i \times AISpecialistExperience_i + + \beta_{23} IndustryLegitimacy_i \times AISpecialistIndustryExperience_i + \beta_{24} IndustryLegitimacy_i \times AISpecialistStartupExperience_i + \beta_{25} IMR_i + \in_i$$

$$\begin{aligned} myHeckit5 = lm(\log(&DateDifference) \sim NumberOfFounders \\ &+ NumberOfEmployees + IndustryLegitimacy + ContinentEurope \\ &+ ContinentAmericas + ContinentOceania + ContinentAsia \\ &+ ContinentAfrica + FundingStatusSeed + FTPreSeed + FTSeed \\ &+ FTSeriesA + FoundedYear2020 + FoundedYear2021 \\ &+ FoundedYear2022 + FoundedYear2023 + AISpecialistEducation \\ &+ AISpecialistExperience + AISpecialistIndustryExperience \\ &+ AISpecialistStartUpExperience + IndustryLegitimacy \\ &* AISpecialistEducation + IndustryLegitimacy \\ &* AISpecialistExperience + IndustryLegitimacy \\ &* AISpecialistIndustryExperience + IndustryLegitimacy \\ &* AIStartUpExperience + IMR, data \\ &= dfregression[dfregression\$AISpecialistPresence == 1,]) \end{aligned}$$

Formula 8 – XGBoosting (Founding Amount)

```
# make this example reproducible
set.seed(0)

# split into training (70%) and testing set (30%)
parts =
createDataPartition(df_xg_dd1$`Last_Funding_Amount_Currency_(USD)`, p =
.7, list = F)
train = df_xg_dd1[parts, ]
test = df_xg_dd1[-parts, ]

# define predictor and response variables in training set
train_x = data.matrix(train[, -1])
train_y = train[,1]

# define predictor and response variables in testing set
test_x = data.matrix(test[, -1])
test_y = test[, 1]

# define final training and testing sets
xgb_train = xgb.DMatrix(data = train_x, label = train_y)
xgb_test = xgb.DMatrix(data = test_x, label = test_y)

# define watchlist
watchlist = list(train=xgb_train, test=xgb_test)

# fit XGBoost model and display training and testing data at each round
model = xgb.train(data = xgb_train, max.depth = 3, watchlist=watchlist, nrounds
= 70)

# define final model
final = xgboost(data = xgb_train, max.depth = 3, nrounds = 17, verbose = 0)

# Obtain predictions for the training set
train_pred <- predict(final, xgb_train)

# Obtain predictions for the testing set
test_pred <- predict(final, xgb_test)

mean((test_y - test_pred)^2) #mse
caret::MAE(test_y, test_pred) #mae
caret::RMSE(test_y, test_pred) #rmse

# Variable importance plot
importance_matrix <- xgb.importance(colnames(xgb_train), model = final)
xgb.plot.importance(importance_matrix, xlab = "Feature Importance", ylab =
"Features")
```

Formula 9 - XGBoosting (Days Until Funding)

```
# make this example reproducible
set.seed(0)

# split into training (70%) and testing set (30%)
parts = createDataPartition(df_xg_dd$Date_Difference, p = .7, list = F)
train = df_xg_dd[parts, ]
test = df_xg_dd[-parts, ]

# define predictor and response variables in training set
train_x = data.matrix(train[, -4])
train_y = train[,4]

# define predictor and response variables in testing set
test_x = data.matrix(test[, -4])
test_y = test[, 4]

# define final training and testing sets
xgb_train = xgb.DMatrix(data = train_x, label = train_y)
xgb_test = xgb.DMatrix(data = test_x, label = test_y)

# define watchlist
watchlist = list(train=xgb_train, test=xgb_test)

# fit XGBoost model and display training and testing data at each round
model = xgb.train(data = xgb_train, max.depth = 3, watchlist=watchlist, nrounds
= 70)

# define final model
final = xgboost(data = xgb_train, max.depth = 3, nrounds = 12, verbose = 0)

# Obtain predictions for the training set
train_pred <- predict(final, xgb_train)

# Obtain predictions for the testing set
test_pred <- predict(final, xgb_test)

mean((test_y - test_pred)^2) #mse
caret::MAE(test_y, test_pred) #mae
caret::RMSE(test_y, test_pred) #rmse

# Variable importance plot
importance_matrix <- xgb.importance(colnames(xgb_train), model = final)
xgb.plot.importance(importance_matrix, xlab = "Feature Importance", ylab =
"Features")
```

7 References

Aldrich, H. (1999). Organizations Evolving. *Social Forces*, 79(4), 1521-1523.

Aldrich, H., & Fiol, C. M. (1994). Fools Rush in? The Institutional Context of Industry Creation. *Academy of Management Review*, 19, 645-670.

Boyd, E. D., Chandy, R., & Cunha, M. (2010). When Do Chief Marketing Officers Affect Firm Value? A Customer Power Explanation. *Journal of Marketing Research*, 47, 1162-1176.

Brown, R., & Rocha, A. (2020). Entrepreneurial uncertainty during the Covid-19 crisis: Mapping the temporal dynamics of entrepreneurial finance. Journal of Business Venturing Insights.

Brush, C., Greene, P. G., & Hart, M. H. (2001). From Initial Idea to Unique Advantage: The Entrepreneurial Challenge of Constructing a Resource Base. Academy of Management Executive, 15(1), 64-78.

Carpenter, M. A., & Fredrickson, J. W. (2001). Top Management Teams, Global Strategic Posture, and the Moderating Role of Uncertainty. *Academy of Management Journal*, 44(3), 533-545.

Cemre, Ü., Ioana, C. (2019). A Machine Learning Approach Towards Startup Success Prediction. IRTG 1792 Discussion Papers 2019-022, Humboldt University of Berlin.

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Retrieved from arXiv:1603.02754v3.

Colombo, M. G., & Grilli, L. (2005). Founders' Human Capital and the Growth of New Technology-Based Firms: A Competence-Based View. *Research Policy*, 34(6), 795-816.

Dellermann, D., Lipusch, N., Ebel, P., Popp, K. M., & Leimeister, J. M. (2017). Finding the unicorn: Predicting early stage startup success through a hybrid intelligence method.

Delmar, F., & Shane, S. (2004). Legitimating First: Organizing Activities and the Survival of New Ventures. *Journal of Business Venturing*, 19(3), 385-410.

Eesley, C. E., Hsu, D. H., & Roberts, E. B. (2014). The contingent effects of top management teams on venture performance: Aligning founding team composition with innovation strategy and commercialization environment. *Strategic management journal*, 35(12), 1789-1817.

Elith, Jl., Leathwick., J. R., & Hastie, T., (2008). A working guide to boosted regression trees. Journal of Animal Ecography, 77, 802-813.

Finkelstein, S. (1992). Power in Top Management Teams: Dimensions, Measurement, and Validation. *Academy of Management Journal*, 35(3), 505-538.

Forbes. (2023). Generative AI Will Continue To Accelerate In 2023: Are You Ready?. Retrieved from https://www.forbes.com/sites/cindygordon/2022/12/29/generative-ai-will-continue-to-accelerate-in-2023-are-you-ready/?sh=162972e3650c.

Friedman, J. (2001). Greedy function approximation: A gradient boosting machine. Annals of Statistics, 29 (5), 1189-1232.

Galanos, V. (2018). Exploring expanding expertise: artificial intelligence as an existential threat and the role of prestigious commentators, 2014–2018. Technology Analysis & Strategic Management. https://doi.org/10.1080/09537325.2018.1518521.

Gorman, M., & Sahlman, W. A. (1989). What do venture capitalists do?. Journal of Business Venturing, 4(4), 231-248.

Güngör, H. (2020). Creating Value with Artificial Intelligence: A Multi stakeholder Perspective. Journal of Creating Value. https://doi.org/10.1177/2394964320921071.

He, T. (2016). An Introduction to XGBoost R Package. Retrieved from http://dmlc.ml/rstats/2016/03/10/XGBoost.html.

Hitt, M. A., Ireland, D., Camp, M., & Sexton, D. (2001). Strategic Entrepreneurship; Entrepreneurial Strategies for Wealth Creation. Strategic Management Journal, 22(6/7), 479-491.

Homburg, C., Hahn, A., & Sandner, P. (2014). The Role of Chief Marketing Officers for Venture Capital Funding: Endowing New Ventures with Marketing Legitimacy. *Journal of Public Policy & Marketing*, 51(5).

Hsu, D. H. (2007). Experienced Entrepreneurial Founders, Organizational Capital, and Venture Capital Funding. Research Policy, 36(5), 722-741.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2017). An introduction to statistical learning (7th ed.) New York: Springer.

Kirsch, D., Goldfarb, & Gera, A. (2009). Form or Substance: The Role of Business Plans in Venture Capital Decision Making. *Strategic Mangement Journal*, 30(5), 487-515.

KPMG. (2021). Global venture capital annual investment shtaaters records following antoher healthy quarter. Retrieved from https://kpmg.com/xx/en/home/media/press-releases/2022/01/global-venture-capital-annual-investment-shatters-records-following-another-healthy-quarter.html#:~:text=Key%202021%20Annual%20Highlights,all%20regions%20of%20the%20world.

Lam, S. (2000). Tacit Knowledge, Organizational Learning and Societal Institutions: An Integrated Framework. Organization Studies, 21(3), 487-513.

Lee, J., Suh, T., Roy, D., & Baucus, M. (2019). Emerging Technology and Business Model Innovation: The Case of Artificial Intelligence. Journal of Open Innovation: Technology, Market, and Complexity. https://doi.org/10.3390/JOITMC5030044.

MacDonald, R. J. (1985). Strategic Alternatives in Emerging Industries. *Journal of Product Innovation Management*, 2(3), 158-169.

Morrissey, K., Kinderman, P., Pontin, E., Tai, S., & Schwannauer, M. (2016). Web based health surves: using a Two Step Heckman model to examine their potential for population health analysis. Social Science & Medicine, 45-63.

Mou, X. (2019). Artificial Intelligence: Investment Trends and Selected Industry Uses.

O'Brien, R. (2007). A Caution Regarding Rules of Thumb for Variance Inflation Factors. Quality & Quantity, 41, 673-690. https://doi.org/10.1007/S11135-006-9018-6.

Rao, R. S., Chandy, R. K., & Prabhu, J. C. (2008). The Fruits of Legitimacy: Why Some New Ventures Gain more from Innovation than Others. Journal of Marketing, 72(4).

Ross, G., Das, S., Sciro, D., and Raza, H. (2021). Capitalvx: A machine learning model for startup selection and exit prediction. *The Journal of Finance and Data Science*, 94–114.

Ruiz, A. What does a Chief AI Officer do?. Retrieved from https://www.nocode.ai/what-does-a-chief-ai-officer-do/#:~:text=A%20Chief%20AI%20Officer%20(CAIO,AI)%20initiatives%20within%20an%20organization.

Schwarzkopf, J., Lévesque, M., & Maxwell, A. (2010). How entrepreneurs-in-residence increase seed investment rates. Venture Capital. https://doi.org/10.1080/13691060903435783.

Scott, R. W. (1994). Institutional Analysis: Variance and Process Theory Approaches. Instituional Environments and Organizations: Structural Complexity and Individualism. *Sage Publications*, 81-99.

Sinaga, K. P., & M. Yang. (2020). Unsupervised K-Means Clustering Algorithm. IEEE Access, 8, 80716-80727.

Suchman, M. (1995). Managing Legitimacy: Strategic and Institutional Approaches. *Academy of Mangement Review*, 20, 571-611.

Taplin, R. (2016). Research involving limited dependent variables: issues in the literature and recommendations for improvement. Quality & Quantity.

Zacharakis, A., McMullen, J., & Shepherd, D. (2007). Venture Capitalists' Decision Policies Across Three Countries: An Institutional Theory Perspective. *Journal of International Business Studies*, 38(5), 691-708.

Zacharakis, A., & Meyer, D. (1998). A lack of insight: do venture capitalists really understand their own decision process?. *Journal of Business Venturing*, 13(1), 57-76.

Zimmerman, M. A., & Zeitz, G. J. (2002). Beyond Survival: Achieving New Venture Growth by building Legitimacy. *Academy of Management Review*, 27, 414-431.

Greene, W. H. (2003) *Econometric Analysis, Fifth Edition*, Prentice Hall.