The Future of Venture Capital Decision Making: The Impact of Quantitative Sourcing and Machine Learning on the VC Investment Process

Philip Schroepel

Dissertation written under the supervision of professor Jose Vasconcelos-Sousa

Dissertation submitted in partial fulfilment of requirements for the MSc in Management with Specialization in Strategy & Entrepreneurship, at the Universidade Católica Portuguesa, 05.01.2022.

Abstract

English

Investing in early-stage startups is a difficult endeavor. Venture Capitalists use heuristics and base their decisions on past experiences, which can lead to biases. Recently, Venture Capitalists are increasingly using artificial intelligence and quantitative sourcing to support their investment process, while others still rely on traditional investment mechanisms. This research investigates the usage and impact of artificial intelligence and machine learning throughout the venture investment cycle to make investment decisions. This dissertation is an exploratory study that employs a qualitative research approach in the form of semi-structured interviews with ten European Venture Capitalists. The results show that Venture Capitalists utilize machine learning and web scraper tools, particularly during the deal origination, firm-specific screening, and general screening stages of the investment process, to solve the identification and selection challenges. As a result, investment processes become more efficient and less biased, allowing for more time to be spent advising and mentoring portfolio startups. It adds to the existing literature on how artificial intelligence and data can augment existing investment mechanisms during the venture capital decision-making process.

Keywords: Venture Capital, Quantitative Sourcing, Artificial Intelligence, Machine Learning, Investment Process, Data-Driven Decision-Making

Title: The Future of Venture Capital Decision Making: The Impact of Quantitative Sourcing and Machine Learning on the VC Investment Process

Author: Philip Schroepel

Portuguese

Investir em startups na sua fase inicial exige um elevado empenho. Os investidores de capital de risco baseiam as suas decisões em pesquisa e experiências passadas, o que pode levar a enviesamentos. Embora muitos investidores de capital de risco ainda utilizem mecanismos de investimento tradicionais, tem havido um aumento na utilização de inteligência artificial e *sourcing* quantitativo para apoiar o processo de investimento. Esta investigação estuda a utilização e impacto da inteligência artificial e de *machine learning* ao longo do ciclo de investimento de risco para tomar decisões de investimento. Esta dissertação é um estudo empírico que utiliza uma abordagem de investigação qualitativa sob a forma de entrevistas semi-estruturadas com dez empresas de investimento de capital de risco europeias. Os resultados mostram que os investidores de capital de risco utilizam *machine learning* e ferramentas de recolha de dados na web, em particular durante o início da oportunidade de negócio, a seleção específica da empresa, e fases gerais de análise do processo de investimento, para resolver os desafios de identificação e seleção. Consequentemente, os processos de investimento tornam-se mais eficientes e menos tendenciosos, permitindo que se utilize mais tempo a aconselhar e a orientar as empresas do portfolio. Este estudo complementa a literatura existente relativamente a como a inteligência artificial e os dados podem elevar os mecanismos de investimento existentes durante o processo de tomada de decisão de capital de risco.

Palavras-chave: Capital de Risco, Fontes Quantitativas, Inteligência Artificial, Aprendizagem de Máquinas, *Machine Learning*, Processo de Investimento, Tomada de Decisões Baseada em Dados

Título: O Futuro da Tomada de Decisão do Capital de Risco: Impacto de Sourcing Quantitativo e de Machine Learning no Processo de Investimento

Autor: Philip Schroepel

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

AI	Artificial Intelligence
CVC	Corporate Venture Capital
DL	Deep Learning Models
DT	Decision Trees
GTB	Gradient Tree Boosting
LP	Limited Partner
ML	Machine Learning
PE	Private Equity
SC	Selection Criteria
VC	Venture Capital
VCs	Venture Capitalists

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

1 INTRODUCTION

1.1 Problem Statement and Research Objectives

Venture Capital (VC) is considered an important instrument to identify and finance high-potential early-stage companies. Companies that have been invested in make an important contribution to the global economy. Famous companies partially funded or invested in by Venture Capitalists (VCs) in the past 30 years include Facebook, Amazon, Google, and Cisco (Kaplan & Lerner, 2010).

It is now widely recognized from various studies that VC is an outlier business, which means that VCs are often subject to high failure rates due to the uncertainty of early stage investing, but the investments that do pay off reap substantial rewards. A missed investment in the next Facebook or Google, however, can determine whether a VC fund is among the best performers (Korteweg & Sorensen, 2010; Retterath & Kavadias, 2020). Thus, VC is regarded as a high-risk, high-return asset class. VCs spend considerable resources and time in their manual screening and decision-making process in order to find the most promising new venture and achieve high returns on investment. This decision-making process is not only resource-intensive but also prone to errors, biases, and heuristics (Kaplan & Lerner, 2010). Hence, numerous studies have been conducted around VC, especially on how VCs make decisions (Gompers et al., 2020), their investment decision criteria (Monika & Sharma, 2015) and new venture evaluation (Hall & Hofer, 1993).

Kaplan and Strömberg (2001) argue that VCs play an essential role in solving the principal-agent problem in financial contracting, with a new venture seeking capital and the VC as an investor providing capital to ventures. Their study also highlights that VCs focus on pre-investment screening and evaluation, contracting, and post-investment monitoring and advising to solve the principal-agent problem. Since 2001, a significant number of empirical studies have explored VC decision-making. Gompers et al. (2020) surveyed 681 VCs and provided detailed information on how VCs make decisions, along with pre-investment screening, structuring the investments, and post-investment monitoring and advising. Throughout the VC management stages and especially in pre-investment screening, VCs are exposed to several problems. The manual investment screening process includes a high degree of information, time pressure in decision-making, uncertainty, and biases. Furthermore, VCs are unable to handle the large number of potential investments that must be screened in such a short period of time (Franke et al., 2006; Zacharakis & Meyer, 1998). As a result, VCs are constantly facing the risk of missing out on promising investment opportunities.

In the current market dynamics, competition among VC firms is increasing. While research suggests that the number of startups in the market is steady, the number of VC funds is increasing while also having more assets under management. (Patel, 2020; Stangler & Kedrosky, 2010). As a result, an increasing number of VC funds are competing for a limited amount of venture investment opportunities. Consequently, company valuations are increasing, requiring VCs to adapt their screening procedures and investment strategies and find innovative approaches to become the first to identify a promising investment opportunity.

In the digital era, many companies are required to respond quickly to the changing competitive landscape. Especially against this background, the business world is using new digital technologies to gain a competitive advantage (Venkatraman, 2017). As one of the technologies, Artificial Intelligence (AI) has taken on a unique role in the competitive environment. As noted by Davenport (2018) and Brynjolfsson & Mcafee (2017), AI, particularly Machine Learning (ML), is considered one of the technologies with the strongest disruption potential of our era (Brynjolfsson & Mcafee, 2017; Davenport, 2018). In addition, a growing body of literature recognizes the importance of data and AI in supporting automated decision-making in an organizational context (Borges et al., 2021). In contrast, the opportunity of AI in the decision-making process of VC is still empirically unexplored.

The overall aim of this research is to provide an understanding of VCs' usage and impact of AI and more specific technologies like ML in their venture investment process to support or automate their investment decisions. This dissertation does not only point out how VCs are using AI but will also provide an insight on the rationale behind the usage. Additionally, it identifies implementation and usage challenges, as well as reasons for reluctance to use it. Furthermore, it adds to existing literature on how data and technology can augment existing investment mechanisms during the decision-making process of VCs. However, what remains unclear is how VCs are using advanced technologies like ML during their investment process, how those technologies affect their decision-making, and what economic effects and value these technologies create. Finally, this research benefits VC firms looking to adopt AI/ML as well as quantitative sourcing approaches in their investment management.

Therefore, the guiding research question for this dissertation is:

How does Artificial Intelligence and Machine Learning impact the Venture Capital investment process?

Taking into account the above-mentioned considerations, the following five research questions emerge to address the main research objective.

Table 1 Research Questions

ID	Research Question
RQ1	What are the reasons for VCs to use quantitative sourcing tools or ML in their investment process?
RQ2	How do VCs utilize AI/ML and data in their venture investment process?
RQ3	How can technology and data within the context of digitization enhance existing investment mechanisms during the decision-making process of VCs?
RQ4	What are the challenges when using quantitative sourcing and AI/ML tools?
RQ5	Where can AI contribute to the decision-making process in the future?

1.2 Research Overview

The methodological approach taken in this study is a qualitative research design to gather primary data. For this purpose, the dissertation conducts qualitative research in the form of interviews that are evaluated with the use of the *qualitative content analysis*. Subsequently, the analysis is triangulated with secondary data to be able to present comprehensive, saturated, and reliable results. By employing qualitative modes of inquiry in the form of semi-structured interviews with VCs, the research questions can be answered.

1.3 Outline

The structure of the research work is divided into five chapters. In order to present the topic in a holistic manner and to clarify its relevance, it is necessary to define the subject matter. After the introduction, the second chapter comprises a theoretical framework. This contains theoretical basics on relevant issues and a delimitation of the central terms of AI. In chapter three, the approaches to qualitative research are presented to justify the decision on the methodological approach. Subsequently, the research concept of qualitative content analysis according to Mayring (2015) is explained in more detail. In chapter four, the results of this analysis are then evaluated and discussed. Finally, the findings of the dissertation are summarized, classified, and reflected upon. An outlook on future developments and further research possibilities on the topic rounds off the work with the final consideration in chapter five.

2 LITERATURE REVIEW

2.1 Venture Capital

2.1.1 Venture Capital Industry and Investment Cycle

VC can be described as an asset class and form of financing as part of private equity (PE) that invests in early-stage companies - also referred to as start-ups - with growth potential. In return for the invested capital, VCs receive an equity stake in the company (Zider, 1998). The VC's goal is to increase the value of the company so that the investment generates a return for the investor at the time of exit. Company exits can either be acquisitions by other companies or via an initial public offering (IPO). However, VCs do not only contribute monetary value, but they also provide value-added services, 'smart capital' in the form of strategic support, management expertise, or technical expertise (Gompers & Lerner, 2004). VC investors manage the assets of a VC fund. For this purpose, the capital is invested in the fund by so-called limited partners. Investors are typically institutions such as banks, insurance funds, or large corporations, and they do not have voting rights in the fund's management. Given that investments in young, growth-oriented companies represent a high risk, VCs diversify their capital into a broad portfolio of start-ups. Investors on the other hand, seek above market returns (Zider, 1998). Fundamentally, VC operates in a cyclical process, the 'venture capital cycle,' as it is referred to in the literature. It all starts with generating a fund, which is then invested in new initiatives, which are then supported by the VC in terms of value-adding services and monitoring to increase the company's value and drive growth (Gompers & Lerner, 2004). As companies grow, they go through various stages of venture financing, with VCs also focusing on specific stages in their investment thesis (Tian, 2011). These stages can be divided into seed, early stage, and late stage. It then continues as successful companies exit, providing a return on investment to the fund and its limited partners (LPs). The cycle is virtuous when new capital is reinvested in the establishment of a new fund (Gompers & Lerner, 2004).

2.1.2 Venture Capital Investment Process

To set the stage of the underlying study, the VC investment process is described in this chapter. The following section also outlines existing hurdles in the investment process that prevent many VCs from further scaling their operations. Decision-making can be thought of as a process that is executed cognitively to make a choice from multiple options, resulting in a final decision. Decision-making theory distinguishes between decisions under risk and decisions under uncertainty (Knight, 1921). As VC firms are confronted with both of these challenges, the decision-making process is considered complex (Kaplan & Lerner, 2010; Kollmann & Kuckertz, 2010).

To reduce the possibility of passing on promising investment opportunities and adverse selection, VCs employ a multi-stage sequential investment process (Hall & Hofer, 1993). Previous studies have explored the model of the VC decision-making process, and those include a consensus on initial stages. However, there is no clear definition of the overall process (Fried & Hisrich, 1994; Hall & Hofer, 1993; Tyebjee & Bruno, 1984; Wells, 1974). In summary, it can be stated that activities take place in three areas: pre-investment, structuring of investments, and post-investment activities (Gompers et al., 2020). Table 2 summarizes the most established models of investment processes and the description of each respective decision-making process stage.

Table 2 Stages of the VC Investment Process

Stages	**Wells (1974)**	**Tyebjee / Bruno (1984)**	**Hall (1989)**	**Boocock / Woods (1997)**	**Fried / Hisrich (1994)**
1.	Search	Deal origination	Generating deal flow	Generating deal flow	Deal origination
2.	Screen	Screening	Proposal screening	Initial screening	Firm-specific screen
3.	n/a	n/a	Proposal assessment	First meeting	Generic screen
4.	Evaluation	Evaluation	Project evaluation	Second meeting	First-phase evaluation
5.	n/a	n/a	n/a	Board presentation	n/a
6.	n/a	n/a	Due diligence	Due diligence	Second-phase evaluation
7.	n/a	Deal structuring	Deal structuring	Deal structuring	Closing
8.	Operations	Post-investment activities	Venture operations	Monitoring of investments	n/a
9.	Cashing out	n/a	Cashing out	Cashing out	n/a

Note. Taken and modified from Boocock & Woods, (1997) and Hall & Hofer (1993)

For the present study, therefore, the model of Fried & Hisrich (1994) is used, as it is the most accepted model in academic literature. Examining the investment process and characteristics of the respective stages, as described below, it becomes clear that this process has hardly undergone any changes to date (Kollmann & Kuckertz, 2010). Stages one to three are considered essential levers for a VC's value creation. They account for up to 60% of the return of a VC fund (Sørensen, 2007), which is why this study focuses on these stages.

As figure 1 indicates, Fried & Hisrichs model consists of six sequential stages defined as follows: (1) *deal originating*, (2) VC *firm-specific screen*, (3) *generic screen*, (4) *first-phase evaluation*, (5) *second-phase evaluation,* and (6) *closing* (Fried & Hisrich, 1994). Details on the stages will be described below.

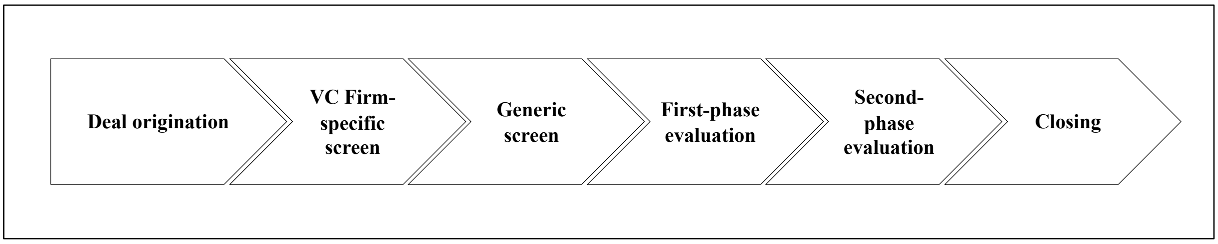

Figure 1 VC Investment Process according to Fried & Hisrich (1994)

Deal Origination

The first stage of the process is deal origination; typically, VCs receive the most significant number of proposals through referrals in their network (Fried & Hisrich, 1994). About 30% are referred to the VCs professional network, another 20% come from co-investors, and 8% are referred by a VCs portfolio (Gompers et al., 2020). Therefore, building a large and active network of people who refer investment proposals is particularly important for VCs (Fried & Hisrich, 1994). The importance of active deal sourcing is underscored by the fact that the company itself proactively generates as much as 30% of deal flow sourcing. Only 10% are initiated by VC's management (Gompers et al., 2020). Although this step of the investment process offers the opportunity for automation and understanding the market dynamics through the data availability in VC databases (Kaplan et al., 2002; Retterath, 2020), it is critical to identify and understand the risks of these data delivery platforms. Publicly sourced databases bear the risk of creating a noisy data, and analysis can result in contradictions for future investment decisions. These vulnerabilities are more prevalent with transaction-level data, implying caution when using publicly available data (Kaplan et al., 2002; Kaplan & Lerner, 2017). However, recent developments suggest that data availability and quality have steadily improved, and that this trend is projected to continue (Kaplan & Lerner, 2017).

VC Firm-specific screen

In the second process step, VCs screen proposals based on the firm-specific investment thesis. According to Fried & Hisrich (1994), this includes generic but hard selection criteria such as investment size, industry, investment stage, or geography. This step of the VC investment process requires limited resources and intellectual magnitude, and thus offers the possibility to be automated via ML or more specifically, supervised learning (Retterath, 2020).

Generic screen

After the firm-specific screening, the generic screen stage refers to the business plan analysis combined with relevant and existing know-how and insights of the VC. This comprises an initial analysis of the generic selection criteria (SC) of VCs, including but not limited to the business model, market characteristics and product characteristics. Those criteria also vary from VC to VC (Fried & Hisrich, 1994). VCs spend a significant amount of time and resources assessing and screening investment opportunities before concluding the deal and designing the term sheet (Kaplan & Strömberg, 2001). The purpose and outcome of the previous two steps is to reject start-ups with the least amount of time investment (Fried & Hisrich, 1994).

Since the investment process aims to find and invest in the uncertain environment, the process is also impacted by several challenges (Kaplan & Lerner, 2010; Sheehan & Sheehan, 2017). In a study conducted by Franke et al. (2006) it was shown that the decision-making process is subject to similarity bias. Similarity biases means that the more similar a VC investors and founders' professional and educational characteristics are, the higher the VC rates the startup. Additionally, availability bias is influencing VCs' decision-making, where they are more likely to weigh and consider information that is easily accessible, while overlooking less interesting information (Zacharakis & Meyer, 1998). Besides these biases, VCs are also subject to other psychological prejudices. Overconfidence of VCs suggests that investors, who typically have an excessive amount and intensity of information (information overload), overestimate their preferred outcome–the likelihood of a start-up's success. While it does not automatically have a negative effect on the decision-making process, it hinders both learning and improving the decision-making process (Zacharakis & Meyer, 2000; Zacharakis & Shepherd, 2001). Collectively, these studies outline that the VC investment process is resource-intensive, slow, and cost inefficient and deals with both incomplete information and information overload. Furthermore, it is prone to errors, biases, and heuristics (Corea, 2019a; Kaplan & Lerner, 2010; Sheehan & Sheehan, 2017). Investors use heuristics to identify startups in the investment process, thereby solving the identification problem. Although heuristics can be successful for investors, they introduce biases, produce invalid information, underestimate elements that are cognitively demanding, or disregard critical data (Åstebro & Elhedhli, 2006; Corea et al., 2021).

2.1.3 Venture Capital Selection Criteria

VC decision making along the investment process is subject to various selection criteria (SC). Researchers and practitioners constantly debate which are the most relevant SC in screening and selection. Prior empirical evidence suggests that VCs have different weights of their SC and differ in their view on investment selection (Gompers et al., 2020). Most recently, a broader perspective has been adopted by Gompers et al. (2020), who's survey of VCs asked for the important and most important factors a VC uses in their investment process. According to this study business model, product, market, industry, valuation, a VC's ability to add value, and the fit between the company and the VC are of the utmost importance among the investment factors. Scholars also identified cross-sectional differences in investment criteria between the stages of investing, for example qualitative factors such as management team being more important for VCs in early stage investing than in later stage investing (Gompers et al., 2020; Hall & Hofer, 1993; Petty & Gruber, 2011). Since SC are determined by humans and the individual thesis of the fund, they are dependent on individual investor views and opinions. Thus, screening outcomes can vary in early-stage investments, in case of involvement of multiple decision makers, as they focus more on the qualitative factors (Retterath, 2020).

Previous studies on VC selection criteria distinguish the SC into a variety of categories. This is due to different definitions of the individual criteria in the literature. For an overall understanding of this dissertation, the SC are classified into five main categories, presented next. This is in line with literature (Gompers et al., 2020; Hall & Hofer, 1993; Kaplan & Lerner, 2010; Petty & Gruber, 2011; Tyebjee & Bruno, 1984; Wells, 1974).

(1) General Company Characteristics: Company specific criteria are rather hard SC. They include stage of financing, industry focus and headquarter location of the potential investment company. As a result, this information allows an assessment of the match between a fund's investment thesis and its portfolio characteristics. It is agreed by scholars, that this is of particular importance in the early stages of investment (Hall & Hofer, 1993).

(2) Entrepreneur & Team Characteristics: This criterion summarizes all founders and team-related SC, including the experience and background of the founders, leadership capabilities, management skills and the personality characteristics of the team. Especially in early-stage investments, the founding team of a start-up is significant to VCs, since it is a strong signal about predicting potential success (Franke et al., 2006, 2008; Gompers et al., 2020).

(3) Funding and Financial Characteristics: The financial characteristics are a rather hard, funding- or shareholder-related SC. It includes factors such as the financial potential of the company, company valuation, the expected rate of return and expected risk, as well as the size of the investment (Petty & Gruber, 2011).

(4) Market Characteristics: This criterion comprises the economic environment of the proposed business, which includes market attractiveness, potential size of the market and market growth. Translated, it answers the question of the size of the problem that the team wants to solve with its product/service (Wells, 1974). Moreover, it is key to understand whether there are rival competitors in the market the company is going to operate (Tewari et al., 2020).

(5) Product/Service Characteristics: This factor is a solution-oriented criteria, besides the product strategy meaning thus it includes the question of a proprietary product to solve the main problem to be addressed. In addition, the product characteristics, the way the product is designed and the innovation, i.e., the technical edge of the product, are subject to the product characteristics (Tyebjee & Bruno, 1984).

Extending the previously mentioned SC, VC firms also consider the *Go-to-Market Strategy, Monetization* and *Technology* as criteria to predict the venture success (Tewari et al., 2020). In conclusion, the literature distinguishes criteria between "jockey vs. horse," with the former being tied to the team and the latter being related to the business model (Kaplan et al., 2009).

2.2 Data and Artificial Intelligence

2.2.1 Definition of Artificial Intelligence and Machine Learning

Since the introduction of the term AI in the 1950s, there is no agreed definition on what constitutes AI (Duan et al., 2019). However, according to a recent systematic literature review on AI, the definition of Russell & Norvig (1995) and its subsequent books were cited the most in academic research (Borges et al., 2021; Russell & Norvig, 1995). This is also in line with Loureiro et al. (2021) that provide a research agenda for AI in business (Loureiro et al., 2021). Therefore, this paper follows the definition of Russell & Norvig (1995), whereas AI is defined along two dimensions: (1) “reasoning-behavior dimension” and (2) “human performance-rationality dimension”. In those dimensions four sections describe AI, (1) AI that thinks like humans, (2) AI that acts like humans, (3) AI that thinks rationally, and lastly (4) AI that acts rationally (Russell & Norvig, 2016). AI systems, by definition, can replicate human cognitive capabilities like speaking, learning, solving complex problems, and decision making (Russell & Norvig, 2016).

As part of AI, ML is considered a sub-category (see Figure 2). ML is improving results based on more input data and its experience gained from execution. This process can be understood as learning. Thereby, ML is refining its own methods and actions (Collins et al., 2021; Marsland, 2014). According to a definition provided by Marsland (2014), ML consists of four different types. (1) Supervised learning, (2) Unsupervised learning, (3) Reinforced learning and (4) Evolutionary learning. Supervised learning is considered the most used type of ML today (Marsland, 2014). Due to the specific application of ML in VC decision making, this paper focuses on the first three types of ML that are described in detail below.

(1) Supervised learning: Correct sample data with right replies is compiled into a training set. Based on that, the ML algorithm is then **generalized** and reacts properly to new input. Thus, it can be described as **task driven**.

(2) Unsupervised learning: Instead of providing correct answers or sample data, the algorithm attempts to discover commonalities and correlations between unlabeled data inputs. Similar input is then afterwards **categorized**. Accordingly, it can be described as a **data-driven** approach to ML.

(3) Reinforced learning: The algorithm is informed whether its response is correct or incorrect, but not how to improve it. As a result, it must repeatedly test a variety of strategies to determine which ones work best. Thus, the algorithm learns according to the **trial-and-error** principle.

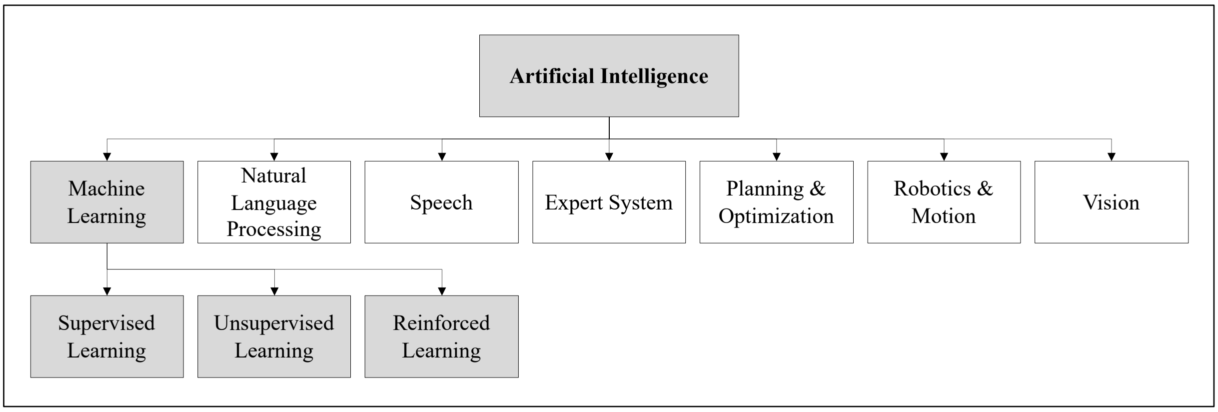

Figure 2 Classification of AI and ML according to (Deloitte, 2017) *and* (Marsland, 2014)

2.2.2 Artificial Intelligence and Data-Driven Decision Making

In today's world, organizations in the business world face increasing uncertainty in decision making. In addition, employees of the companies are exposed to a vast amount and a complex variety of data. Since the value of data in decision making became clear, data-driven decisions have been adopted by many companies (Brynjolfsson et al., 2011; Brynjolfsson & McElheran, 2016).

Due to increasing uncertainty within business organizations, support in the strategic decision making is a notable application of AI and ML (Haefner & Morf, 2021; Smeets et al., 2021; Trunk et al., 2020). Beyond the decision support, AI and ML can also support businesses in creating new innovative solutions, products, or services or to overhaul and improve internal processes. This could lead to an increase in revenue on the one side, or positively impact the costs on the other side (Haefner & Morf, 2021).

In decision making, AI and ML can be used in two ways: first, to provide machine support for the decision-making process, and second, to completely take over the decision making (Duan et al., 2019; Edwards et al., 2000). According to Edwards et al. (2000) depending on the level of decision-making, algorithmic decision-making has several strengths and weaknesses, they do however not distinguish between sub-categories of AI such as ML. The authors distinguish between strategic, tactical, and operational decisions, as well as between human support and human replacement. Additionally, they differentiate between structured and unstructured decisions. They conclude that AI has a particularly positive effect in replacing decision making when decisions are rather operationally and tactical. Likewise, they list that structured decision-making processes can be replaced by AI better than rather unstructured decision-making processes. They also point out that the strength of AI in supporting decisions lies primarily in unstructured and strategic decisions, which include decisions that incorporate external information and specific criteria that have their roots in long-term visions and ambitions (Edwards et al., 2000).

Even though some decisions made by AI completely disengage humans, this does not imply that mere automation is the goal of AI. The true value of artificial intelligence lies in its ability to make better decisions than a human could by itself. Humans that interact with the possibilities created by AI's data processing rather than the data itself can make better decisions. This is what is called hybrid or combined decision making (Colson, 2019).

AI and algorithm paradigms are however linked to adoption restraints and challenges. As biggest challenges for the application of AI in decision making, literature mentions trust, permission, the required evidence on correctness of AI, and the requirements and security to AI to completely act on the decision to be made (Dietvorst et al., 2015; Phillips-Wren, 2012). In the influential paper from Dietvorst et al. (2015), the authors argue that people rely more on human judgment than on that of an algorithm, especially after algorithms err. In contrast to Dietvorst et al., Logg (2018) argues that under certain scenarios people adhere more to the recommendations of a computational algorithm. In the literature, this disparity is referred to as "algorithm aversion" (Dietvorst et al., 2015) or "algorithm appreciation" (Logg et al., 2018).

2.2.3 Status Quo of AI and ML in Venture Capital

Despite more recent attention of business magazines, newspapers, and discussions in VC related forums about the usage of AI for investment decisions (Corea, 2019b; Council, 2021; Warnock, 2021; Wiggers, 2021), there is a relatively small body of academic literature that is concerned with AI, ML, and data-driven sourcing in VC decision making. Furthermore, VCs do not disclose details about specific ML use cases, the application of algorithms, and ML approaches. As one reason, Burgess (2018) mentions, companies keep information about the usage of AI in their operations under wraps supposing to retain the competitive edge that technology provides. Moreover, the internal organization of VCs is generally poorly understood, as VCs are secretive about internal operations and their structure (Gompers et al., 2020). A possible explanation is due to the already existing competitive pressure in the early stages of the investment process. Despite the growing number of academics publishing on the use of ML for VC decision making, real-world use of this technology and, as a result, VC acceptance is lagging (Gompers et al., 2020). In recent years, there were different approaches to predicting startup's success using various approaches or technologies (Arroyo et al., 2019; Bai & Zhao, 2021; Catalini et al., 2018; Dellermann et al., 2017; Krishna et al., 2016). However, literature mainly provides insights on technical AI and ML approaches to assist in decision making or in predicting startup success in VC and PE.

Krishna et al. (2016) compared six ML techniques and evaluation methods, such as decision trees (DT), the data of successful and unsuccessful companies of CrunchBase and 20 prediction factors (some are funding, seed stage, valuation) to predict if a company becomes successful (Krishna et al., 2016). Arroyo et al. (2019) tested five different ML approaches, like Decision Trees (DT) or Gradient Tree Boosting (GTB) which belongs to *supervised learning* and over 100 prediction variables on their performance to support VCs in their decision-making process (Arroyo et al., 2019). Both studies researched different ML models in broader context of the VC investment process, especially in the selection of startups.

Catalini et al. (2018) investigated on the prediction of startup success, using data of an US accelerator to train ML models. Ghassemi et al. (2020) researched on how specific entrepreneurial aspects predict the fate of companies within an entrepreneurship competition (Ghassemi et al., 2020). Most recently, Corea et al. (2021) published a research article presenting an investment framework for data-driven VCs. Using CrunchBase data of over 600,000 early-stage companies, they developed a tree-based ML model using a relevant set of investment signals. Namely they used a GTB, which shows similarity to the study of Arroyo et al. (2019) that already proved that GTB outperforms other ML tree classifiers (Corea et al., 2021).

The goal for this extensive review of prior research and analysis of AI and ML in VC is to provide a baseline on which to build the qualitative questionnaire that was used in this dissertation and has forward-thinking academic as well as practical relevance. While most studies provide valuable insights and contribute to academic literature by offering technical guidelines or open-source code for prediction models, they lack the business perspective and strategic insight or information about the long-term success, as well as insights about practical implications for VC funds (Retterath, 2020). Therefore, this dissertation contributes to academic research and explores the potential use and impact of so-called AI/ML and quantitative sourcing tools for VC decision making.

3 RESEARCH METHODOLOGY

3.1 Design

In chapter two a comprehensive literature review was carried out to compile relevant findings and studies from recent years. For this purpose, relevant international journals in the fields of business administration, business venturing and business informatics were examined and analysed. These findings serve to achieve the objective and form the basis for future research. Following the theoretical background of the previous chapters, the methodological approach regarding the empirical part of this thesis will now be explained.

This dissertation uses the qualitative research approach, namely the methodology of qualitative interviewing. Tailored to this methodology, the concept of qualitative content analysis, which is based on scientific text analysis, is considered appropriate (Mayring, 2015). The principles of qualitative content analysis were proposed over 35 years ago and continue to enjoy popularity within qualitative scientific approaches (Mayring, 2015). According to Mayring (2015), qualitative content analysis is characterized by some specific features. In general, the procedure is systematic, guided by rules and theory that form the foundation for the empirical cachets of objectivity, reliability, and validity. Categories act as a central element of content analysis procedure which specifies the objectives of the analysis and justifies each of the evaluation steps to make the results comparable. Besides, the analysis works question related. The text-analytical questions of the analysis arise from the superordinate research questions of the dissertation. To guarantee the systematics, the text analysis works with a flow model which rules are revised after a first test phase (Mayring, 2015). Based on these principles, the data analysis and its subsequent interpretation are transparent, verifiable, and comprehensible, despite subjective influences. Furthermore, the synthetic category construction contributes significantly to the comparability of the results and to the estimation of the reliability of the analysis. Accordingly, this is an open and descriptive approach (Gläser & Laudel, 2010; Mayring, 2015).

According to Flick (2017) qualitative research methods are predestined for largely unexplored questions, as the open, semi standardised approach can be used to investigate the new and unknown. The resulting findings then serve as a foundation for new topic exploration. Thus, the study uses qualitative analysis to gain insights on the experiences and knowledge of selected experts with different industry backgrounds who are directly affected by the change. A further advantage of qualitative content analysis is that the available material can be analysed step by step in a methodologically controlled manner (Mayring, 2015). Semi-structured interviewing serves as an instrument of qualitative research (Rowley, 2012), which will be further detailed in the following subchapters.

3.2 Research Methodology and Approach

After developing a draft questionnaire, the script was circulated among academics and VCs for comments and feedback. Afterwards the draft questionnaire was tested with VCs to gain further feedback. Moreover, advice by marketing research experts was sought to support the survey design, format, and phrasing of the interview questions. As a result of the combined efforts, several changes in language, format and style were made to the interview guideline.

Selection of Interview Partners

For qualitative research, the optimal sample size of interviewees is of particular importance and must be sufficiently considered when developing the research design (Boddy, 2016; Fusch & Ness, 2015). Thus, this dissertation aims for data saturation in this specific field of research. Multiple interviewees are selected to complete the semi-structured interview script, so that with each additional interview, further data saturation is achieved, and validity is assured (Fusch & Ness, 2015). For this study, a sample size of ten interview partners was planned, thus it fulfilled the recommendations of Glaser & Strauss (1976) . They recommend the number ten, because an adequate amount of information is obtained from this number (Glaser & Strauss, 1976). The interview participants were selected according to the so-called *purposive sampling*. With this technique, interviewees are selected based on the professional characteristic they possess. This technique is particularly suitable for qualitative research (Alkassim et al., 2016) and thus was applied in this dissertation. For the selection of interview subjects and the sample size, it is recommended to refer to: the scope of the study, the related nature of the subject, the duration of the interviews with each participant, and the characteristics of the interview participants (Boddy, 2016).

To consider the selection criteria for the present work, it is necessary to define a relevant interviewee or the experts. In this study, experts are defined as individuals who have relevant responsibility or experience in their current position for the research subject. This includes individuals with decision-making authority in the VC investment process, investment committee members, (AI/ML) technology managers, (data) engineers within a fund, as well as topic-appropriate researchers. Subsequently, the study contains a *relevant set of interview* *subjects* predestined to answer the research question. However, the study does not claim to be fully representative for the whole VC market.

Anonymizing interviews is subject to tension between research quality and protection of identification of corporations or individuals. For this interview approach VCs were asked about their willingness to share information non-anonymized. Respondents indicated they did not want information traceable to their company due to competitive reasons. Thus, the anonymity of the interviewees has the highest priority, and the respective companies and names of the research subjects are censored if they were mentioned in the interview. The consent forms between the author and the interview partners are not included for the same reason. They serve as a precautionary measure for the author.

To encourage interview inquiries, the network of Católica Lisbon School of Economics and Business (CLSBE) was used. Thus, alumni and academic workforce were contacted. Additionally, the author sourced from the professional network of the university's student VC club. To encourage responses and completion, the participants were offered to get early access to the study results and to option to discuss potential findings for their VC fund or business. Finally, the author gathered contacts manually by researching professional networks such as LinkedIn and databases like Crunchbase.

Presentation of Interview Partners

In order to prove the professional suitability of the interviewees, the person and his or her professional background are briefly introduced in Table 3.

Table 3 Overview of VC Interview Partners

ID	Type of VC	Investment Stage Focus	Usage of AI or ML	Position of Interviewee
IP01	Independent	Pre-Seed to Series A	Yes	Investment Manager
IP02	Independent	Pre-Seed to Series A	Yes	Partner
IP03	Independent	Seed to Series A	Yes	Investment Manager
IP04	Independent	Seed	No	Investment Manager
IP05	Corporate (CVC)	Seed to Series A	No	Investment Director
IP06	Independent	Seed to Growth	Yes	VC Research Lead Data Projects
IP07	Independent	Seed to Growth	Yes	VC Data Engineer
IP08	Independent	Series A	No	Investment Director
IP09	Independent	Seed	Yes	Investment Manager
IP10	Independent	Seed	Yes	Investment Manager

3.3 Data Collection

For the interviews a semi-structured qualitative interview approach is used. This means, that the interview guideline is not fully structured and there is room for improvisation and flexible questions depending on the interview progress (see Appendix A). In this context, the questions are divided into two parts: First, the predetermined essential questions of the semi-structured interview are referred to as mandatory questions. Second, questions that are asked based on the previous answer or the course of the interview are called contingent questions.

For the semi structured interview, a research questionnaire was created. The interview script is divided into four consecutive sections, which are divided into main questions and subquestions. The script is designed openly and depending on the experience with AI/ML of the interviewee, the questionnaire can be adapted during the interview. The questionnaire is structured in such a way that the individual sections gradually build on each other and create a virtuous cycle to attain deep insights into the manner.

1. Introduction
2. Categorization of Fund and Specialization Questions
3. Artificial Intelligence, Machine Learning and Data in the Investment Process
4. Future of VC Decision Making & Outlook

In order to classify the interview partner, the company (VC) and its investment thesis, a web search is conducted prior to the interview. The **first section** serves to query ambiguities or necessary verifications about the person and the fund, as well as an introduction to the topic of the dissertation. As a result, a common understanding of the topic and then a common knowledge base for the interview is guaranteed. This ensures that all information provided by the interview partner is relevant to the research question and topic. The **second section** includes questions about the specific investment approach of the VC to categorize and differentiate the VC. In **section three**, comprises questions about the status and use of AI and ML in the investment process of the VC. Lastly, the **fourth section** is future-oriented, and the questions are thus intended to investigate future developments and expectations in more detail.

Before conducting the interviews, the broad aim and scope of the dissertation was presented to the interview partners. The interviews were conducted in the period between November 2021 and December 2021 via online video calls to minimize situational disruptive factors such as disturbance of the interview situation. All interviews took place during the interviewees' working hours. A consent to record the entire interview was obtained and attention was drawn to the anonymization of the recording, thus subjects are not inhibited from the interviewer. On average the interview duration was 30 minutes.

3.4 Data Analysis

Data analysis is then carried out based on the transcription material. The basis of the data analysis is the analytical procedure which is also called content structuring. For the structured procedure of the qualitative content analysis a flow model according to Mayring (2015) is applied. This ensures a synthetic penetration and understanding of the content. Figure 3 shows the seven-step approach to structured content analysis following Mayring (2015).

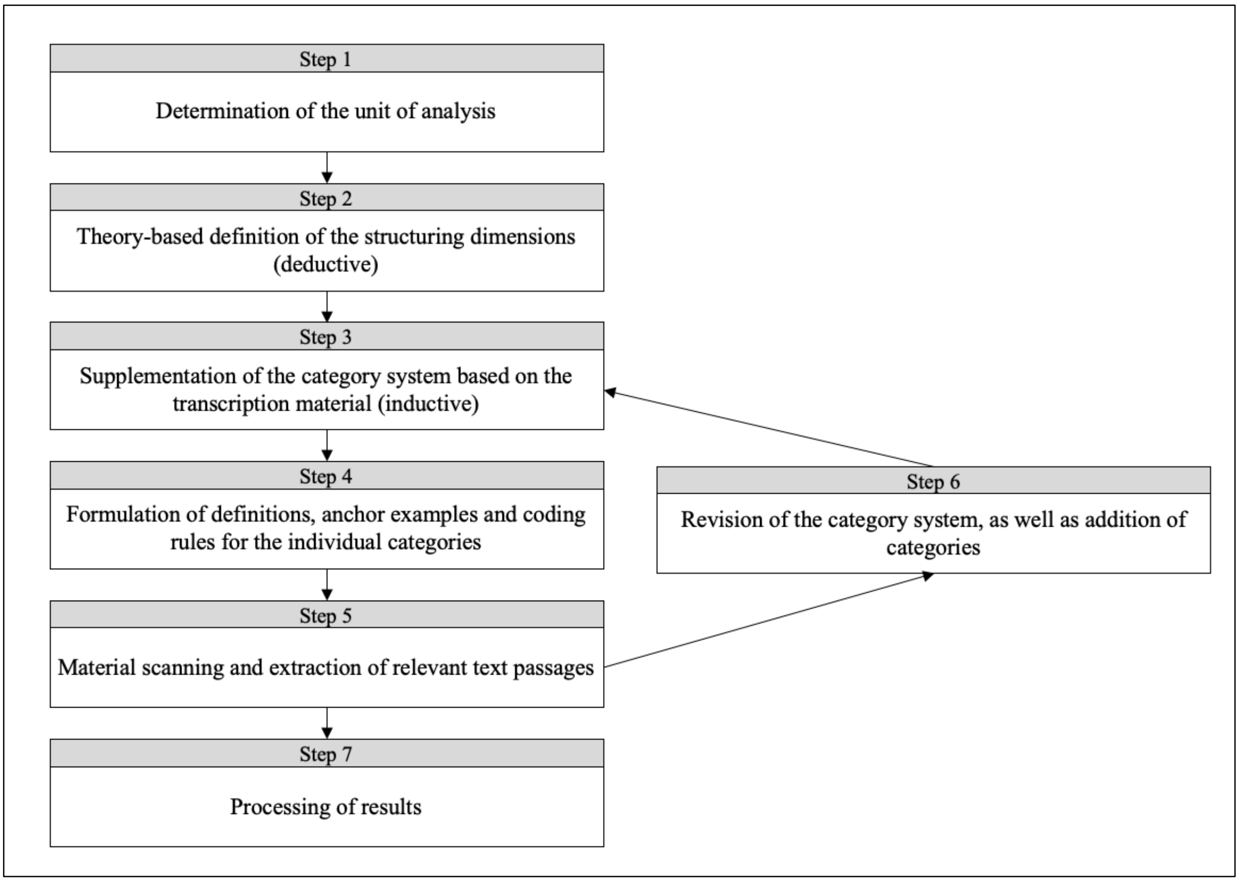

Figure 3 Process Model of Structuring Content Analysis according to Mayring (2015)

Step 1, the determination of a precise unit of analysis, includes the definition of the coding unit and context unit, which defines the relevant textual components for the analysis, as well as the evaluation unit. Step 2 generally foresees a theory-based category formation (deductive). Due to the elaborated research problem and the relevant research gap, a mixed form of deductive and inductive category formation is applied here. In detail, this implies that deductive categories derived from the theory were formed, which were then inductively refined and augmented following the review of the transcription material in step 3. The fourth process step according to Mayring (2015) is defined according to which prerequisites text components can be comprehensibly assigned to a category. This is a three-part process, in which firstly the categories are defined. Then concrete text excerpts, so-called anchor examples of the respective category are given. And thirdly, the associated coding guidelines are formulated, which enable the assignment of the text components in code components to categories in a distinctive manner. In the fifth step, the comprehensive review of the material and the assignment of the individual text passages to the categories is carried out according to the coding rules with the help of the qualitative data analysis software MAXQDA 2022. After the complete material run, the existing category system is then revised, checking whether the categories serve the goal of the analysis. In the seventh and final step, the text extracts and the resulting category system are analyzed and processed in a structured manner. The specific empirical findings that ultimately resulted from this process are presented in following chapter.

According to Mayring (2015), the foundation of qualitative content analysis is a categorical system. Due to the complexity and specificity of the research question elaborated at the beginning of this thesis, a mixed form of deductive-inductive category formation is used. In the first step, the main categories of the category system derived from the theory are roughly captured via the interview guidelines, thus categories are formed deductively. It is derived from two relevant areas. Firstly, from the phases of the investment process, as well as the effects on the individual process steps. Second, from the technology perspective and data used within the investment process. These categories also serve as foundation for the qualitative examination of the data material. Subsequently, and after reviewing the material, the categories are supplemented and revised, which implies an inductive category formation. Table 4 shows the resulting categories and coding rules for the analysis of the text material. The full category system including the subcategories can be found in Appendix B.

Table 4 Overview of Categories and Coding Rules

Category	**Coding Rule**
C1: Decision Making	ALL statements that relate the decision-making process of VCs.
C2: Rationale & Trigger	ALL statements that relate to the organizational rationale and trigger of AI/ML implementation.
C3: Requirements	ALL statements that relate to requirements of VCs towards AI/ML.
C4: Data	ALL statements that relate to the topic of data (sources, types, categories, etc.).
C5: Automated Investment Process	ALL statements that relate to automation and technology used within the investment process.
C6: Business Value	ALL statements that relate to the business value that gets created using quantitative sourcing and AI/ML.
C7: Challenges	ALL statements that relate to the challenges for the implementation and usage of quantitative sourcing and AI/ML.
C6: Business Value	ALL statements related to business value created using quantitative sourcing and AI/ML.
C8: Future of VC Investments Data	ALL statements that relate to the future of how VCs will make decisions.

Transcription & evaluation and qualitative content analysis

The transcription of the interviews serves as the basis for the qualitative text content analysis of the interviews. The aim is to capture what was said and make it accessible for the subsequent analysis. The interviews were recorded using the video software Zoom and a voice recording application. Thereafter, the recordings were transcribed according to the methodology of *selective protocol* according to (Mayring, 2014). All parts pertinent to the research are transcribed. Sections that are not vital for the text analysis and interpretation are not transcribed, for example, the explanation of the research question or extensive introduction sections.

Data Triangulation

The quality of data is at the heart of the criticism leveled at qualitative research. The nature of the analysis in qualitative studies has thus sparked the controversy (Patton, 1999). In this study's analysis, the triangulation process is used to protect against systematic bias (Patton, 1999)and to ensure data saturation (Denzin, 2017). To ensure a thorough understanding of the results, data sources like very recent and pertinent articles, press reports and studies are used and cross-checked against the findings of the qualitative analysis. In other words, data triangulation increases data reliability and quality of the research.

4 RESULTS & DISCUSSION

The present study was designed to determine the impact of AI, more specific ML, and other data-driven approaches on the VC investment process. The quotations that follow come from a series of interviews carried out with experts in the field of VC and advanced technologies such as AI and ML from various organizations. Participants are referred to pseudonymously after each quote. The main transcription quotes of the interviews can be found in Appendix C.

4.1 Rationale for using Machine Learning Tools

The second research question of this research was set out with the aim to understand reasons and the rationale for VCs to use quantitative sourcing tools or ML in their investment process. VC companies are constantly on the search for new promising investment opportunities. The interviews reveal that the rationale for VCs using or investing in data-driven tools and ML are different in nature. As can be seen in Figure 4, VCs rationale for using intelligent tools and data in the decision-making process is tied to the current market dynamics, previously presented. VCs mentioned that such tools help them to manage the competitive situation between funds. This has mainly two reasons: the overall amount of capital in the market and the market speed, leading to increased competitive pressure of VCs needing spotting investment opportunities early on and then securing deals. Additionally, the target to achieve greater operational efficiency by unbundling tied resources of Investment Managers in the very manual process. Another notable finding from the interviews is the need for smaller funds, such as those characterized by lower AUM or lower fund visibility and reputation, to invest in such tools. This is due to the disadvantages of smaller funds compared to larger funds that have a stronger network and can land better deals through their brand, resources, and value-added services.

Moreover, the interviews reveal that the initial trigger of VC internal technology approaches are coming from two main directions: human driven and technology driven. Human driven means, that either executives from a fund bring up AI and ML support in the investment process, or the technological perspective, meaning that data and technological solutions are available. *"So, we are getting right now at a specific moment in time where the critical mass of data is good enough to start these things. So I think it's a conjunction of two different aspects, a human one and a proper data one" (IP06).*

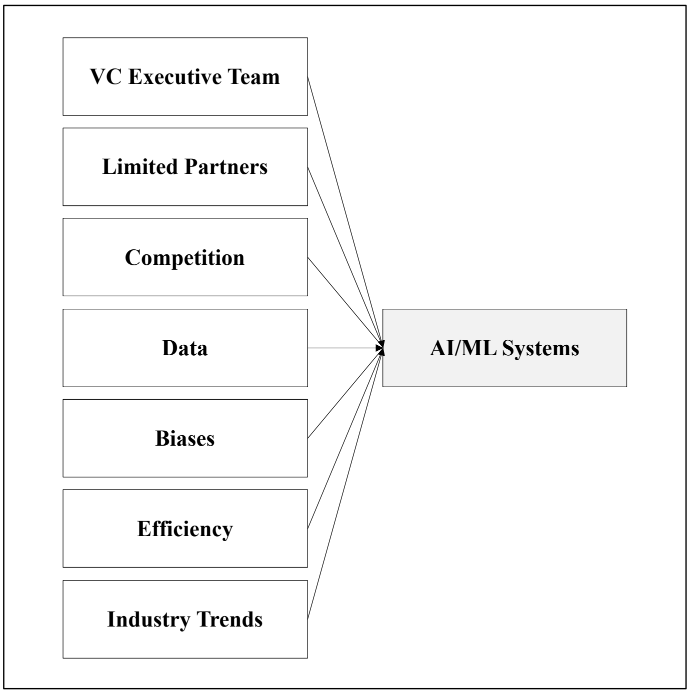

Figure 4 Drivers of VCs using AI/ML Systems

It can be concluded that while the initial trigger for the utilization of AI and ML tools are different, funds target to achieve an improved fund performance by identifying new ventures early and more efficient, thus solving the identification.

4.2 Data-Driven and ML Systems in Venture Capital

The first question in this research was *"How do VCs utilize AI and ML along their venture investment process?"*. The interviews of this dissertation reveal that VCs use both an underlying Customer Relationship Management (CRM) system and data-driven deal sourcing systems, referred to as *web crawler* or *AI/ML systems* that support investors along their investment process and beyond. The CRM system is often an external system used as an underlying relationship intelligence tool that stores relevant data about the company, founders, and contact information. It can also be used to get in touch with founders, make appointments, maintain updated contact details, and communicate between colleagues within a VC (IP01; IP08; IP09; IP10). The main objective of a web crawler or web scraper is to automatically extract information from the internet and store the data in a structured manner. This can include structured as well as unstructured data from various web pages. In this way, it can automatically replicate the human search for information on the web, consequently speeding it up and making it more efficient (Diouf et al., 2019; Saurkar & Gode, 2018) and supports in the 'deal origination' phase.

Both systems are interlinked and interact with each other in a complementary way, mainly in three ways: a) Based on the interviews, the baseline CRM system manually and constantly gets updated and maintained along the investment process. b) The AI/ML Web-Platform, which can range from a 'web crawler' to an ML platform, accompanies the whole decision-making process as a support system. It complements the CRM tool with up-to-date and automatically generated data from various data sources. c) Those systems have several variations, as the interviews found. In addition, VCs develop proprietary in-house tools that include several data sources to provide intelligence for the investors (see Figure 5)

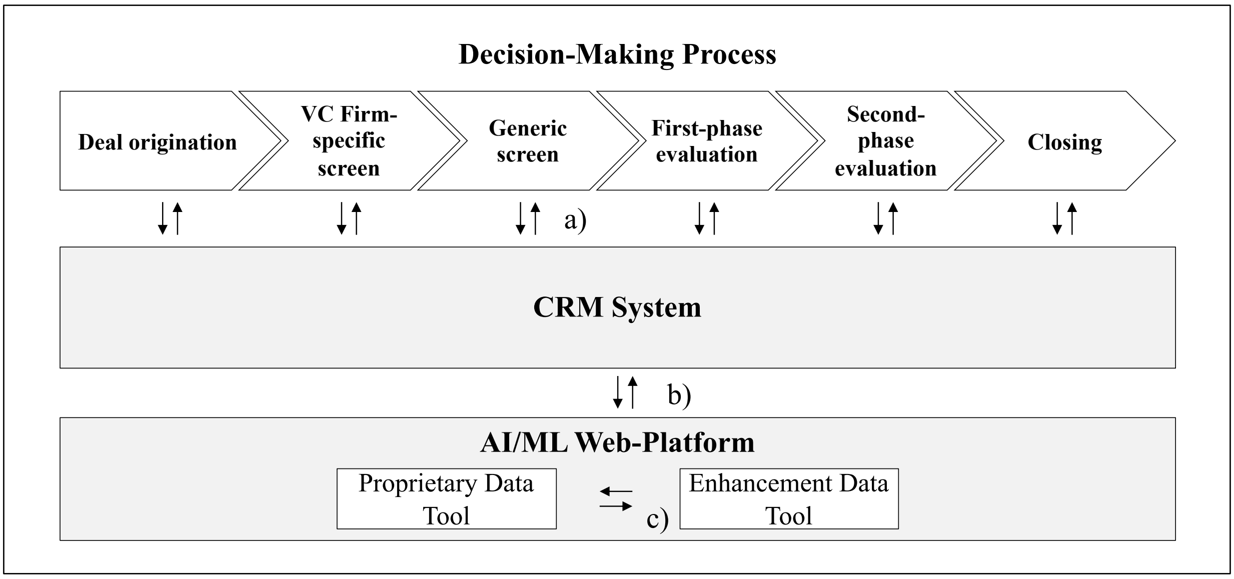

Figure 5 VC Data System Interaction

"We join the data together and we give a full picture to the investment team, so they don't need to jump between different platforms from different services. They just have everything in one view, and they can easily do CRM operations directly from that platform" (IP07).

VCs also mentioned that they also implemented 3rd party enhancement tools with specific data intelligence, for example Natural Language Processing (NLP) to analyze further web data that provides insights on potential investment opportunities (IP02; IP06). NLP is a subform of AI and is used to understand human communication, which is ambiguous due to its complexity

(Russell & Norvig, 2016). NLP in combination with AI and ML is already applied in finance, e.g., for predicting stock market movements by analyzing news articles or web postings (Fisher et al., 2016). In the VC investment process, NLP could be used particularly to automatically analyze unstructured data from the web, such as news articles or company updates, or to validate and understand existing data for VC investment decisions (IP02; IP06).

The combination postulates various external data sources to create a comprehensive and structured information intelligence resource. In contrary interviews IP04 and IP10 revealed that the VC firm decided to use an external provider of deal sourcing and deal flow support systems instead of developing a proprietary system.

4.3 Impact on the VC investment process

How technology and data in the context of digitization enhances existing VC investment mechanisms was the second research question, building on the previous subchapter. As mentioned in the previous section, VCs commonly use a CRM system to manage investment-related data along the investment funnel. Additionally, VCs use AI or ML web-platforms connected to the CRM system. This section addresses how those tools impact the three stages of the investment process and what specific data sources are used.

Deal Origination

Deal origination, also called sourcing, is the step in the decision-making process where VCs rely most on technology assistance. The rationale is that this step has the most significant potential for automation, since the critical decisions about a company are being taken in the screening stages. In addition, broad and extensive sourcing activity potentially leads to value creation when data is intelligently analyzed. *"It's like putting together different data sources, like in the same format and having signals that automatically trigger if something happens" (IP06).*

As mentioned, VCs face a variety of biases along their investment process. The most mentioned bias that VCs reflect upon is the so-called "network bias". The interviewees said they are aware that startups that become visible to them are inherently biased by their personal and professional network often shaped over the years. Technology can come into play in this specific stage to take over one primary task—automatically identifying investment opportunities by searching the web. The interviews reveal that VCs use identification 'web crawler' and data science to spot two types of companies: first, companies pertinent to the VC firm are based on past investment history and various criteria that train the data model, and second, companies that are relevant to the VC firm‘s without the VC firm being aware of them.

“As you can imagine, if we get a very good model, that's going to be a bit of a game changer for the investment team. Because we exclude all the deals that we know that they are not interested in. And the investment team is focused on a subset, which is most likely very interesting deals for them. So, they are going to spend much less time on, wasting less time on reading all the deals. They are speeding up the investment process. They are speeding up the due diligence. They have much more time to do research or to focus on specific companies and work on those. So that's mainly what we do for sourcing. But I guess as far as we say it's, that deal sourcing and kind of investing in VC is always going to be at least the foreseeable future, a human-driven interaction” (IP07).

It can be concluded that data-driven sourcing is of great importance to the VC industry and significantly impacts outbound activities. Overall, it reduces the effort required to march for investment opportunities manually and extends the investment team’s information content and potential research reach.

VC firm-specific screen and general screening

In the screening stage, VCs screen investment opportunities for potential and whether they fit their investment thesis. According to the interviews, when screening for the most prospective investment to pursue, technologies like ML assist investors in identifying and prioritizing the most promising start-ups to investigate. *“It doesn’t tell us, what is the company we have to sign a check on, but it tells us, what are the companies that we should spend time on” (IP02).*

The VCs that use AI in their investment process indicated that an AI-based tool significantly aids them intelligently prioritize investment opportunities. Prioritization is done by first screening for the hard criteria of the investment thesis, which include, as described at the beginning, industry, geography, stage, and investment size. Subsequently, softer SC, such as market, product, or technology, are screened. Since most data is already collected in the ‘Deal Origination’, additional information is added to start-ups already listed in the web-based system. The existing data is mainly supplemented by unstructured data, such as news mentions. In the case of screening or monitoring prior an investment over a more extended period, data from start-up databases and social network data are also added. In this way, data remains current and promising companies that have already been identified in early phases can be evaluated with current information. "*What AI can do for us is actually help me prioritize, understand who to really reach out to, which founder I might get a warm intro to. Our crawler might show me, by the way, this founder there has already someone been in touch within the fund" (IP01).*

In summary, these results show that data-driven tools such as web crawlers or ML tools have a substantial impact on the screening phase. Taken together, VCs get a prioritization of the companies to be screened based on predictive statistical models and an initial screening of the potential for success. *"So, scorecards are not coming out from my mind. They are coming out of a sample of data coming from past failures and successful companies, by the way. And this is something we continuously update" (IP02).* This leads to a more effective and efficient process for screening startups by reducing the overall amount of information to be analyzed manually and allowing the investor to spend more time on relevant investment opportunities.

Impact and Summary

The combination of predictive tools like AI/ML systems and the CRM systems can result in positive business value for VCs on all stages of the investment process. The interviews reveal that the main positive impact of AI and ML platforms and their algorithms is increase of efficiency. The interviewees commonly ranked 'increased efficiency' as main positive impact. Controversial results of interviewees were reported when discussing 'reduced bias'. This can be drawn back to the fact that the system of one VC was directly linked to their professional network and analyzes professional connection signals, thus the interviewee stated this can increase bias. *"I think it increases my bias, because I get flagged a lot of things that are in my inherent ecosystem" (IP01).* In controversy, one interviewee mentioned their overall goals were to decrease bias when initially discussing the use of AI and ML in their investment process. They ranked it as his VC firm's second most positive impact. Overall, interviewees mentioned speed increase as the second most positive impact on the investment process. Investors at VC funds can spot companies and entrepreneurs faster and spend less time on declining start-ups since the systems provide them with a prioritized list of start-ups to look at.

"*Obviously, these are estimates, don't take them as face value. But on average, we've been decreasing the time from sourcing to identifying a deal by five times, which is a lot if you think about our saved of my time" (IP02).*

Regarding research question number five, *"Where can AI have an impact on the venture investment cycle to reduce the common VC problems?"*, table 5 summarizes the value-creating impact. To cluster the business value, the framework of Borges et al. (2021) is adapted. Thus, our results fall into the following two dimensions: i) Decision Making; and ii) Automation.

Table 5 Business Value of AI Tools

Sources of Value Creation	**Affected Investment Stage**	**Business Value Impact**
Decision Making	Deal Origination Firm-specific screen General Screen	• Exclusion of investment opportunities based on data • Quantitative screening of investment opportunities • Increased efficiency and speed
Automation	Deal Origination Firm-specific screen General Screen	• Automated deal origination of potential companies or founder teams based on data intelligence • Automation of the screening based on data, to the extent that human interaction is reduced to the minimum • Automatic prioritization of investment opportunities • Increased efficiency and speed of execution

Note. Sources: (IP01; IP02; IP03; IP04; IP06; IP07; IP09; IP10)

Based on the interviews and the qualitative content analysis, it was found that VCs use web-based systems for their outbound sourcing, while network-related activities remain the major inbound channel. Along the investment process, multiple automated systems, data sources, and manual tasks are combined to create a holistic human-machine interaction. In the 'deal origination' stage, VCs commonly stated that, additionally to their inbound activities and manual research activities they use *web crawler systems* to identify possible investment opportunities or at least consider using quantitative sourcing technology. These findings are in contrast with previous research (Gompers et al., 2020) and can be explained with a number of reasons. Various structured and unstructured data sources are scraped from the Internet and intelligently combined to identify potential early-stage companies or even founders before the company is officially established. In all cases, respondents reported that "talent data," which often comes from professional social networking sites like LinkedIn, is a valuable signal for spotting investment opportunities early on. As Figure 6 indicates, after the initial step of scraping web data, the existing information gets augmented by further data sources. This data is rather unstructured and data sources like news feeds or industry relevant tech magazines get included and searched. Once potential start-ups are identified in the web, the systems also do initial screening tasks. VC company specific criteria are used as a first filter, these can include hard selection criteria like industry, but also are subject to history data from a trained model. Thereby the system is constantly updated and re-trained by data from previous decisions.

"*By the way, the scorecards that we mentioned have been informed by predictive by statistical models. We've been working on the variables and the weights behind the scorecards are developed by making predictive models about company success. By the way. So scorecards are not coming out of my mind. They are coming out of a sample of data coming from past failures and successful companies, by the way. And this is something we continuously update" (IP02).*

As last step before the more detailed due diligence, VCs use available data from the web and from company specific data to evaluate the start-up. According to the interviews, this step, however, is not as automated and requires more individual and human analysis. Once the stages are processed, the investment team can access a prioritized list of potential start-ups to look at for further screening and selection.

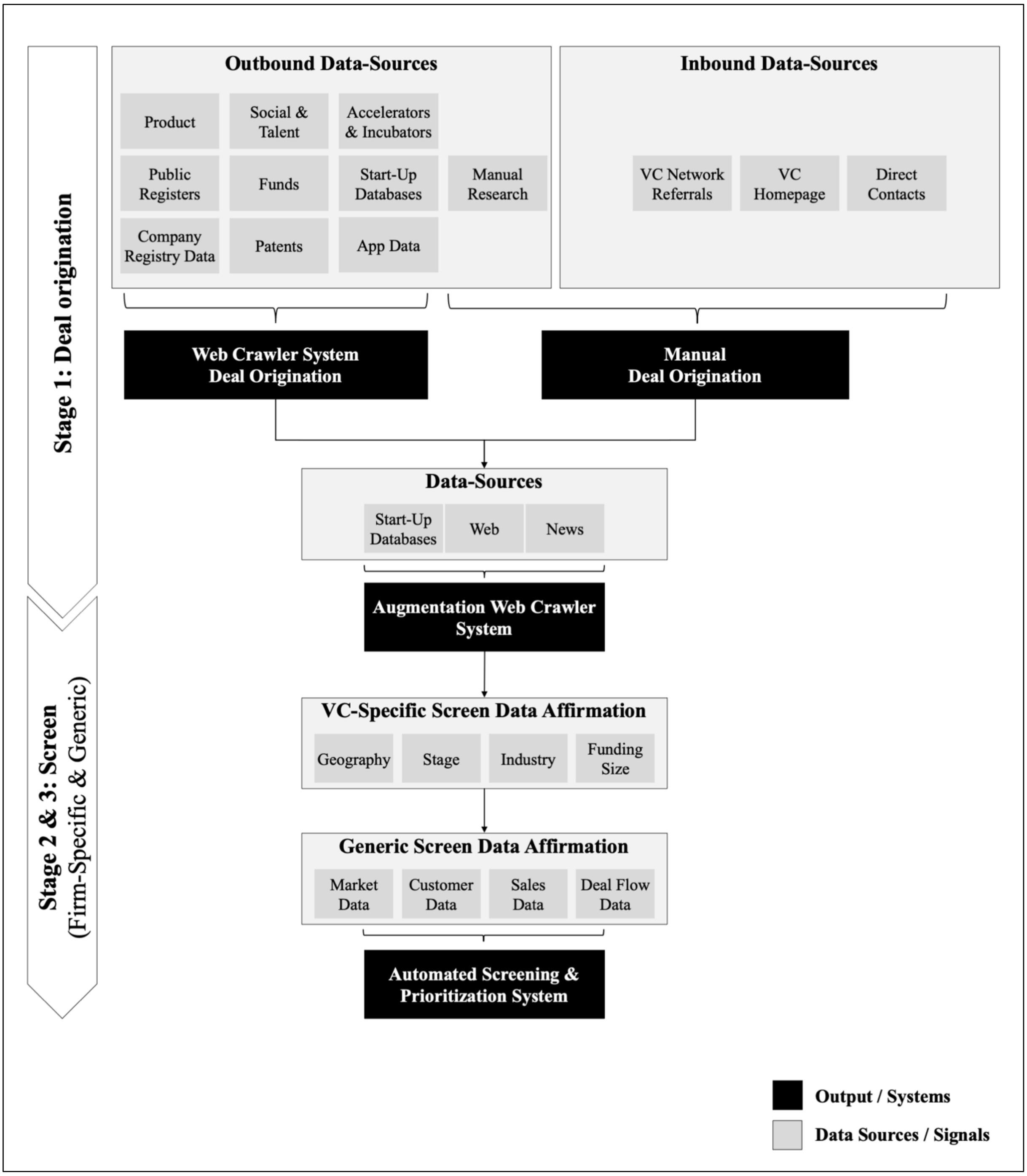

Figure 6 Interplay and Flow of Deal Origination and Screening Tools

Table 6 summarizes the information on the extent of data utilization and data sources in each phase according to the interviews and is triangulated with existing research to provide accurate, exemplary data sources from practice. It also indicates which data sources VCs use to identify signals from promising startups. Finally, the table outlines the VCs' perception of the improvements on each stage.

Table 6 Overview of Data Usage and Impact of Investment Stages

Investment Stage	**Data Usage**	**Relevant Data Sources**	**Improvements**
Deal Origination:	High	• Funds, Accelerators & Incubators (e.g., YCombinator) • VC databases (e.g., Crunchbase) • Social networks (e.g., LinkedIn, Twitter) • Talent data (e.g., LinkedIn) • Product libraries (e.g., Product Hunt) • Developer platforms (e.g., GitLab, GitHub, StackOverflow) • App data (e.g., AppAny, SensorTower) • News (e.g., Google News Feed) • Public Registers • Academic Publishing • Patents	• Increased Efficiency • Increased Speed • Deeper market insights • Expanded sourcing capabilities and resources
VC Firm-Specific Screen	High	• VC databases (e.g., Crunchbase) • Investor network data • Social networks (e.g., LinkedIn, Twitter) • News (e.g., Google News Feed)	• Increased Efficiency • Increased Speed • Reduced Uncertainty
Generic Screen	Medium	• Customer data • Deal flow data • Sales data • Market intelligence data (e.g., trends, news mentions, web traffic, etc.)	• Increased Efficiency • Reduced Uncertainty • Augmented Investment Insights

Note. Sources: (IP01; IP02; IP03; IP04; IP05; IP06; IP07; IP08; IP10; Arroyo et al., 2019; Corea, 2019a; Ghassemi et al., 2020; Liang & Yuan, 2016; Trocha, 2019; Weibl & Hess, 2019)

4.3.1 Challenges and Limitations

With respect to the fourth research question, about *challenges when using quantitative sourcing and AI/ML tools,* it was found that the usage of sophisticated technological systems and tools in the decision-making process also comes with some challenges. This study's interviews reveal that challenges can be clustered in four areas: i) Organizational ii) Cultural, iii) Technological, iv) Aversional.

Under i) Organizational, the interviewees unanimously reported that allocated resources in the technological development of a VC firm are limiting the possibilities of AI and ML in the decision-making process. Often, VCs are subject to a lack of focus or allocate only a few resources on the internal developments of artificial technologies and prioritize capacity spent on analytical tasks like deal flow support. *"Some funds just don't commit resources. [...] That's something that I've seen many companies suffering from." (IP02)*

While ii) culture is not particularly a challenge for VCs on the more junior levels like analysts in their operational deal flow work, the interviews show that executives at VC companies rely more on their gut feeling. VC companies themselves are not innovative and use outdated technologies. *"The priority is making investments, and this may cause some sort of misallocation of resources when it comes to that" (IP02).* Consequently, VCs should consider to adapt their corporate culture to overcome their reluctance to use algorithms and automate their work. Not only because algorithms can outperform human investors (Retterath, 2020), but also because there is a new generation of VCs with a purely data-driven focus and culture (Corea, 2019b).

On the iii) technological and data level, VCs mention limitations in analyzing and predicting team-related data and stage-related data. Interviewees noted that predicting data on the founding team is of significant importance but also complex. Overall, it can be summarized that data availability, data quality, and filling parse data were mentioned as challenges in the usage and implementation of AI and ML tools. *"So the hardest part for us is to actually avoid parse data and filling all the data points for any deal that we have, for any opportunities that we look at. And data quality is the biggest problem that I guess every VC is facing." (IP07)*

When automating investment decisions, VCs face the dilemma between managing automation and control of the investment process—iv) algorithm aversion. *"I'd say that's where my distrust comes from, having a tool doing screening for me with respect to decisions of three to four million euros" (IP05).*

More technology and decision-making support will automatically result in less human control. While the impact on decision control of AI and ML is limited in the early stages of the investment process, such as in "deal origination" or "firm-specific" screening, the risk of a machine "miss" increases with each subsequent step in the decision process. Thus, it is evident from the interviews that VCs rely on algorithms, especially in the early stages. VC investors' reliance on AI also increases in later stages of the investment process when data-tested ML algorithms outperform investors' decisions, and VCs become increasingly aware of the impact of their heuristics and biases.

4.3.2 Future of VC Investment

In accordance with the requirements of the project and research question number five, the interview partners were also asked for their assessment of future application scenarios. In sum, the VCs interviewed reported that they expect advanced technologies such as ML or web scraper to become a compulsory technology for VCs. This refers to the increasing importance of technologies in VC business processes, which are necessary for a VC to remain competitive (IP02; IP06; IP08). VCs are commonly aware of the changes and innovation in their industry. This includes knowledge of VC-specific AI offerings, VCs increasingly investing in their own technology stack, and fully data-driven VCs are emerging (IP06).

With consideration to prospective application scenarios of advanced technologies, the interviewees stated automation of several internal and external processes. Those can be clustered into automation of communication and automation of investment processes (see Figure 7).

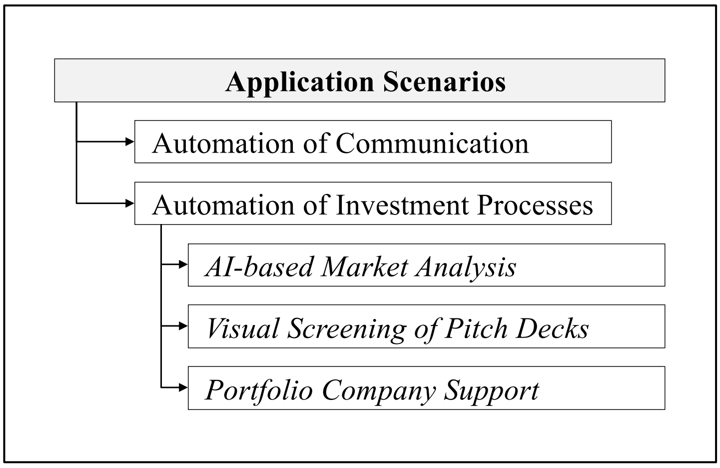

Figure 7 Overview of Identified Application Scenarios

As mentioned in the interviews, *communication with founders* is a manual and resource-intensive process, so VCs anticipate automating and smoothing communication processes utilizing data. For example, an integration of extended contacts into a CRM system would be possible, where a trigger is used to send an automated contact inquiry and invitation to a pitch meeting (IP01; IP09). Internal communication between employees of a fund can be enhanced by intelligent data processing as well (IP10).

AI could also be expanded in terms of an *analysis of general market trends* and market changes in the view of the interview partners. For example, the detection of emerging sectors in respect of newly founded companies is conceivable. This could be accompanied by the simultaneous identification of possible gaps between start-ups in a market and investors' money flows. This information is particularly relevant, as it allows early anticipation of trends and developments and enables VC investors to adapt and more accurately guide their deal sourcing activities.

Application of *computer vision* could thus mean the automatic extraction and recognition of relevant information in pitch decks or deal flow specific documents. For the analysis of pitch decks, this could mean to analyze and benchmark the structure, word count, key words, and quantitative information, such as financials, competitors, or market relevant data like market sizing.

Following the field of application before an investment decision, AI can also play a significant role in the *management of portfolio companies* (IP02; IP03; IP07). In addition to the general management of portfolio companies, the interview partners see potential in the use of recruiting and sales (IP02). This is exemplified in the work of quantitative VCs such as *SignalFire* (IP02), since they state on their homepage “We take a human + technology approach to supporting our companies” (*SignalFire*, n.d.). SignalFire mentions that they use AI to support their portfolio companies in hiring talent and to provide competitive intelligence in areas such as market analysis, pricing, and product benchmarking. In summary, the use of AI by traditional VCs may expand to downstream stages of the decision-making process in the future. However, the focus lies on complementing human performance, as the human component plays an important role, especially for early-stage startups (*SignalFire*, n.d.).

5 Conclusion

5.1 Summary

VCs invest in innovative companies and technologies of the future. Therefore, they must remain innovative themselves to identify promising business and not pay more than the real valuation of a company. VC investors rely heavily on their network contacts, heuristics, gut feeling and past investment successes when making decisions. As a result, VCs are prone to several challenges, such as biases in their investment process and the potential to exclude promising investment opportunities. There is currently little academic understanding of how VCs use advanced technologies such as data-driven sourcing solutions or ML in their investment process. This research aimed to understand the impact of AI and more specific ML approaches on the VC investment process, with a focus on deal sourcing and company screening. Based on an explorative qualitative research approach using semi-structured interviews with 10 VCs and subsequent qualitative content analysis, this research provides insights on the practical usage of data and quantitative driven decision making.

The study found that VCs use technological support in deal origination in the form of web crawlers that search relevant data sources for signals of potential success. The investigation of the rationale of data-informed decision support revealed that there are several internal and external initiations for the implementation, with all being drawn back to the rising competitive situation of VC fund performance and, consequently, their need to identify and make investments into startups early on. The research has also shown that VCs synthesize their information and sourcing efforts in a CRM system that is linked to their proprietary or third party developed web scraper to enable efficient and automated processing of deal flow data along the investment process. In the case of proprietary solutions, they also turn to third-party providers for technological support in the acquisition and processing of signal-relevant data. With regard to data sources, VCs use a broad spectrum of data along the investment stages to spot signals of venture success and potential new startups. The further down the six-stage investment process, the less automation and data sources are used. In the screening stages, their systems automate initial screening decisions for investment thesis and provide support in general screening efforts. The investigation showed that VCs especially rely on data-driven support and ML in the 'deal origination' and 'firm-specific screen' stages. To summarize, VCs don't just utilize web scraping technologies to harvest data from the web, they also combine CRM and intelligence systems, as well as apply AI and ML to identify new ventures to investigate with priority. This leads to increases in efficiency and speed of execution due to automation, and augments the information available to VC firms. Thus, it can lead to untied resources and, as a result, lower operating costs for VC funds.

The approach of qualitative research was used to get a deep understanding of the impact of data-driven and ML approaches on the investment process as well as the human-technological interaction of VC firms. Interestingly, VCs assume that AI approaches to decision making will become a commodity in the industry. The future will tell when and at which stages of the investment process VCs can let go of the steering wheel of data-driven technology to not only support humans but also make autonomous decisions.

5.2 Implications for theory and practice

This study extends the knowledge within the management and entrepreneurial finance literature as previous research lacks an understanding of AI/ML and quantitative approaches to VC investment decisions. It provides an empirical qualitative investigation into technology usage in the decision-making process of VC investors. Accordingly, this study extends the findings from Gompers et al. (2020) as it can be argued that quantitative sourcing is used or planned in our sample. Moreover, it contributes to the study of Retterath (2020) as it extends his research by conducting empirical research interviews and provides insights on the rationale of VCs using algorithms and ML in their investment process. Additionally, it sheds light on business value impact in the respective stages. Consistent with Franke et al. (2006b), Zacharakis & Meyer (1998), and Zacharakis & Shepherd (2001), the interviewed VCs report facing biases and identification problems in their investment process. Furthermore, VCs revealed that they are aware of these challenges and see ML applications and data-driven sourcing as solutions. Lastly, this dissertation contributes to existing computer science literature (Arroyo et al., 2019; Bai & Zhao, 2021; Catalini et al., 2018; Dellermann et al., 2017; Krishna et al., 2016), as it adds a management perspective on how VCs use technology to predict new venture success and provides insights into what signals and data are used in practice. Based on these conclusions, it is suggested that VCs investigate quantitative approaches and ML in deal sourcing, screening, and selection to remain competitive and to reduce their biases. Practitioners should also consider investing in a holistic AI strategy as younger VC firms have the data-driven approach integrated into their processes and company DNA from day one. However, the quantitative approach is a costly endeavor for VCs based on the required resources. This study also has implications for entrepreneurs and new venture founders seeking financial investments from VC organizations, as they will need to update their online presence and published information with quantitative metrics.

5.3 Limitations and avenues for future research

It is acknowledged that this dissertation has some limitations. To begin, because the empirical study was conducted in the form of qualitative research, caution should be applied regarding the generalizability of the findings. Although the sample size of the interviewed VCs met the requirements of Glaser & Strauss (1976), a larger sample size or a broader geographical spread could provide more detailed insights. Furthermore, the geographical focus of VCs in Europe limited the scope of this study; thus, the study does not claim to be fully representative of the global VC industry. This is particularly important as the VC industry in the US, for example, is potentially much more data-driven and has a different starting point in terms of data access.

Following the findings of this dissertation, there are four avenues for future research. A natural progression of this work is the removal of the identified limitations, such that the results can be validated to a higher degree by interviewing more VCs in Europe, as well as understanding how VCs in the US are approaching advanced technologies, which would provide further insights. Furthermore, a greater focus on the impact of AI in later stages of the investment process, such as due diligence and post-investment activities, could be usefully explored in further research. Future empirical studies could investigate the impact and necessity of increasing data volumes (e.g., volume, number of data sources, types of data sources) on the informativeness and human usability of AI and ML tools in the investment process. This research has also raised the question of the long-term effects of quantitative sourcing and AI on fund performance and business value. Further research on these questions would be a useful way of providing a better understanding of the performance of machines versus human investors.

Bibliography

Alkassim, R. S., Tran, X., Rivera, J. D., Etikan, I., Abubakar Musa, S., & Sunusi Alkassim, R. (2016). Comparison of Convenience Sampling and Purposive Sampling. *American Journal of Theoretical and Applied Statistics*, *5*(1), 1–4. https://doi.org/10.11648/j.ajtas.20160501.11

Arroyo, J., Corea, F., Jimenez-Diaz, G., & Recio-Garcia, J. A. (2019). Assessment of Machine Learning Performance for Decision Support in Venture Capital Investments. *IEEE Access*, *7*. https://doi.org/10.1109/ACCESS.2019.2938659

Åstebro, T., & Elhedhli, S. (2006). The Effectiveness of Simple Decision Heuristics: Forecasting Commercial Success for Early-Stage Ventures. *Http://Dx.Doi.Org/10.1287/Mnsc.1050.0468*, *52*(3), 395–409. https://doi.org/10.1287/MNSC.1050.0468

Bai, S., & Zhao, Y. (2021). Startup Investment Decision Support: Application of Venture Capital Scorecards Using Machine Learning Approaches. *Systems*, *9*(3). https://doi.org/10.3390/systems9030055

Boddy, C. R. (2016). Sample size for qualitative research. *Qualitative Market Research: An International Journal*, *19*(4), 426–432. https://doi.org/10.1108/QMR-06-2016-0053

Boocock, G., & Woods, M. (1997). The Evaluation Criteria used by Venture Capitalists: Evidence from a UK Venture Fund. *International Small Business Journal: Researching Entrepreneurship*, *16*(1). https://doi.org/10.1177/0266242697161003

Borges, A. F. S., Laurindo, F. J. B., Spínola, M. M., Gonçalves, R. F., & Mattos, C. A. (2021). The strategic use of artificial intelligence in the digital era: Systematic literature review and future research directions. *International Journal of Information Management*, *57*, 102225. https://doi.org/10.1016/J.IJINFOMGT.2020.102225

Brynjolfsson, E., Hitt, L., & Kim, H. (2011). Strength in Numbers: How Does Data-Driven Decisionmaking Affect Firm Performance? *International Conference on Information Systems 2011, ICIS 2011*, *1*, 541–558. https://doi.org/10.2139/SSRN.1819486

Brynjolfsson, E., & Mcafee, A. (2017). Artificial Intelligence, For Real. *Harvard Business Review*.

Brynjolfsson, E., & McElheran, K. (2016). The Rapid Adoption of Data-Driven Decision-Making. *American Economic Review*, *106*(5), 133–139. https://doi.org/10.1257/AER.P20161016

Burgess, A. (2018). *The Executive Guide to Artificial Intelligence*. Springer International Publishing. https://doi.org/10.1007/978-3-319-63820-1

Catalini, C., Foster, C., & School, H. B. (2018). *Machine Intelligence vs. Human Judgement in New Venture Finance*. Mimeo.

Collins, C., Dennehy, D., Conboy, K., & Mikalef, P. (2021). Artificial intelligence in information systems research: A systematic literature review and research agenda. *International Journal of Information Management*, *60*, 102383. https://doi.org/10.1016/J.IJINFOMGT.2021.102383

Colson, E. (2019). *What AI-Driven Decision Making Looks Like*. Harvard Business Review. https://hbr.org/2019/07/what-ai-driven-decision-making-looks-like

Corea, F. (2019a). *AI and Venture Capital*. Springer, Cham. https://doi.org/10.1007/978-3-030-04468-8_15

Corea, F. (2019b). *Data-Driven VCs: Who Is Using AI To Be A Better (And Smarter) Investor*. https://www.forbes.com/sites/cognitiveworld/2019/05/02/data-driven-vcs-who-is-using-ai-to-be-a-better-and-smarter-investor/?sh=1f20293b1d34

Corea, F., Bertinetti, G., & Cervellati, E. M. (2021). Hacking the venture industry: An Early-stage Startups Investment framework for data-driven investors. *Machine Learning with Applications*, *5*, 100062. https://doi.org/10.1016/J.MLWA.2021.100062

Council, J. (2021). *VC Firms Have Long Backed AI. Now, They Are Using It. - WSJ*. https://www.wsj.com/articles/vc-firms-have-long-backed-ai-now-they-are-using-it-11616670000

Davenport, T. (2018). *AI Advantage: How to Put the Artificial Intelligence Revolution to Work* (1. Edition). The MIT Press.

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. *ICIS 2017: Transforming Society with Digital Innovation*. https://doi.org/10.2139/SSRN.3159123

Deloitte. (2017). *Artificial Intelligence Defined*. https://www2.deloitte.com/se/sv/pages/technology/articles/part1-artificial-intelligence-defined.html

Denzin, N. K. (2017). *The research act: A theoretical introduction to sociological methods*. Transaction publishers.

Dietvorst, B., Simmons, J. P., & Massey, C. (2015). Algorithm Aversion: People Erroneously Avoid Algorithms after Seeing Them Err. *Journal of Experimental Psychology: General*, *144*(1), 114–126. https://doi.org/10.1037/xge0000033

Diouf, R., Sarr, E. N., Sall, O., Birregah, B., Bousso, M., & Mbaye, S. N. (2019). Web Scraping: State-of-the-Art and Areas of Application. *Proceedings - 2019 IEEE*

International Conference on Big Data, Big Data 2019, 6040–6042. https://doi.org/10.1109/BIGDATA47090.2019.9005594

Duan, Y., Edwards, J. S., & Dwivedi, Y. K. (2019). Artificial intelligence for decision making in the era of Big Data – evolution, challenges and research agenda. *International Journal of Information Management*, *48*, 63–71. https://doi.org/10.1016/J.IJINFOMGT.2019.01.021

Edwards, J. S., Duan, Y., & Robins, P. C. (2000). An analysis of expert systems for business decision making at different levels and in different roles. *European Journal of Information Systems*, *9*(1). https://doi.org/10.1057/palgrave.ejis.3000344

Fisher, I. E., Garnsey, M. R., & Hughes, M. E. (2016). Natural Language Processing in Accounting, Auditing and Finance: A Synthesis of the Literature with a Roadmap for Future Research. *Intelligent Systems in Accounting, Finance and Management*, *23*(3), 157–214. https://doi.org/10.1002/ISAF.1386

Flick, U. (2017). *The SAGE Handbook of Qualitative Data Collection*. SAGE Publications.

Franke, N., Gruber, M., Harhoff, D., & Henkel, J. (2006). What you are is what you like—similarity biases in venture capitalists' evaluations of start-up teams. *Journal of Business Venturing*, *21*(6), 802–826. https://doi.org/10.1016/J.JBUSVENT.2005.07.001

Franke, N., Gruber, M., Harhoff, D., & Henkel, J. (2008). Venture Capitalists' Evaluations of Start–Up Teams: Trade–Offs, Knock–Out Criteria, and the Impact of VC Experience. *Entrepreneurship Theory and Practice*, *32*(3). https://doi.org/10.1111/j.1540-6520.2008.00236.x

Fried, V. H., & Hisrich, R. D. (1994). Toward a Model of Venture Capital Investment Decision Making. *Financial Management*, *23*(3), 28. https://doi.org/10.2307/3665619

Fusch, P. I., & Ness, L. R. (2015). Are We There Yet? Data Saturation in Qualitative Research. *The Qualitative Report*, *20*(9), 1408–1416. http://www.nova.edu/ssss/QR/QR20/9/fusch1.pdf

Ghassemi, M. M., Song, C., & Alhanai, T. (2020). *The Automated Venture Capitalist: Data and Methods to Predict the Fate of Startup Ventures*. https://github.com/ghamut/automated-venture-capitalist

Glaser, B. G., & Strauss, A. L. (1976). *The discovery of grounded theory: strategies for qualitative research / B and Anselm L. Strauss* (A. de Gruyter, Ed.). Aldine Transaction.

Gläser, J., & Laudel, G. (2010). Experteninterviews und qualitative Inhaltsanalyse. *Experteninterviews Und Qualitative Inhaltsanalyse*. https://doi.org/10.1007/978-3-531-91538-8

Gompers, P. A., Gornall, W., Kaplan, S. N., & Strebulaev, I. A. (2020). How do venture capitalists make decisions? *Journal of Financial Economics*, *135*(1). https://doi.org/10.1016/j.jfineco.2019.06.011

Gompers, P. A., & Lerner, J. (2004). *The venture capital cycle*. MIT press.

Haefner, N., & Morf, P. (2021). AI for Decision-Making in Connected Business. In F. Gassmann Oliver and Ferrandina (Ed.), *Connected Business: Create Value in a Networked Economy* (pp. 215–231). Springer International Publishing. https://doi.org/10.1007/978-3-030-76897-3_12

Hall, J., & Hofer, C. W. (1993). Venture capitalists' decision criteria in new venture evaluation. *Journal of Business Venturing*, *8*(1). https://doi.org/10.1016/0883-9026(93)90009-T

Kaplan, S. N., & Lerner, J. (2010). It Ain't Broke: The Past, Present, and Future of Venture Capital. *Journal of Applied Corporate Finance*, *22*(2). https://doi.org/10.1111/j.1745-6622.2010.00272.x

Kaplan, S. N., & Lerner, J. (2017). *Venture Capital Data* (Vol. 75). University of Chicago Press.

Kaplan, S. N., Sensoy, B. A., & Strömberg, P. (2009). Should Investors Bet on the Jockey or the Horse? Evidence from the Evolution of Firms from Early Business Plans to Public Companies. *The Journal of Finance*, *64*(1), 75–115. https://doi.org/10.1111/J.1540-6261.2008.01429.X

Kaplan, S. N., & Strömberg, P. (2001). Venture Capitals As Principals: Contracting, Screening, and Monitoring. *American Economic Review*, *91*(2), 426–430. https://doi.org/10.1257/AER.91.2.426

Kaplan, S. N., Strömberg, P., & Sensoy, B. A. (2002). How well do venture capital databases reflect actual investments? *Available at SSRN 939073*.

Knight, F. H. (1921). *Risk, uncertainty and profit* (Vol. 31). Houghton Mifflin.

Kollmann, T., & Kuckertz, A. (2010). Evaluation uncertainty of venture capitalists' investment criteria. *Journal of Business Research*, *63*(7), 741–747. https://doi.org/10.1016/J.JBUSRES.2009.06.004

Korteweg, A., & Sorensen, M. (2010). Risk and Return Characteristics of Venture Capital-Backed Entrepreneurial Companies. *Review of Financial Studies*, *23*(10). https://doi.org/10.1093/rfs/hhq050

Krishna, A., Agrawal, A., & Choudhary, A. (2016). Predicting the Outcome of Startups: Less Failure, More Success. *IEEE International Conference on Data Mining Workshops, ICDMW*, *0*, 798–805. https://doi.org/10.1109/ICDMW.2016.0118

Liang, Y. E., & Yuan, S. T. D. (2016). Predicting investor funding behavior using crunchbase social network features. *Internet Research*, *26*(1), 74–100. https://doi.org/10.1108/INTR-09-2014-0231/FULL/XML

Logg, J., Minson, J., & Moore, D. A. (2018). Algorithm Appreciation: People Prefer Algorithmic To Human Judgment. *SSRN Electronic Journal*. https://doi.org/10.2139/SSRN.2941774

Loureiro, S. M. C., Guerreiro, J., & Tussyadiah, I. (2021). Artificial intelligence in business: State of the art and future research agenda. *Journal of Business Research*, *129*, 911–926. https://doi.org/10.1016/J.JBUSRES.2020.11.001

Marsland, S. (2014). *Machine Learning: An Algorithmic Perspective* (2nd ed.). CRC.

Mayring, P. (2014). *Qualitative content analysis: theoretical foundation, basic procedures and software solution*.

Mayring, P. (2015). Qualitative Inhaltsanalyse - Grundlagen und Techniken. In *Beltz* (12th ed.). Beltz Verlag.

Monika, & Sharma, A. K. (2015). Venture Capitalists' Investment Decision Criteria for New Ventures: A Review. *Procedia - Social and Behavioral Sciences*, *189*. https://doi.org/10.1016/j.sbspro.2015.03.195

Patel, N. (2020). *Pitchbook European Venture Report 2020*. https://files.pitchbook.com/website/files/pdf/2020_Annual_European_Venture_Report.pdf

Patton, M. Q. (1999). Enhancing the quality and credibility of qualitative analysis. *Health Services Research*, *34*(5 Pt 2), 1189.

Petty, J. S., & Gruber, M. (2011). "In pursuit of the real deal": A longitudinal study of VC decision making. *Journal of Business Venturing*, *26*(2), 172–188. https://doi.org/10.1016/J.JBUSVENT.2009.07.002

Phillips-Wren, G. (2012). AI Tools in Decision Making Support Systems: A Review. *International Journal on Artificial Intelligence Tools*, *21*(2). https://doi.org/10.1142/S0218213012400052

Retterath, A. (2020). Human Versus Computer: Benchmarking Venture Capitalists and Machine Learning Algorithms for Investment Screening. *SSRN Electronic Journal*. https://doi.org/10.2139/SSRN.3706119

Retterath, A., & Kavadias, S. (2020). How to Hit Home Runs: Portfolio Strategies and Returns in Formal and Informal Venture Capital. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3527412

Rowley, J. (2012). Conducting research interviews. *Management Research Review*, *35*(3–4), 260–271. https://doi.org/10.1108/01409171211210154

Russell, S., & Norvig, P. (1995). *Artificial Intelligence: A Modern Approach* (1st ed., Issue 2). Prentice Hall.

Russell, S., & Norvig, P. (2016). *Artificial Intelligence: A Modern Approach* (3rd ed.). Prentice Hall.

Saurkar, A. v, & Gode, S. A. (2018). An Overview On Web Scraping Techniques And Tools. *International Journal on Future Revolution in Computer Science & Communication Engineering*, *4*(4), 363–367. http://www.ijfrcsce.org/index.php/ijfrcsce/article/view/1529

Sheehan, P., & Sheehan, A. (2017). *The Paradox of Experience*. https://etfpartners.capital/wp-content/uploads/2017/01/ETF-Partners-Investment-Insights-Decision-making.pdf

SignalFire. (n.d.). Retrieved December 27, 2021, from https://signalfire.com/

Smeets, M. R., Roetzel, P. G., & Ostendorf, R. J. (2021). AI and its Opportunities for Decision-Making in Organizations: A Systematic Review of the Influencing Factors on the Intention to use AI. *Die Unternehmung*, *75*(3), 432–460. https://doi.org/10.5771/0042-059X-2021-3-432

Sørensen, M. (2007). How Smart Is Smart Money? A Two-Sided Matching Model of Venture Capital. *The Journal of Finance*, *62*(6), 2725–2762. https://doi.org/10.1111/J.1540-6261.2007.01291.X

Stangler, D., & Kedrosky, P. (2010). Exploring Firm Formation: Why is the Number of New Firms Constant? *SSRN Electronic Journal*. https://doi.org/10.2139/SSRN.1585380

Tewari, A., Gabarro, J., Sole, J., Lapouble, B., & Montull, L. (2020). Artificial Intelligence Based Decision Making for Venture Capital Platform. *Lecture Notes in Business Information Processing*, *384 LNBIP*, 136–149. https://doi.org/10.1007/978-3-030-46224-6_11

Tian, X. (2011). The causes and consequences of venture capital stage financing. *Journal of Financial Economics*, *101*(1), 132–159. https://doi.org/10.1016/J.JFINECO.2011.02.011

Trocha, B. (2019). *Data-Driven VCs: How 83 Venture Capital Firms Use Data, AI & Proprietary Software to Drive Alpha Returns*. https://medium.com/hackernoon/winning-by-eating-their-own-dogs-food-83-venture-capital-firms-using-data-ai-proprietary-da92b81b85ef

Trunk, A., Birkel, H., & Hartmann, E. (2020). On the current state of combining human and artificial intelligence for strategic organizational decision making. *Business Research*, *13*(3). https://doi.org/10.1007/s40685-020-00133-x

Tyebjee, T. T., & Bruno, A. v. (1984). *A Model of Venture Capitalist Investment Activity*. *30*(9), 1051–1066. https://doi.org/10.1287/MNSC.30.9.1051

Venkatraman, V. (2017). *Digital Matrix: New Rules for Business Transformation Through Technology*. LifeTree.

Warnock, E. (2021, October 25). *How data is helping European VCs find their next hidden gems*. https://sifted.eu/articles/data-driven-sourcing-vcs-funding/

Weibl, J., & Hess, T. (2019). Finding the next unicorn: When Big Data meets venture capital. *Proceedings of the Annual Hawaii International Conference on System Sciences*, *2019-January*, 1075–1084. https://doi.org/10.24251/HICSS.2019.132

Wells, W. A. (1974). *Venture Capital Decision-Making*. Carnegie Mellon University.

Wiggers, K. (2021). *Gartner: 75% of VCs will use AI to make investment decisions by 2025*. https://venturebeat.com/2021/03/10/gartner-75-of-vcs-will-use-ai-to-make-investment-decisions-by-2025/

Zacharakis, A. L., & Meyer, G. D. (1998). A lack of insight: do venture capitalists really understand their own decision process? *Journal of Business Venturing*, *13*(1), 57–76. https://doi.org/10.1016/S0883-9026(97)00004-9

Zacharakis, A. L., & Meyer, G. D. (2000). The potential of actuarial decision models: Can they improve the venture capital investment decision? *Journal of Business Venturing*, *15*(4), 323–346. https://doi.org/10.1016/S0883-9026(98)00016-0

Zacharakis, A. L., & Shepherd, D. A. (2001). The nature of information and overconfidence on venture capitalists' decision making. *Journal of Business Venturing*, *16*(4), 311–332. https://doi.org/10.1016/S0883-9026(99)00052-X

Zider, B. (1998). How venture capital works. *Harvard Business Review*, *76*(6), 131–139.

Appendices

Appendix A Semi-Structured Interview Guide

VC currently using AI/ML or Quantitative Sourcing Tools	**VC currently NOT using AI/ML or Quantitative Sourcing Tools**
1. Categorization of Investor by Type and Specialization Questions	**1. Categorization of Investor by Type and Specialization Questions**
1.1. Name of VC Fund	1.1. Name of VC Fund
1.2. What is your job title and role in the fund?	1.2. What is your job title and role in the fund?
1.3. Approximately, what is your most recent funds total committed capital or assets under management?	1.3. Approximately, what is your most recent funds total committed capital or assets under management?
1.4. Do you target a specific stage, industry, or geography?	1.4. Do you target a specific stage, industry, or geography?
1.5. What are your most important decision criteria when deciding whether to invest? Rank important items/criteria (rank them by most important first)	1.5. What are your most important decision criteria when deciding whether to invest? Rank important items/criteria (rank them by most important first)
1.6. What challenges do you have within your decision-making process?	1.6. What challenges do you have within your decision-making process?
2. Status of Quantitative Sourcing or AI and ML in the Investment Process	**2. Status of Quantitative Sourcing or AI and ML in the Investment Process**
2.1. What is your perspective on AI/ML tools in VC decision making?	2.1. What is your experience and perspective on AI/ML tools in VC decision making?
2.2. Are you currently using AI/ML within your investment process?	2.2. Are you currently using AI/ML within your investment process?
2.2.1. If **YES**, please elaborate how and where AI/ML is integrated in your investment process?	2.2.1. If no, why did you decide **NOT** to use AI/ML for support in your investment process?
2.2.2. Why did you decide to use ML for support in your investment process?	2.2.1.1. Why did you consider or discuss the usage?
2.2.3. What was the initial trigger for your VC-Fund to explore quantitative sourcing opportunities? [GP, Market Research etc.]	2.2.1.2. What was the initial trigger for your VC-Fund to explore quantitative sourcing opportunities?

2.2.4. What data sources do you use in the respective stage? 2.3. Where did you experience the biggest challenges when implementing AI/ML tools? 2.3.1. What measures did you take in order to overcome those problems? 2.4. What are your internal requirements to a ML algorithm for its implementation? 2.5. How did technology improve your decision-making process? Rank important items/criteria (rank them by most important first) (Increase Speed, Increased Efficiency, Increased Return, Reduced Bias, Reduced Uncertainty, Augmented Investment Decisions) 2.6. How does AI/ML change the way you make decisions / execute your investment process? **3. Outlook** 3.1. Where do you think AI/ML can contribute most significantly to your investment process in the future? 3.2. How does your fund plan on making organizational changes in the future in order to achieve this? 3.3. How do you plan on using AI/ML for your investment process in the next 5 years? 3.3.1. Could you please specify on the specific stage of your process? *3.3.2.* If you could wish anything, what would your ideal technological support in the decision-making process look like?	2.2.1.3. What problems could it solve? 2.2.1.4. Why did you decide not to realize it in the end? / What are your concerns? 2.3. Where did you experience the biggest challenges in the discussions for AI/ML tools? 2.4. What are your internal requirements to a ML algorithm for its implementation? 2.5. What would get you to use such tools? **3. Outlook** 3.1. Where do you think AI/ML can contribute most significantly to your investment process in the future? 3.2. How does your fund plan on making organizational changes in the future in order to achieve this? 3.3. How do you plan on using AI/ML for your investment process in the next 5 years? 3.3.3. Could you please specify on the specific stage of your process? *3.3.4.* If you could wish anything, what would your ideal technological support in the decision-making process look like?

Appendix B Category and Coding System

Category	Coding Rule
C1: Decision-Making	ALL statements that relate the decision-making process of VCs.
C1.1: Challenges	ALL statements that relate the decision-making process of VCs AND challenges.
C1.2: Biases	ALL statements that relate the decision-making process of VCs AND biases.
C1.3: Criteria	ALL statements that relate the decision-making process of VCs AND criteria.
C2: Rationale & Trigger	ALL statements that relate to the organizational rationale and trigger of AI/ML implementation.
C3: Requirements	ALL statements that relate to requirements of VCs towards AI/ML tools.
C4: Data	ALL statements that relate to the topic of data (sources, types, categories, etc.).
C4.1: Co-Investor Data	ALL statements that relate to the topic of data (sources, types, categories, etc.) AND co-investor data.
C4.2: Third-Party Data	ALL statements that relate to the topic of data (sources, types, categories, etc.) AND third-party data.
C4.3: Data Platforms	ALL statements that relate to the topic of data (sources, types, categories, etc.) AND data platforms.
C4.4: Talent Data	ALL statements that relate to the topic of data (sources, types, categories, etc.) AND talent data.
C5: Automated Investment Process	ALL statements that relate to automation and technology used within the investment process.
C5.1: Market Situation	ALL statements that relate to automation and technology used within the investment process AND market situation.
C5.2: Systems	ALL statements that relate to automation and technology used within the investment process AND systems.
C5.2.1: Third-Party Tools	ALL statements that relate to automation and technology used within the investment process AND third-party tools.
C5.2.2: CRM	ALL statements that relate to automation and technology used within the investment process AND CRM.
C5.2.3: Web Crawler	ALL statements that relate to automation and technology used within the investment process AND web crawler.
C5.2.4: NLP	ALL statements that relate to automation and technology used within the investment process AND NLP.
C5.2.5: AI/ML	ALL statements that relate to automation and technology used within the investment process AND AI/ML.
C5.3: Overall Usage	ALL statements that relate to automation and technology used within the investment process AND the overall usage.
C5.3.1: Sourcing	ALL statements that relate to automation and technology used within the investment process AND sourcing.
C5.3.2: Screening	ALL statements that relate to automation and technology used within the investment process AND screening.

C5.3.3: Investing	ALL statements that relate to automation and technology used within the investment process AND investing.
C5.3.4: End of Usage	ALL statements that relate to automation and technology used within the investment process AND the end of usage.
C6: Business Value	ALL statements that relate to the business value that gets created using quantitative sourcing and AI/ML.
C7: Challenges	ALL statements that relate to the challenges for the implementation of AI/ML.
C7.1: Bias	ALL statements that relate to the challenges for the implementation of AI/ML AND bias.
C7.2: Automated Denial	ALL statements that relate to the challenges for the implementation of AI/ML AND automated denial.
C7.3: Interpretation	ALL statements that relate to the challenges for the implementation of AI/ML AND interpretation.
C7.4: Technology & Data	ALL statements that relate to the challenges for the implementation of AI/ML AND technology & data.
C7.5: Algorithm Aversion	ALL statements that relate to the challenges for the implementation of AI/ML AND algorithm aversion.
C7.6: Organization & Culture	ALL statements that relate to the challenges for the implementation of AI/ML AND organization and culture.
C7.7: Resources	ALL statements that relate to the challenges for the implementation of AI/ML AND resources.
C6: Business Value	ALL statements related to business value created using quantitative sourcing and AI/ML.
C8: Future of VC Investments Data	ALL statements that relate to the future of how VCs will make decisions.
C8.1: Data	ALL statements that relate to the future of how VCs will make decisions AND data.
C8.2: Competitive Landscape	ALL statements that relate to the future of how VCs will make decisions AND the competitive landscape.
C8.3: Commodity	ALL statements that relate to the future of how VCs will make decisions AND commodity of developing AI/ML.
C8.4: Market Intelligence	ALL statements that relate to the future of how VCs will make decisions AND market intelligence.
C8.5: Technology	ALL statements that relate to the future of how VCs will make decisions AND technology. .
C8.6: Due Diligence	ALL statements that relate to the future of how VCs will make decisions AND due diligence.
C8.7: Post-Investment	ALL statements that relate to the future of how VCs will make decisions AND post-investment activities.
C8.8: Future requirements	ALL statements that relate to the future of how VCs will make decisions AND future requirements.
C8.9: Organizational Changes	ALL statements that relate to the future of how VCs will make decisions AND organizational changes.

Appendix C Interview Insights

Interview IP01 – 11.2021

IP	Category	Transcript Segment
IP01	C1.3: Criteria	And then apart from the team, second branch is obviously how they fit into the market. And especially in the early stage it always looks like a crowded landscape. I think the number one thing you can always say for declining anything is saying it is a crowded market. Because everything is crowded, right. So then the question becomes, what do they really think to breakthrough. What is the product differentiation, so it doesn't have to exist today but at least do they know exactly which niche they are targeting?
IP01	C1.3: Criteria	And then the question that comes out, and that the third is, we had team and product now the third thing is going to be around is this vision that they have enough for us to really get a big outcome. And there is nothing wrong with companies that want to be bootstrapped, but if they really want to get our money and sign up to our terms, in the cycle that we are under.
IP01	C2: Rationale & Trigger	I think two things. The number one thing is the liquidity. So there is so much money in the market. And number two, things are going so quick. I mean unless I see what i actually need to lock at, I always stick to the party. It's almost like, if this founder is connecting at the same time to ten of my closest network people on LinkedIn. It means he is speaking to all of them. And I need to be in the game right now.
IP01	C2: Rationale & Trigger	But the fact that you can say that you are having this tool is a good story to the LPs, to the founders, and yes, sometimes you do catch a view things that you wouldn't have caught otherwise.
IP01	C3: Requirements	You shouldn't miss obvious things, but I guess it is impossible to see everything. It shouldn't be the idea to see everything. What AI can do for us is actually help me prioritize, understand who to really reach out to, which founder I might get a warm intro to. Our crawler might shows me, by the way, this founder there has already someone been in tough with in the fund. So you totally reach out to Christian, so it almosts sets up an automatic E-Mail, where I can just hit send to Christian with the top five ot the week, saying: hey can you give me an intro to founder one, two, three, four, five. So this is the exptectation on the system. Because things move so quick. It is such a personal relationship, and that why I sent you the article, is only so much that AI can replace this human touch with. And the same thing for warm intro.
IP01	C4: Data	But if I see there is some traffic on website traffic, also on ProductHunt, where you can see GitLab, numbers, stars, forks, all of these smaller pieces of data are aggregated.
IP01	C4: Data	So one he identified for me, pre-seed even, that there is a founder, he's getting some tracktion and the website i exploding and maybe he got included in YCombinator, he signed up at YC, or something else.
IP01	C4.3: Data Platforms	There is a couple of pieces now, so there is Crunchbase as official data platform people are using. I wouldn't necessarily call it AI, but almost like a dump of every kind of information they find. We subscribed to it on a licenced basis.
IP01	C4.4: Talent Data	It's almost like, if this founder is connecting at the same time to ten of my closest network people on LinkedIn. It means he is speaking to all of them. And I need to be in the game right now.
IP01	C5.1: Market Situation	I think more and more funds have started with like out-of-the-box platform solutions, as if in some sort of pre-excisting solutions they have tried to plug-in. I think what has happened now has been a shift. I think every VC right now is using affinity or is about to use affinity. And then they use definitley base to build around AI features, to detect anormalies, website traffic. So there are a couple of things that feed into the system. But then automatically gets linked into an action to-do list. There is a couple of pieces now, so there is Crunchbase as official data platform people are using. I wouldn't necessarily call it AI, but almost like a dump of every kind of information they find. We subscribed to it on a licenced basis. And then there is an internal view, I think we build one, Notion builds one, every fund I know builds one inhouse to some degree.

IP01	C5.1: Market Situation	So I guess everyone has started to realize we can tell about the fund, if we are selling it as "we are so modern, we are using - the technology we are looking at everyday - on a day to day basis ourselves". So VCs saying that they were using that technology, it made them stand out a little bit. And obviously founders as well, when they were like, you were pinged by our database. It made them feel good. It's part of a selling story.
IP01	C5.1: Market Situation	It totally get this AI is taking over and become more prevalent. The truth is, right now it is not yet, and no matter how big the fund is and how many resources you put in, the same is true across a lot of funds, it is not the most effective way yet to do quick deals.
IP01	C5.2.1: Third-Party Tools	We are trying to staying away from any third party providers. Because they change quite a bit over time. You always have to map it again, integrate again, it can fail.
IP01	C5.2: Systems	Good point, so I think, if you think about the workflow, right. So I come into work in the morning and I see my affinity tool, because thats the essential data point. And I think that the same for pretty much every fund today. So now, where it comes in, behind Affinity, there is almost like this black box, of AI capabilities, that is essentially helping us navigate the to-do list. So, it is essential coming into the place, where if i login, it helpes me move up and down my priorities. And tells me, you know what this company that you got on your tracking list just got a lot of traffic, the founder had a nother 50 of your friends connected on your LinkedIn, so it pushed it up on my agenda.
IP01	C5.2.3: Web Crawler	Yes, so you can cut it in different pieces. So the crawler, essentially reads through anything across Europe, even though to be honest we do 90% UK, western Europe, Germany. So it is a bit of a mix here, but then it kind of flags all things across Europe. Now for me I get a tailored view, only for FinTech, Insurance and Crypto. My colleagues are sperated, so they would only get things in Healthcare. But everyone of us is tracking their own sectors. So the crawler allocates the information for us individually.
IP01	C5.2.3: Web Crawler	There is a trend to really focus on really working continuously on the crawler to really be part of the core solution in the "black box". So apart from the email pings it gets centralized even more.
IP01	C5.2.2: CRM	I think every VC right now is using affinity or is about to use affinity. And then they use definitly base to build around AI features, to detect anormalies, website traffics. So there are a couple of things that feed into the system. But then automatically gets linked into an action to-do list.
IP01	C5.2.2: CRM	So if I see something happening, I need to tell the team after me to have a look. But for the ones I actually did as a deal, I don't think that there is any point in having an AI tool. I mean we have some new reporting tools, internal tracking. That field is really well covered.
IP01	C5.2.2: CRM	So what happens is, I have an Affinity login, you might have one at a different one, and we decide that we are so close friends, that I share all of my Affinity contacts with you and you share them with me. It is the most efficient sales machine. Better than any tool that i have seen.
IP01	C5.3.1: Sourcing	And tells me, you know what this company that you got on your tracking list just got a lot of traffic, the founder had another 50 of your friends connected on your linkedin, so it pushed it up on my agenda.
IP01	C5.3.1: Sourcing	Obviously I see things are picking of, now its time to get in touch. So yes prioritization, but it is also like pick me up the signals I wouldn't see on a day to day basis.
IP01	C5.3.1: Sourcing	So the truth is, most of the deals are not coming from AI. 95% are still coming through personal relationships. And I don't think it's going to change maturely, it's maybe going to change a little bit. This is the same for all funds.
IP01	C5.3.2: Screening	What AI can do for us is actually help me prioritize, understand who to really reach out to, which founder I might get a warm intro to. Our crawler might shows me, by the way, this founder there has already someone been in tough with in the fund.
IP01	C5.3.4: End of Usage	No, I mean I guess it is because it is so early. Seed deals are done in two phone calls. So one he identified for me, pre-seed even, that there is a founder, he's getting some tracktion and the website i exploding and maybe he got included in YCombinator, he signed up at YC, or something else. It takes me one or two calls and the deal is done. So I don't really need anything else, rather then establishing selling and just being in a good place with the founder at

		that time and the deal is done. In the aftermath I mean now if the company is part of the portfolio there is absolutely no need for me to have any need for me to know, how is the employee base growing.
IP01	C6: Business Value	Obviously I see things are picking of, now its time to get in touch. So yes prioritization, but it is also like pick me up the signals I wouldn't see on a day to day basis.
IP01	C6: Business Value	What AI can do for us is actually help me prioritize, understand who to really reach out to, which founder I might get a warm intro to.
IP01	C7.1: Bias	Last part, i think the systems are inherently biased the moment we set it up, because the people that i have in my network obvisouly look at excactly the same deals as me, right? My core network is fintech and crypto people. So just because all of them are connecting with the new kind of founder, it doesn't mean that we aren't overlooking other hot deals in Bulgaria. It just means my personal network is so biased towards one specific kind of founder, in on kind of geo, in one borader field.
IP01	C7.1: Bias	Does it reduce bias? I think it increases my bias, because I get flagged a lot of things that are in my inherent ecosystem. That we already have a touchpoint with, that we already know somebody, that are already in the core markets, that are part of our portfolio companies they are customers of. It actually works the other way around. It deprioritizes things that we wouldn't have maybe look at.
IP01	C7.2: Automated Denial	It's a pit of a pain, and you kind of decline it because you looked at it before and yet it is a totally different company.
IP01	C7.3: Interpretatio n	The hardest part is to understand which stage the company is at. So just because the company is getting a lot of website traffic, I have absolutely no idea how much money they would need. So there is a gap between something is happening, but I have no idea if its early, or growth. Or even for the multi stage funds, which pockets it would fit. So there is this allocation thing, where it gets allocated to early stage. But it turns out to be a 100 million round, so it should go to the growth team. So there is this back and forth allocation, but no crawler today can understand how much money the company is looking for.
IP01	C7.4: Technology & Data	And there is minor things where things get picked up the wrong way. I mean you know how many times websites are being sold again to someone else. If a company gets bankrupt today, there is no need to keep the website. After half a year a new company takes the website. Especially 5 letter, 3 letter and short accronyms of websites. So now my affinity is totally packed, because it still loads Crunchbase data from the previous one. But I have to overwrite it.
IP01	C7.4: Technology & Data	And in early-stage there is a thing calles stealth mode, where there is people that don't want to disclose what they do. And it's impossible to pick up who excactly is building what. And the AI can typically not understand who is the founder even though he might be registered on LinkedIn, with the signals they are getting through random websites. There is no way to connect the two dots, thats the whole logic of not disclosing on what the website kind of operating on.
IP01	C8: Future of VC Investments Data	If I could dream, I think what I would probably want to do is almost have an automatic sales machine, that the moment the system understands this is a really hot company there is a way to simplify the integration to communicate with those people. It's not enough to just have Affinity. In a CrunchBase you maybe have a founder name, but no E-Mail, no phone number, no contacts, I mean, in a perfect world, it would be quite an exclusive way to access that information. But this is a dream, it would never actually happen, but to actually understand how can I reach this person best, so that I'm creditable enough to whatever way to get in touch. Have an automated sales machine to then almost reaches out to that person to say, "Hey here is my Calendly, I'd love to talk to you." And I don't need to touch it. Right I have to check LinkedIn manually, and search for name, company, country. Sent him a message, get a message back, tell him again to please reach me by e-mail. (...) This is where AI can help a lot more, especially understanding a little bit of those, how can I make my sales part more efficient. Just because I flagged it, doesn't mean I contacted this guy.
IP01	C8: Future of VC Investments Data	So there is almost this lost in translation piece, that I think could massively help, for everything I didn't do. If the headcount is growing, they hired some fantastic new people, they are trending on XYZ, it won a price, it was at an event, the crawler saw it on five different event websites. So if I see something happening, I need to tell the team after me to have a look.

IP01	C8.9: Organizational Changes	We do have two people in the team that do nothing else than figuring out how to make the process slicker.
IP01	C8.9: Organizational Changes	So I think our internal effort is to develop our own in-house system. Get inspiration of others. Not going to lie. Don't forget, every fund has their own reporting, their ongoing field where they tracking things as.

IP02 – 11.2021

IP02	C1.1: Challenges	VCs suffer from a network bias, which can be a good thing or a bad thing. "I know a friend who knows a partner, who knows a friend that has a company, because I know this guy, then a trust is founder team." This is how the narrative usually is in an investor, which again yielded amazing results. You can see the Silicon Valley bubble that we have there, but yet exactly the same here. But in my opinion, from inequality, you have just amazing teams with amazing talent that do not know someone. Therefore, they get under the microscope of an investor later and it's too late and they have less chances of getting money.
IP02	C1.2: Biases	One is the directive about biases that I told you about. Biases on the long run harm investors. And we need to find ways to get rid of that fast. And that data offered the answer.
IP02	C1.2: Biases	Let's try to be less biased. Obviously, our recommendations, our references, it builds more trust. This is about trust. But to give higher chances to things that are not as exposed to be seized as a graduate from WHU in Germany, for example, it's something that we really do care, right. So we go beyond that. And this is especially important for emerging managers, in my opinion.
IP02	C1.2: Biases	Like a 2nd, 3rd, 4th generation fund. The natural has become quite big. If you're an emerging manager, one of your key issues is access to quality deals flow, right? That is usually eaten away by the competitor investors.
IP02	C1.3: Criteria	Market timing. What I mean by that is one thing that we find it out, that it doesn't matter how amazing the team is. It doesn't matter how amazing the product is. If the timing of the market is not there, a company is not going to fly, right. Obviously, the question is, what is exactly market timing is not just how big a market is, how much is growing. That's not what we mean. What we mean is whether there is a set of technological and consumer trends that make the product scalable now, instead of before or after. Right. So it's a systemic variable. And this is why it's hard to identify.
IP02	C1.3: Criteria	I mean, obviously team is probably another top factor. Right. You need a team that is resilient, making and going through an early-stage company has an enormous amount of ups and downs, and you need a team that is psychologically motivated, committed and able to go through the cycles.
IP02	C1.3: Criteria	We also found that there are obviously correlations with your background and your personal professional history, professional rather than personal and your latitude of success in making a new company. Right. It's something obviously you can look at. And perhaps one thing that we also noticing, but it's hard to measure. This is a bit softer is also the culture in the company, right.
IP02	C2: Rationale & Trigger	One is the directive about biases that I told you about. Biases on the long run harm investors. And we need to find ways to get rid of that fast. And that data offered the answer.
IP02	C2: Rationale & Trigger	The other one is purely operational efficiency. We manage now seven funds. One is now fundraising. It's a lot. And we are 35 people. Right. And the investment team is perhaps 25. So it's a lot of money to manage for a small team, you need to be efficient. So efficiency was really, can I reduce the time spent for a member of my investment team into sourcing and finding and analyzing deals? And the answer was yes. We also did our own back of the envelope calculations of how it was before and after. Obviously, these are estimates, don't take them as face value. But on average, we've been decreasing the time from sourcing to identifying a deal by five times, which is a lot if you think about our saved of my time.

IP02	C2: Rationale & Trigger	I mean the trigger was, I come from quant finance. I'm a data scientist, I did a PhD on that stuff, and I've seen all the fields and finance that quantity analysis may help you understand where a certain value of an asset is going. So when I was hired, there was okay, let's try to make some scores that was it right. But what I wanted to bring on the table was, there are ways, in my opinion, to model to understand where innovation is going and then you can harness that to define and elaborate investment strategies. This is a hard thing to do, but definitely doable. This is the long run. Short run, those two drivers that I told you about. But that's why we do it here.
IP02	C2: Rationale & Trigger	Let‘s try to be less biased.
IP02	C4: Data	Layer number one is composed of structured and unstructured data that tells us any piece of information about any company that we might find.
IP02	C4: Data	And unstructured data is mainly I mainly refer to. This is some news and articles to social media. Right. So this is a bit of first layer, obviously, to make sense of the structured data, you need more sophisticated approaches in natural language processing, which we do. This is the first layer.
IP02	C4: Data	The second layer is data mainly about events. Did a company launch a new product? Did a company sign a new partnership? Did a company raise money with whom? This is data that for the most part is unstructured. So you need to make sense of it generally from the unstructured data that you collect already on the first right. And this helps us, especially when you do not have access to a company's financial records, which is true for most of the countries of the world, except for Scandinavia. So this is a bit how we do it.
IP02	C4.1: Co-Investor Data	What we also analyze is we try to source companies, not just by looking at data sets. Also looking at who is invested in what, right.
IP02	C4.1: Co-Investor Data	Why is that important? Because over time, you get investors that get specialized to become the ones with privileged access to certain type of companies that we might be interested in. So we try to identify who does are and we try to build relationships with them. So in a way, this is also helping us escaping the network that you only stick with the people, you know, because then I need to expend my network towards that side. With those investors, with those teams in order to obtain investments that I'm interested in.
IP02	\C4.4: Talent Data	And this is something that with data, whether, as we were talking "teams background", whether that's some form of traction or information about previous investors or any other signal is something that helps us save a lot of time.
IP02	C5.2.4: NLP	So there are two layers that we use. Layer number one is composed of structured and unstructured data that tells us any piece of information about any company that we might find. Structured data might be databases such as CrunchBase and alike. We use all of those. And unstructured data is mainly I mainly refer to. This is some news and articles to social media. Right. So this is a bit of first layer, obviously, to make sense of the structured data, you need more sophisticated approaches in natural language processing, which we do. This is the first layer. The second layer is data mainly about events. Did a company launch a new product? Did a company sign a new partnership? Did a company raise money with whom? This is data that for the most part is unstructured. So you need to make sense of it generally from the unstructured data that you collect already on the first right. And this helps us, especially when you do not have access to a company's financial records, which is true for most of the countries of the world, except for Scandinavia. So this is a bit how we do it.
IP02	C5.2.4: NLP	NLP, this is quite a lot of stuff. On predicting which company is going to be more interesting or not, we have been doing that. We have been testing that. To be firm with you, not a lot of statistical results.
IP02	C5.2.5: AI/ML	By the way, the scorecards that we mentioned have been informed by predictive by statistical models. We've been working on the variables and the weights behind the scorecards are developed by making predictive models about company success. By the way. So scorecards are not coming out of my mind. They are coming out of a sample of data coming from past failures and successful companies, by the way. And this is something we continuously update.

IP02	C5.2.5: AI/ML	We use a lot of machine learning when making sense of the data and identifying what a company is working on.
IP02	C5.3.2: Screening	So, as you said, we screen for the investment strategy. Is this company doing something that fits the mandate [investment thesis] and then the additional screening is screening for potential. Right.
IP02	C5.3.2: Screening	On the screening side, what I mean by that is you have 1000 companies that fit the investment thesis. What are the top ten you need to talk to. Here again, data has been working very interesting for us in the sense that it's not a crystal ball. It doesn't tell us what is the company we have to sign a check on. But it tells us what are the companies that we should spend time on or the top 1000? There are maybe 20 with the ingredients that we want to look for.
IP02	C5.3.2: Screening	The screening part is rather important for efficiency, to be efficient from sourcing to really get to the teams that you want to really spend time with the due diligence site.
IP02	C5.3.2: Screening	And I think this is the one where not a lot of work has been done, because then you look at one company, you need to decide: are there any red flags here? What is it that we make decided about it? And here it opens up a little box, because also potential applications that you can have of data, statistical models to make a better understanding of a market, of a team, of a technology.
IP02	C5.3.3: Investing	Structured data might be databases such as CrunchBase and alike. We use all of those.
IP02	C5.3.4: End of Usage	I think it happens right after making the shortlist. As I told you, if you have 1000 potential deals, what are the top 20? I need to spend my time on. Those 20, this is where the traditional VCs work starts more, because you need to be the relationship with them. You need to know whether what looks like being a cool thing is not right. And vice versa. That's very start. So after this screening, when you have, like, a shortlist on interesting deals, that's where the manual work in a way starts. Okay.
IP02	C6: Business Value	The other one is purely operational efficiency. We manage now seven funds. One is now fundraising. It's a lot. And we are 35 people. Right. And the investment team is perhaps 25. So it's a lot of money to manage for a small team, you need to be efficient. So efficiency was really, can I reduce the time spent for a member of my investment team into sourcing and finding and analyzing deals?
IP02	C6: Business Value	We also did our own back of the envelope calculations of how it was before and after. Obviously, these are estimates, don't take them as face value. But on average, we've been decreasing the time from sourcing to identifying a deal by five times, which is a lot if you think about our saved of my time.
IP02	C6: Business Value	So definitely first will be increased efficiency. The second will be reduced bias, third, augmented investment, fourth, increased speed, fifth, reduced uncertainty.
IP02	C6: Business Value	No, it's hard to say we've been actively. I'll tell you how about half of our deals come from, at least from some of the funds come from our data sourcing strategy, right. In my own funds. The last deal we've made came just like that. Right. But the fund is too young, so I don't have numbers. I can tell you that it's cheaper to run an investment team because we are more efficient. I cannot tell you whether that's going to give me more returns in the end.
IP02	C6: Business Value	This is what is giving us a bit of an edge into looking at deals that nobody else is looking at. Having a contrary strategy
IP02	C6: Business Value	Because over time, you get investors that get specialized to become the ones with privileged access to certain type of companies that we might be interested in. So we try to identify who does are and we try to build relationships with them. So in a way, this is also helping us escaping the network that you only stick with the people, you know, because then I need to expend my network towards that side
IP02	C6: Business Value	The screening part is rather important for efficiency, to be efficient from sourcing to really get to the teams that you want to really spend time with the due diligence site.

IP02	C7: Challenges	I would say the main one is a bit of a strategic slash operational one to be frankly with you. A VC is not a technology company, right? It's not a product company. Yet, we have been trying to make a product, an internal product, right. My experience. And it is also the experience of all the VC funds that have been trying something similar is that this is not the priority. The priority is making investments, and this may cause some sort of misallocation of resources when it comes to that. If I have a data scientist, then I might need them also for a technical due diligence sometimes or to do something very short term, because this is my priority. So definitely one thing to do as far as I know, 99% of VC funds trying to do something with data suffering from, is this lack of focus. Because once again, this long term project that definitely yields results, but it takes time. There's very short term, high-pace priorities that come and distract you.
IP02	C7.5: Algorithm Aversion	You mean people who are skeptical about it? Yeah. I think it's also very important to add the skepticism because VCs suffers from an innovation trap. Why do I need to change the way I do things? If for the last 20 years, I've been making money this way. This is what people ask themselves. I think it's a challenge because you need to show that you are yielding better results or just as good results, but in a shorter time frame. And that was definitely a challenge at the beginning for us because I myself was not (...) it was an experiment for us at the beginning. We were seeing whether it worked or not. And indeed. So my answer to that is skepticism. As long as it's not close minded, is good. It's only healthy to overcome it with hardcore data results which come with patients. They will not come in six months of work.
IP02	C7.6: Organization & Culture	My experience. And it is also the experience of all the VC funds that have been trying something similar is that this is not the priority. The priority is making investments, and this may cause some sort of misallocation of resources when it comes to that. If I have a data scientist, then I might need them also for a technical due diligence sometimes or to do something very short term, because this is my priority. So definitely one thing to do as far as I know, 99% of VC funds trying to do something with data suffering from, is this lack of focus. Because once again, this long term project that definitely yields results, but it takes time. There's very short term, high pace priorities that come and distract you.
IP02	C7.7: Resources	Yeah. Some funds just don't commit resources. They might hire one person that becomes head of something, but nothing happens. It's like going into something with just half of your foot. Right. That's something that I've seen many companies suffering from.
IP02	C8: Future of VC Investments Data	I think right now we and other VCs will be focusing on the perceived low hanging fruit. Can I make stuff more efficient? Because this brings hard KPIs that then you can show to the rest of the partnership and say, this is what we do, what we do, right. Which is fair. I like it. Long term I think that will become commoditized. I think that if it works, as you say, I think this will just be a matter of if you can attract the right talent to make that happen. But at that stage in the long run, I think really what is still missing and what I really want to achieve this. And I was foreshadowing that is making a quantitative modeling of where innovation is going. Innovation is an emerging phenomenon in a complex system where you have talent that works on something, where you have resources flowing into something, where you have a need. Right. This triangle of things. It's a complex process that we have the tools to moderate the data for it and ways to make prediction on what kind of technology will impact the world in XX years. If you are able to do that, you are able to design entire investment strategies, create new financial products just because of that. So this is really beyond the sourcing, beyond the actual decision. It's really on the defining investment strategies correctly, not just based on what I think I know. That's one. The other long term thing, but I don't think it will be automated as much in the next ten years, is really the due diligence and final decision-making process, which is at the end of the value chain. Definitely statistical learning will automate quite a lot of it. But the risk, as with any other predictive model, is that you become biased to a certain type of people and all others. And that's something that I'm very afraid of.
IP02	C8.2: Competitive Landscape	Yes. I think that's part of a bigger picture. This is one of the tools that right now VCs are using to really fight each other for deals in a climate where there is too much capital around. Right. I think if this becomes commoditized, if, it will boil down to two aspects. One is speed to sign a check, which I mean, you see already right now. Right. With the, let's say context. And the second one is what you bring it to the company after you make an investment, which, again, is something that we see in the seeds of this already happening right now.
IP02	C8.3: Commodity	Long term I think that will become commoditized. I think that if it works, as you say, I think this will just be a matter of if you can attract the right talent to make that happen. But at that stage in the long run, I think really what is still missing and what I really want to achieve this.

IP02	C8.7: Post-Investment	And that's another thing that we haven't discussed, right? How can I use data to support my companies? There are huge challenges that we all suffer from. How can we can help them recruiting, finding the right sales leads? That's something the data can just offer so much. So that's probably going to be that once people get no longer excited about, let's use this stuff for sourcing. It would be: how can I use this stuff to help my portfolio companies? But that requires a product mentality, which I think, in my opinion, only very few funds right now have. One of which is probably SignalFire in San Francisco. I'm sure you heard of them.

IP03 – 12.2021

IP03	C1.3: Criteria	Okay. Yeah. So that's a very good point. And I guess you heard it at this early stage because we do, like pre-seed, seed and Series A. So basically, tickets between 250k and 1.5 million euro. So what it means the most important part by far is the team. But that's the most important part. And then the product when it comes to the product, it should be somehow proprietary. So, for example, we don't really focus on, like, network effects startups. There should be some progressive technology involved and then the market itself. So does it really make sense from the market perspective? Is the market big enough that's it really. But I would say the most important one is the team.
IP03	C2: Rationale & Trigger	To be honest, I wasn't at the firm when they first started using that because it was a few years ago already. So that might be important information. But from what I see, it's basically to distinguish ourselves from other companies. So even in our company presentation, we have this active tool or proactive for sourcing. We have it as one of the things that we have an edge on when it comes to other funds. So I would say the reason was really to first of all, the obvious one just to find more startups. Really. The second one is to sort of like, show it to our LPs that we have something that order funds don't have.
IP03	C2: Rationale & Trigger	I was just going to say our main LP is a bank. So basically, I don't think that that was anything from their side, but yeah.
IP03	C4.3: Data Platforms	But what I know what they use is, for example, Crunchbase for scraping the web, because something and I know that basically they just scraping CrunchBase, Pitchbook and all the big data providers.
IP03	C5.1: Market Situation	So actually, I would say it might even be the other way around. We actually want to sort of focus more on the partnerships with, especially when it comes to the Eastern part. So the Croatia, Hungary, Bulgaria in these countries because basically there isn't that many VCs. So there is not as much money, but there are very good opportunities. So actually, what we are planning to do more is to really exchange with them. And again, it's not for the automated thing. So I would say that the trend might be contrary to this.
IP03	C5.2: Systems	But what I know what they use is, for example, Crunchbase for scraping the web, because something and I know that basically they just scraping CrunchBase, Pitchbook and all the big data providers. Like you come as I mentioned earlier, you come you say, the geography, the revenue that you're expecting, the size of tickets they might be looking for. What the year it was funded, how much, how much money they need and this sort of stuff. So you just, like, add a bunch of criteria and then you will basically get a list of the startups with the contacts as well. And then what I think they are doing right now is they do sort of pre sales as well, meaning that they can call up the startups and do like the sort of the initial, like sales or in sales it's called sales-qualified leads, right. But qualifying the startups and then they just give you a list. It's basically "warm interest" already.
IP03	C5.2.5: AI/ML	So it's actually both. So it's finding good startups, but also then making a decision, as you said. So I see the future in a way that machine learning will actually learn how to quantify the startups much better than we can. Meaning looking at the right metrics for the right startups with the right business models in the sort of right sector. That's something that I think there is a good potential. And if you know how to do that, please do that. But yeah, I think there is definitely future.
IP03	C5.3: Overall Usage	I can actually look it up exactly. Let's say we've been through around 1000 startups and out of these give me a second. I can look it up. (...) I would say maybe 20% of that comes from this automated part. The rest comes (...) there is a good bunch that's coming just to info mail. There is a good part just coming from network VCs.

IP03	C5.3.1: Sourcing	So when it comes to really the sourcing, we do. So as I mentioned, we use company called [AI Sourcing Tool] sometimes because they are also our portfolio company.
IP03	C5.3.1: Sourcing	And basically how it works is that we just come with a set of criteria in our case that we focus on see. So we sure about it the stage of the startups. And then it basically gives us a list of startups that we can approach. That's it really. And to my knowledge, what they do is they basically just scrape the web using different sources and basically just information this way about the startups.
IP03	C5.3.1: Sourcing	So when it comes to sourcing, I would say most of the stuff is still happening manually. I'm not sure if you were asking about the automated part because there is also the part of the sourcing that's not automatic. And that's basically, as I mentioned, for example, we get the leads from other VCs that I've talked to or from conferences, obviously, or from different meetings or our mail as well. But it's not really automated.
IP03	C5.3.1: Sourcing	Otherwise, I think there is a big space for improvement when it comes to sourcing, but also when it comes to managing the portfolio startups.
IP03	C5.3.4: End of Usage	But what happens after is that when we have the list, the investment analysis, they usually just reach out to the startups. And then basically they have a set of 15 to 20 questions that they ask them. They have a chat with them and then they basically write the initial memo [investment memo]. But it's really just like the one for pipeline meeting. Then we have a pipeline meeting where we either agree that we don't proceed with the project or that we proceed with the project. And in that case, then investment managers basically take over. So when it comes to sourcing, I would say most of the stuff is still happening manually. I'm not sure if you were asking about the automated part because there is also the part of the sourcing that's not automatic. And that's basically, as I mentioned, for example, we get the leads from other VCs that I've talked to or from conferences, obviously, or from different meetings or our mail as well. But it's not really automated.
IP03	C5.3.4: End of Usage	To give you an example, a B2B [Business to Business] business model has completely different metrics to B2C [Business to Consumer] business model. But so that means that this sort of information is again, something that the analysis look up by himself and then just based on experience or on different benchmarks, say, if he's good or not. So things like customer question costs [CAC], LTV. And this sort of stuff is not something that the platform provides. So basically it's like the very high level information, but the rest, it's something that you need to figure.
IP03	C6: Business Value	So I would say maybe first and second. So first properly, like increased efficiency, because when it comes to searching for startups, and I think that's where it's pretty useful, meaning that it also increases speed. So I would say first "efficiency", second "speed". Then I would say that maybe the third one will be the augmented investment decisions, if you mean in the way that it sort of helps at a very early stage, like, additional source of information to make a decision. So they will probably be the third one, then the fourth one is probably "reduced uncertainty" in a way that you at least get a little bit of information reduced bias. Well, I don't think so to much extend, just because you still need to do lots of work, because before I form my opinion and also like your bias from the very start, I don't think that the data that you can find from an AI Tool is going to make difference when it comes to increased return. I don't know yet, because the investment procedure takes a few years. And you know you've been using that for or at least I know of this for a year and a half. So it's just still very early in the process. So there isn't any recent return whatsoever just because you would need to wait ten years to give some interest.
IP03	C6: Business Value	Because there's still going to be a bit of bias, even though it might sort of decrease as we come up with better tools.
IP03	C6: Business Value	Otherwise, I think there is a big space for improvement when it comes to sourcing, but also when it comes to managing the portfolio startups.
IP03	C7: Challenges	It's really the problem that every business model is very different. So the data you gather might work for one startup, but not going to work for another one.

IP03	C7: Challenges	So that's one thing, the other thing is, as I mentioned, the most important fact is founders, like the human part can't really be measured as well, how we think about founders is super important, right. And their track record. So I think these are the two things that are going to stay.
IP03	C7: Challenges	The biggest issue right now is that every startup is different. So if you want to, like, sort of quantify the performance, I guess there is a way, but it's just not as easy as it sounds.
IP03	C7.1: Bias	Because there's still going to be a bit of bias, even though it might sort of decrease as we come over with better tools.
IP03	C7.4: Technology & Data	And actually, it's a good point that you're making. But the problem is that the business model of startups are so different. That's not something that you can't really, or you could possibly. But, for example, the [AI Sourcing Tool] doesn't really provide. It doesn't really provide you with granular information on the business model.
IP03	C7.4: Technology & Data	To give you an example, a B2B [Business to Business] business model has completely different metrics to B2C [Business to Consumer] business model. But so that means that this sort of information is again, something that the analysis look up by himself and then just based on experience or on different benchmarks, say, if he's good or not. So things like customer question costs [CAC], LTV. And this sort of stuff is not something that the platform provides. So basically it's like the very high level information, but the rest, it's something that you need to figure.
IP03	C7.4: Technology & Data	And then I think that the main problem with these tools is that if you really want to have super granular data, that would really give you a very good idea of how they are doing. You would need to have, like, so many variables when it comes to the database. That's something that I haven't really seen, like, any tool that it provides. I think that's the main struggle there really. That the data is not tailored to the startups and their business model.
IP03	C7.6: Organization & Culture	What I also heard from other VCs is the fact that they invest in super innovative technologies, but themselves they are not really innovative. So sometimes it's hard for them to start using new tools. When it comes to onboarding, because they are the old school guys that just like to do what they have been doing over the last 20 years. I don't think in our case it was a thing.
IP03	C8: Future of VC Investments Data	But there is still going to be this human aspect of things. And I think it's like two levels. One is the fact that a human do the decision at the end. Most likely. Right. Because there's still going to be a bit of bias, even though it might sort of decrease as we come up with better tools.
IP03	C8: Future of VC Investments Data	So it's actually both. So it's finding good startups, but also then making a decision, as you said. So I see the future in a way that machine learning will actually learn how to quantify the startups much better than we can. Meaning looking at the right metrics for the right startups with the right business models in the sort of right sector. That's something that I think there is a good potential. And if you know how to do that, please do that. But yeah, I think there is definitely future.
IP03	C8.5: Technology	So I see the future in a way that machine learning will actually learn how to quantify the startups much better than we can. Meaning looking at the right metrics for the right startups with the right business models in the sort of right sector. That's something that I think there is a good potential. And if you know how to do that, please do that. But yeah, I think there is a definitely future.
IP03	C8.7: Post-Investment	Otherwise, I think there is a big space for improvement when it comes to sourcing, but also when it comes to managing the portfolio startups.
IP03	C8.9: Organizational Changes	I'm trying to do that. And it's one thing to sort of starting very easily because you have lots of groups when it comes to data analytics and trying to find the right metrics for different business models and sectors. And that's what I'm trying to standardize within the company. But then also when it comes to sourcing. I have an MVP of an app that aims to do something like this, but not really internally, externally. But for now, I think this excel spreadsheet with checklist is what we can do at this point in the near future. Maybe one thing that might be good to remember as well. I want to just highlight it because I mentioned it before, and it's making the VCs who have been in the industry for 15-20 years, making them to use new tools might be quite difficult because I guess they are just used to something. And that's what I hear quite often. It's like we are ever on the fact that we should use some tool to make it easier for us when it comes to sourcing, but also

		the portfolio management. But we've been doing this for years and sort of works out, so why to change. So I think there will also needs to be some push when it comes to this as well.

IP04 – 12.2021

IP04	C1.1: Challenges	The analysis are very qualitative and not quantitative.
IP04	C1.2: Biases	And we are doing the opposite. So we are investing from our feelings.
IP04	C1.3: Criteria	Well, absolutely. So team is, of course, one of them. And I believe if you already talk with some friends, all of them will tell you the same, then we evaluate the market opportunity if it's big enough for us, if there is a market trend pushing for the market in that direction, and if the timing is right, and then we evaluate the product, the technology that is being built, if the technology is robust enough, innovative and has competitive advantages in the market. So those are basically the three criteria that are the most important for us. Then if we have traction metrics and if we have other good VCs on board joining the round and also providing their value to the company, we are investing. It's always a plus.
IP04	C2: Rationale & Trigger	So there's a lot of room to grow in terms of developing technology, both to do the scraping of investment opportunities and also to assess and select the best investments in the next step. But why I think that nobody was able to accomplish this in the past, and there are some VCs doing some efforts in this direction.
IP04	C2: Rationale & Trigger	Well, it popped up like two years ago and we did some investigation. We looked in the market for tools that were doing the same. We didn't find any tool that was doing what we wanted. We've been using a software called Spector [tryspecter.com].
IP04	C4.3: Data Platforms	We are pushing information from, like, CrunchBase, DealRoom, Startup Lisboa, DealRoom, Portugal, DealRoom Spain.
IP04	C4.4: Talent Data	And it's not easy to evaluate the founding team with an algorithm. You really need to meet them. You really need to feel the chemistry between founders. You have to really feel that they have the fire in their eyes, that drive to push the project forward. So in this stage that we invest the seed stage, I don't believe it's possible to remove people from the equation and do investments with an algorithm, maybe at another stages, like a Series B, Series C, where numbers play a bigger role. It can be possible, but not in our stage.
IP04	C5.1: Market Situation	So there's a lot of room to grow in terms of developing technology, both to do the scraping of investment opportunities and also to assess and select the best investments in the next step. But why I think that nobody was able to accomplish this in the past, and there are some VCs doing some efforts in this direction. But the main issue here is the market is very opaque [unclear].
IP04	C5.1: Market Situation	Then there's this initiative from CB Insides, which is called Mosaic [https://www.cbinsights.com/company-mosaic], where they basically rate startups according to the information they have from them in terms of market management, timing, et cetera. And then there is this initiative from a fund called CircleUp. It's like a DTC [consumer product focused] fund, and what they do is they invest from another (...) it's like algorithm decision based investments.
IP04	C5.2: Systems	But what we wanted was to build a centralized platform where we can aggregate all our deal flow sources in one single platform. There was no tool to do this, and we have not the technical expertise to build it inhouse. So we went to some software houses to build it. At the end, we were able to hire a person to develop our tool. So we did it this summer. But as I told you, we are in the early beginnings of it, it's basically a web crawler where we have seven or eight information sources.
IP04	C5.2: Systems	It was three months project. The next step, and that's where AI and machine learning enters the equation and is basically helping us to do our screening and filter the startups that can be the most interesting ones for us, and they eliminate the ones that do not match our criteria. For example, giving you a very simple example. If there's one startup that already raised €5 million, they are not in our investment scope, because next round will be a ten €20 million round. And we are seed stage investors. It does not fit our investment criteria. That's one basic thing we can do. The next step is basically we have one column in our dashboard, which is for us to rate the startup according to our investment criteria. So what we want in the future is that we have an algorithm running behind this, and this algorithm learns from our classifications. So if the

		algorithm finds a startup similar to other startup that we gave a five star rating, probably it will be an interesting company for us to speak with.
IP04	C5.2.1: Third-Party Tools	Well, it popped up like two years ago and we did some investigation. We looked in the market for tools that were doing the same. We didn't find any tool that was doing what we wanted. We've been using a software called Spector [tryspecter.com]. I don't know if you know them, but basically they are integrated with LinkedIn, and they spot profiles, launching a company that have a good background. We use this tool.
IP04	C5.2.3: Web Crawler	And we have one single platform where we can see all this information. But that's a basic web crawler.
IP04	C5.2.5: AI/ML	It was three months project basic. The next step, and that's where AI and machine learning enter the equation, and is basically helping us to do our screening and filter the startups that can be the most interesting ones for us, and they eliminate the ones that do not match our criteria. For example, giving you a very simple example. If there's one startup that already raised €5 million, they are not in our investment scope, because next round will be a ten €20 million round. And we are seed stage investors. It does not fit our investment criteria. That's one basic thing we can do. The next step is basically we have one column in our dashboard, which is for us to rate the startup according to our investment criteria. So what we want in the future is that we have an algorithm running behind this, and this algorithm learns from our classifications. So if the algorithm finds a startup similar to other startup that we gave a five-star rating, probably it will be an interesting company for us to speak with.
IP04	C5.2.5: AI/ML	And I believe that's here where ML and AI can play a role, because when we are talking about investing, the team part appears. And it's not easy to evaluate the founding team with an algorithm.
IP04	C5.3: Overall Usage	First in the scraping stage [sourcing/identification]. So I think there's already a lot of tools you can use to do the scraping and find new projects. You have CrunchBase, you have Pitchbook, then you have national sources by country. I believe that at the end of the day you will not be able to fully automate this process and rely on automation 100% because this is a market where personal relationships are very important and there are good projects in sales mode that they do not appear in CrunchBase or other sources, but it's like 20% of the market, the other 80%. You can rely on these sources and automate your scraping process. And that's one thing that we are doing here at FUND NAME. So we are internally developing a platform where we have our deal flow sources and we are aggregating the information from those deal flow sources in one single dashboard. So imagine a dashboard where you have different columns, and those columns are fulfilled automatically through the deal flow sources. We are integrated and pushing information from.
IP04	C5.3.4: End of Usage	We can try to be more quantitative rather than qualitative when evaluating a project. But where I believe that technology will play a major role is before meeting one entrepreneur. Until that part, a computer or an algorithm or a software can access the same information as us, can access the information which is publicly available, can access the information from a pitch deck, for example. But after that you need to meet entrepreneurs. They will tell you a story. You need to have the feeling of the people.
IP04	C6: Business Value	What I would say is that what it would improve is our internal processes and the way we do our first step of filtering the projects we are going to need. So basically the final goal here is to fully automate the process before meeting one entrepreneur. After that, we can use some technology. We can try to be more quantitative rather than qualitative when evaluating a project. But where I believe that technology will play a major role is before meeting one entrepreneur. Until that part, a computer or an algorithm or a software can access the same information as us, can access the information which is public available, can access the information from a pitch deck, for example. But after that you need to meet entrepreneurs. They will tell you a story. You need to have the feeling of the people.

IP04	C7.3: Interpretation	And it's not easy to evaluate the founding team with an algorithm. You really need to meet them. You really need to feel the chemistry between founders. You have to really feel that they have the fire in their eyes, that drive to push the project forward. So in this stage that we invest the seed stage, I don't believe it's possible to remove people from the equation and do investments with an algorithm, maybe at other stages, like a Series B, Series C, where numbers play a bigger role. It can be possible, but not in our stage.
IP04	C7.4: Technology & Data	There is no information available, no reliable sources
IP04	C7.4: Technology & Data	So we are talking about private markets which are not the same as public markets, where companies have the obligation to publish their results, their data, their future plans. The main issue here is to assess information, but what can we do to go around this and where I see technology being applied to the industry.
IP04	C7.6: Organization & Culture	Well, my opinion is that the VC industry is kind of "stopped in time". So VC's are basically using the same processes, the same frameworks they were using 10 or 15 years ago. So I believe there's a lot of room to improve within the industry. Basically, VCs invest in tech startups with a strong innovation component. Well, robust products, et cetera.
IP04	C7.7: Resources	I would say that the most challenging part was that we don't have anyone in the team with a technical expertise to develop this tool, which is fine. But we didn't also understood the requirements to build it. Meaning, is it pushing information from CrunchBase easy or not? Is it developed developing this algorithm easy or not? How much time is going to take? So we talk with different people. Like, for example, CTOs of our portfolio companies, people more in the technical part. And everyone has a different opinion and a different way to build it basically. And well, it took us some time to really understand the requirements, the timings, the budget.
IP04	C8: Future of VC Investments Data	I believe that at the end of the day you will not be able to fully automate this process and rely on automation 100% because this is a market where personal relationships are very important and there are good projects in sales mode that they do not appear in CrunchBase or other sources, but it's like 20% off the market, the other 80%.
IP04	C8.4: Market Intelligence	Okay, first of all, it can point me the direction to look at. So having, for example, a market analysis tool that can tell me what are the sectors that are more hot. Where there is a gap between investment and new startups being founded, to have an idea about what are the sectors booming and that will grow at the fast pace in the future?
IP04	C8.6: Due Diligence	Then what I would say is in the assessment part. Help me evaluating every factor which is not human-related. For example, nowadays we look at metrics. That part of the process could be fully automated if I have, for example, comparison of metrics between all the projects that I evaluated in terms of growth rates, customer acquisition costs, lifetime value, et cetera. If I have that information and I can have a scoring of every project that I am evaluating, the only part of the work I would need to focus would be meeting the entrepreneurs and check if they have the skills to execute their vision.

IP05 – 12.2021

IP05	C1.1: Challenges	The key challenge here is to get a high quality funnel unless you get stuck with the typical AI of "garbage in garbage".
IP05	C1.1: Challenges	So for us, challenge number one, getting the best companies. With respect, if you go to the next stage of the funnel, which is typically screening, this is where we feel that those companies that have made it past the first conversation.
IP05	C1.1: Challenges	That's where you feel that is your thesis and your ideas about the company are right or wrong. But that's part of it. So if you want two pains, one good deal flow at the beginning, second, it's in the execution part.
IP05	C1.3: Criteria	We, when we look for companies, we look for exactly the same thing as any other VC. So we do invest with the objective to get the financial return on our investments. So financial return is also important to us. So we typically look at all the things like team, technology, traction, all of that. I would say our criteria are not different.

IP05	C1.3: Criteria	This is where we take a good look into that. We take a look at the team, we take a look at the technology, we take a look at the traction, the market size.
IP05	C2: Rationale & Trigger	We've seen it as a trend. It's been new. We've got work from other VCs that they are using AI. But after we've talked to them and see what kind of what are they using, we've come back to these questions and you seed it's not AI. So we have PitchBook and CB Insights like everyone. We just put an alert on whatever new companies that are added to the platform, and that's the AI that they have. It's the AI that we have as of today. Again, correct me if I'm being very simplistic, but so far this is what we've seen. So we are looking at it because we've seen other VC using it.
IP05	C3: Requirements	Yes. I actually expected a more curated insight from these companies. I expect them to say something along the lines of in the past five months, in the past week, we've come across these 20 new companies, and we are ranking them in accordance to what our model says are the ones with the best probability of success based on criteria X-Y-Z and whatever. This is, what I was expecting to see.
IP05	C4.3: Data Platforms	So we have PitchBook and CB Insights like everyone. We just put an alert on whatever new companies that are added to the platform, and that's the AI that they have. It's the AI that we have as of today.
IP05	C4.3: Data Platforms	Nevertheless, sourcing tools and let me go back to things like PitchBook, CBInsights, CrunchBase. I think they are extremely useful because they allow you to easily do significant amount of filters with respect to company location, company maturity amounts raised number of employees, estimated revenues they can really support and do fast track a lot of work, but you still need the human in the loop to interpret them. That is our vision as of today.
IP05	C5.2: Systems	So we have PitchBook and CB Insights like everyone. We just put an alert on whatever new companies that are added to the platform, and that's the AI that they have. It's the AI that we have as of today.
IP05	C5.2.1: Third-Party Tools	We had the meeting with ScopeAI. I think it was last Friday or yesterday. Not sure. I'm not the one managing it. And the meeting with RaizedAI, it was two weeks ago. So this is very fresh for us. And so we began to look, there's a trending about this. Let's see what they are doing. We got some referrals and we did our own web research. Okay. These are the companies providing the service. Let's see what they do to see if there's value added. We've come to a realization that what they do. We can do it ourselves.
IP05	C6: Business Value	I'm absolutely sure that they will use it to improve their own selection capabilities, algorithms if they have, to their other client base. I'm pretty sure that they will do it.
IP05	C7: Challenges	It happens quite a lot in VC. I may be wrong. I may be a bit old fashioned in that, but I do believe that this still is a work that will require a significant amount of human labor if you'd like or not so much labor, but at least human mind to interpret what you are actually seeing.
IP05	C7.4: Technology & Data	And I don't believe that there is enough information available online to assess how good the team is.
IP05	C7.4: Technology & Data	At least I haven't seen one that works so far, and they don't work because what I've told previously, you need large data sets. You need to have information about what works and what companies work and do not, and you need to train those algorithms based on that information that is not shared among funds, and therefore it's hard.
IP05	C7.5: Algorithm Aversion	I don't like the fact that I particularly do not like the fact. I'd say that's where my distrust comes from in having a tool doing screening for me with respect to decisions of three to four million euros.
IP05	C7.5: Algorithm Aversion	And two, you are letting someone else do your own work about decisions that will be in the two to 3 million euros in a portfolio just to give you an idea. We have 100 million capacity to invest in the next five years. So a two, 3 million or 4 million decision weighs a significant amount.

IP05	C8: Future of VC Investments Data	You actually need to meet them, and you need to understand their background story and how they met and interact with them. And get talked to the clients that they have been working to see how they interact with the clients, to have at least a bit of a judgment about team. That's not something that with available information. You contain a model, an AI model to tell you that they see, from the scale from one to ten, that this team is a ten or this team is a six or this team is one. So if you are not able to do that, you believe in that. I think the human dimension is very important for you to continue doing this investment.
IP05	C8.4: Market Intelligence	However, it might be interesting for AI, and I think that's not it's to anticipate investment trends as a whole, not so much companies, but I would call trends. But you can do Google analysis to see what kind of words are popping up at increasing their search rate or the number of views that contain any particular reference.
IP05	C8.4: Market Intelligence	I think it would be interesting to as much as you could anticipate those trends. And again, I reinforced general trends. I think it would be very useful for us to begin searching even before it begins.

IP06 – 12.2021

IP06	C1.1: Challenges	Which is using data in a more intelligent fashion, to try to spot this company in the first place, because honestly, we don't really have revenue data for many of these things. As you as you probably know, like all these major databases, Crunchbase, Pitchbook, all these things they don't really have, like balance sheet or income information that are accurate. It means that you got to find ways to proxy those information without the variables.
IP06	C1.1: Challenges	The final point is the market is becoming incredibly competitive and incredibly fast, which means that there are two major players, like smaller fish and big fish.
IP06	C1.3: Criteria	So it's very hard that you buy a company in Series A with less than 1 million in revenues. That's a pretty clear threshold in Europe, for the states [USA] it's a bit different, but in European markets, I think like one 1 million in the recurring revenues per year. It's pretty much like the accepted threshold for Series A, which of course, even there, which is a major issue with venture, generally speaking, as an industry, the main issue is that value that I told you 1 million. Sometimes you can stretch it for the right condition. Meaning if the company that you're looking at is like half a million revenues but it's growing like 50% month over month, probably like you don't care that it's not like a 1 million already, because it would get to 1 million, like in three or four months time. So it's still, like, good for Serious A.
IP06	C1.3: Criteria	If you are trying to understand before talking to a company, if that thing is really something that you want to interact with, then the conversation is completely different, which is why we do what we do, right? Which is using data in a more intelligent fashion to try to spot this company in the first place, because honestly, we don't really have revenue data for many of these things.
IP06	C2: Rationale & Trigger	I think that usually happens for two main reasons. Right. One is there is someone in the partnership, so that's always like an executive decision to partner level. There is someone to wakes up and says, hey, we should probably like, we are not doing enough with the data that we have or we are losing ground, or we might lose ground in the future if we don't do something today. Right. And this is pretty much what happened with the hedge fund industry, like, 15 years ago. Of course, venture is a bit different. Actually, it took a while to get there, but usually it's someone at a very high level that gets like a trigger. But I can't tell you what the bad trigger is because I have no idea.
IP06	C2: Rationale & Trigger	Okay. At the same time, I think that the other interesting thing is that right now we have the last three or four years, we got more means and more tools to track online footprint of companies and people. So long story short, we have more data these days.
IP06	C2: Rationale & Trigger	So we are getting right now at a specific moment in time where the critical mass of data is good enough to start these things. So I think it's a conjunction of two different aspects, a human one and a proper data one.

IP06	C2: Rationale & Trigger	The final point is the market is becoming incredibly competitive and incredibly fast, which means that there are two major players, like smaller fish and big fish. The big fish, big funds or funds with more of a solid record. They're looking at these things right now because they have the means. So usually if you raise, like, 500 million in a fund, you might have, like, a decent management fee to invest. Right. That usually means that you can get one person dedicated to these kind of things. And you do that because you want to keep your position and possibly not losing ground in the future at the same time. It's crazy because the smaller funds are the ones that really need these things the most, because you want to get up to speed with bigger funds. In the shortest amount of time possible. So you really need these capabilities, but you cannot afford it, because honestly, paying someone only for that thing, it's crazy. Let's say, you have less than 100 million in AUM, they will probably tell you that they prefer to hire an associated or principle. Then, like, a data guy, that makes sense because you don't really have massive cost base. So there is always this conundrum between people that really need data, like the smaller fund, but they cannot get it. And people that don't really need that much, which is the bigger funds. But they're getting it right now. So potentially, this gap between big and small can diverge a lot and get much larger, unless the small funds find a way to get that data DNA from day one.
IP06	C4: Data	And if the company is a consumer company, you might try to understand how many people are downloading the app, how many people are paying things within the app or using the app on a monthly, weekly, daily basis and trying to build a model that tells you where the company is going or how much is growing and trying to infer from there the revenues level.
IP06	C4: Data	You might get more creative if you want. So you can study tech stacks. You can study GitHub Stars and Forks. Of course not everything applies to all the companies. It really depends on the company that you're looking at. So GitHub stars as I just mentioned, it applies much more to open source companies than to others.
IP06	C4: Data	That could be like, easier right now. You can track how many people are "forking" the thing, how many people are getting their repos [Repository] and using gate and all those kinds of things.
IP06	C4: Data	And you can expand this type of thinking, like in multiple ways. They could be done, like an academic level, Labs, if you're making a deep tech. And of course, you can attach on top of this, like patents data, you can attach software data, so GitHub, Tech Stack and all those kinds of things. So yeah, this is pretty much what we are doing these days. It's like putting together different data sources, like in the same format and having signals that automatically trigger if something happens.
IP06	C4: Data	So let's say you were mentioning media and news and all those kind of things. We have a system for that. But it's not something that we build. It's something that we are integrating on top of it from another source.
IP06	C4: Data	So everyone is going to use, like, ten different data sets and patent and GitHub all these kinds of things to scout companies.
IP06	C4.3: Data Platforms	So we've been using, like, a bunch of different data from traditional data sets, like Crunchbase, PitchBook, Dealroom, all these guys.
IP06	C4.4: Talent Data	More esoteric kind of things like talent data. So like checking whether someone is changing status or LinkedIn. Of course, this sounds like trivial, but made it a scale and made it with specific features. It make it be more difficult because we don't want to check everyone that changes status to LinkedIn with everything. Of course, you want to check for people that change status with new tech companies or stealth companies in software that has specific background, specific characteristics, specific features. Right. So of course, if you are leaving uni [University] or leaving your fund or your company to get and start a bar in Thailand, that's probably not for us either way. So we don't want to waste time understanding that thing because it's completely outside our business model, right.
IP06	C5.1: Market Situation	This is what many funds are doing. And the reason why it is so, is because that's the low-hanging fruit for everyone. It's probably the thing that has the less technical complexity is because doing, like, an automated DD tool or screening is actually much harder, like putting there together, like finding signals. It's not as hard as other things. And it's also the one that could prove to have value in a shorter amount of time, because potentially, if the system like, let's say that you actually put the system together, like in a week and getting more data into one place and you find straight away a company potentially depending on your fund, it might take,

		like a month to invest in the company. So in two months time, you see real-life results from what you're doing, which is pretty impressive. Again, this is the best case scenario.
IP06	C5.1: Market Situation	So if you look at new funds like Moonfire or Inreach Ventures, these guys, they got in the partnership, one person that was a data tech guy. So Ben in Inreach Ventures or Mike and Moonfire, they are tech guy. Right? So they are partners and they are tech guys. So they actually build infrastructure from day one to be a data driven fund. In that case, it's different because it's really like a partner level thing. And you get the data DNA so you can cover that gap with bigger fund but small funds in a traditional way and trying to put data afterwards. I think that I probably going to be left behind at some point.
IP06	C5.1: Market Situation	The first one is that, if we were like in equilibrium two years ago and right now we are moving towards a new equilibrium stay in five years. As you were saying, everyone is going to do the same things. Right. So everyone is going to use, like, ten different data sets and patent and GitHub all these kinds of things to scout companies. And you are completely right. That probably like at some point in time, the advantages of using this might erode and might decay. At the same time, the problem is that if you wouldn't be invested in that, you would be very much behind. So it's not about getting benefits or advantages from using these tools. It's about not getting disadvantages from not using that thing. Right. So it's being like, at the bare minimum of what everyone is looking at.
IP06	C5.2.1: Third-Party Tools	So that's probably going to be different for different funds. I try not to build these kinds of things internally, but to use it from some external providers, if you can.
IP06	C5.2.1: Third-Party Tools	So let's say you were mentioning media and news and all those kind of things. We have a system for that. But it's not something that we build. It's something that we are integrating on top of it from another source.
IP06	C5.2.4: NLP	NLP guy, a competitive vision guy and doing like, a bunch of different experiments.
IP06	C5.3: Overall Usage	So first of all, we use it mostly for the early stage team. Why? Because, honestly, growth [Growth Fund] is still too early.
IP06	C5.3: Overall Usage	We are planning to use it more for due diligence and execution for the growth team. But this is still something that we are tuning these days.
IP06	C5.3: Overall Usage	Listen, everyone is looking at data in due diligence, but I don't think that we are doing anything, which is incredibly AI-driven or whatever it is for DD purposes. Like that thing is actually very much classic analytics stuff. So like benchmarking things, getting everything like in the right format. It's not anything like super innovative. I think that it's very well structured in terms of business processes. And the way we automate things like due diligence is not really that much.
IP06	C5.3.1: Sourcing	For the early team, it is mostly sourcing. It is a bit of screening.
IP06	C5.3.1: Sourcing	So we got a bit smarter, I think, sourcing and screening. So sourcing means, as I was telling you before, getting as much data as possible from different sources, different things that can help you spot companies that you wouldn't see otherwise. Right. So the major idea there is identifying things that are relevant for us, or maybe things that we didn't know that they were relevant, but they could actually tell you: "Okay, this is like a weak or strong signal. You should reach out to this company." That's like the bottom line of what happened.
IP06	C6: Business Value	I think it's mostly speed, efficiency, coverage.
IP06	C6: Business Value	Overall, I don't think that we are at the point where our decision process is improved by these tools. I think that we simply get to see more in a much faster way in a much more efficient way. But eventually the final decision processes. It's a more informed decision, right. Because you have more data points, more things like to base your analysis on. But I think that simply like a consequence of having more information is not really that the final decision is better or worse. It's simply more complete and more grounded.

IP06	C7.1: Bias	One thing is the funds that I mentioned before, like Inreach and Moonfire. They are fantastic data-driven fund, but they are much more earlier than us, so they usually do pre-see and seed stage, which is different than looking at what we look at Series A. Which means that actually, instead of eroding that advantage, it could be used both ways. They can source companies. We can actually tap into the companies they source, like their portfolios and getting already like a different level of signaling. So it doesn't go like in a direction saying: if someone is going to do the same thing, it's probably going to get not useful. That might happen. That's the reality. It could definitely happen. Then at some point in time, everyone is doing the same things and no one is really finding anything new.
IP06	C7.6: Organization & Culture	So usually this happens, like many other funds. They're not like a real team. It's usually one or two persons, which means that building all the crawlers, building, all web scrapers, building NLP or computer vision for scanning decks. It is something that makes a lot of sense by maintaining those things is a full time job by itself, which means that you got to sacrifice other things.
IP06	C7.7: Resources	So usually this happens, like many other funds. They're not like a real team. It's usually one or two persons, which means that building all the crawlers, building, all web scrapers, building NLP or computer vision for scanning decks. It is something that makes a lot of sense by maintaining those things is a full time job by itself, which means that you got to sacrifice other things.
IP06	C7.7: Resources	But there is always, like, one or two people that are dedicated to this and not really like, six person teams that you can get like, an NLP guy, a computer vision guy and doing like, a bunch of different experiments.
IP06	C8.2: Competitive Landscape	It's crazy because the smaller funds are the ones that really need these things the most, because you want to get up to speed with bigger funds.
IP06	C8.2: Competitive Landscape	So potentially, this gap between big and small can diverge a lot and get much larger, unless the small funds find a way to get that data DNA from day one.
IP06	C8.2: Competitive Landscape	I think it's not completely true either way, because, first of all, not everyone is using the same type of data sets. Not everyone is looking at the same companies. Not everyone has the same investment thesis. The reality is that I know that you already have been speaking to other funds so [INVESTOR NAME] and other people, even in that case, even if people in VC, like doing data, like they talk a lot with each other. Right. So the reality is that even if I share everything I do with [VC FUND NAME], or [VC FUND NAME] or [VC FUND NAME] or anyone else, there is very clear possibility that they wouldn't be able to replicate everything in the same way.
IP06	C8.2: Competitive Landscape	And two, even if they could, they wouldn't be like, as useful to them as it is useful to us. Because there is first, a different investment thesis overall, and second, there's always like a degree of customization and personalization depending on your team
IP06	C8.2: Competitive Landscape	One thing is the funds that I mentioned before, like Inreach and Moonfire. They are fantastic data-driven fund, but they are much more earlier than us, so they usually do pre-see and seed stage, which is different than looking at what we look at Series A. Which means that actually, instead of eroding that advantage, it could be used both ways. They can source companies. We can actually tap into the companies they source, like their portfolios and getting already like a different level of signaling. So it doesn't go like in a direction saying: if someone is going to do the same thing, it's probably going to get not useful. That might happen. That's the reality. It could definitely happen. Then at some point in time, everyone is doing the same things and no one is really finding anything new.
IP06	C8.2: Competitive Landscape	So even if you find all the companies, like all the best companies in this world, it doesn't mean that you get access to those companies, like in the first place.
IP06	C8.2: Competitive Landscape	It means that eventually, even if the starting point is pretty much the same for everyone, the ending point will be completely different. That's the major twist there.
IP06	C8.3: Commodity	That probably like at some point in time, the advantages of using this might erode and might decay. At the same time, the problem is that if you wouldn't be in invested in that, you would be

		very much behind. So it's not about getting benefits or advantages from using these tools. It's not getting disadvantages from not using that thing.

IP07 – 12.2021

IP07	C1.1: Challenges	And not just relying on networking connections or just signals, manual signals from other pieces and companies.
IP07	C1.1: Challenges	So I guess in that respect, we're more so creating these models to aid and kind of make time of the analysts and the associates more valuable so they don't spend a lot of time trolling through like hundreds of thousands of deals, and we're using it to cut out the noise at the beginning.
IP07	C2: Rationale & Trigger	(...) effort of [FUND NAME] to become more data-driven and structure more on the engineering side.
IP07	C2: Rationale & Trigger	The very beginning came like two or three years ago, but the time it was harder for the investment team to have proper communication with founders. And it was old tech like un-automated stuff. So that's where we actually had a push for it. Of course, because we want to give a parallel stream to the investment team with technology and the adherent solution, which is necessarily going to improve decision-making and deal flow possibilities. So to increase the quality and the quantity of our deal flow.
IP07	C4: Data	So what we do, for example, we rely on third-party services to get deals and company information. For example, we use PitchBook, CrunchBase. We can use LinkedIn, we can use Twitter, any external social and non-social platforms. And so we have the list of all the deals that are happening broken down by countries, thesis, financial indicators, and other columns and fields that we are interested in. So we have a in-house database were we store all those deals happening every week. And first of all, we show all possibility to the investment team to see all these deals. They can filter by whatever they want to filter by, all the attributes. And that's the basic engineering solution we provide.
IP07	C4.2: Third-Party Data	For example, we are in contact with another third party service, which is going to provide us, hopefully with some more additional data points, which are not strictly deal data or financial data. But there are more signals about founders, for example, because it's one topic we are really interested in.
IP07	C4.4: Talent Data	But there are more signals about founders, for example, because it's one topic we are really interested in. Because if it's basically former employees of top companies, top consultancy companies like McKinsey, PwC staff, probably they are going to be more successful than the average.
IP07	C5.1: Market Situation	As you know, the VC industry is still quite behind on the data-driven and AI & ML engineering site.
IP07	C5.2: Systems	And then we started building both engineering applications and software for us and machine learning solutions for the investment team.
IP07	C5.2: Systems	It's called "[VC Fund] Discovery Dashboard", for example, is where we, as I told you earlier, we gave them the possibility to manage the deal flow, pre-investment deal flow from a proprietary platform. And they can filter by any attribute that the third party data provider gives us, and they are able to look at the status in our CRM. So the deal flow status of those companies in our CRM as well as joined data that we got from other data providers. So we join basically what we have internally what we get from external parties. We join the data together and we give a full picture to the investment team, so they don't need to jump between different platforms from different services. They just have everything in one view, and they can easily do CRM operations directly from that platform. That's going to speed up.
IP07	C5.2: Systems	They want to be able to only see what they care about. So we pre-filter that, then they can see the status of those selected companies in our CRM, how they are doing what data we have about them, and that you can just change the deal flow status and attributes so that we can reach even more the data.

IP07	C5.2.5: AI/ML	Then on the other side, that's how we are using ML, AI, data to improve and to speed up the process for analysts and associates. It's where we are developing predatory machine learning solutions, where we actually know and learn what investment team likes, what's the good fit for [FUND NAME]. And so we filter. So we don't show all the deals to the investment team, but we already pre-filter it. Only the interesting ones. According to our model, that's one of the solutions they are using. As you can imagine, it's very tough because we need to learn what [FUND NAME] likes and whats the proper deal ecosystem and startups world.
IP07	C5.2.5: AI/ML	And from that perspective, maybe our use of ML and what we'd like to do with it is not necessarily in the actual investment decision at the end game, but at the start and helping with the filtering. So to help cut out the noise from all the rubbish companies that we don't want to look at and just go straight into find the ones that will fit [FUND NAME] and ones that we want to look at. So I guess in that respect, we're more so creating these models to aid and kind of make time of the analysts and the associates more valuable so they don't spend a lot of time trolling through.
IP07	C5.3: Overall Usage	On one side we have the engineering side where we provide the investment team with applications both and other software just to make the deal flow and the due diligence and other investment flows easier and faster. And on the other side, we are working on machine learning and AI solutions to improve the decision making of investment team and to help them to become more data-driven. And not just relying on networking connections or just signals, manual signals from other pieces and companies. So that's what we are doing. We are working on different projects at the moment. As I told you, they vary from more like mostly engineering proper applications and web development. And then we have some more automations. Like deal flow automation and another one with actual intelligence that we are creating. So proprietary machine learning models to have the investment team making better decisions and having better signals.
IP07	C5.3: Overall Usage	So we use data basically in two main processes, of course, in sourcing and then in portfolio management. So after we invest in companies, so you're more interested in the sourcing part, right. In the sourcing process, the process from the investment team is usually structured in this way for investment in the associate, senior associate, principal and partners.
IP07	C5.3: Overall Usage	And then the analyst actually do the manual search. And that's where technology and data can intervene. And what we are working now, for example, in [FUND NAME].
IP07	C5.3: Overall Usage	Like hundreds of thousands of deals, and we're using it to cut out the noise at the beginning. And then from there on it's still a very similar investment process doing regular due diligence. And obviously we'd like to help out with that as well. In addition, we want to provide companies that human research probably wouldn't get right, because there is too much data, too much deals happening, too many companies. And the analyst can't look at all the companies in a few weeks, in a few days or so. But we are trying to pre-select them and actually take them to their intention because otherwise they would be lost or they would be missing opportunities and stuff like that.
IP07	C5.3.1: Sourcing	So its a pre-defined campaign. So the analysts are going to look into that framework. That's called campaign in general. So they are decided by the more senior employees of the VC, partners, principals, or senior associates. And then the analyst actually do the manual search. And that's where technology and data can intervene. And what we are working now, for example, in [FUND NAME].
IP07	C6: Business Value	And on the other side, we are working on machine learning and AI solutions to improve the decision making of investment team and to help them to become more data-driven. And not just relying on networking connections or just signals, manual signals from other pieces and companies.
IP07	C6: Business Value	Then on the other side, that's how we are using ML, AI, data to improve and to speed up the process for analysts and associates.
IP07	C6: Business Value	As you can imagine, if we get a very good model, that's going to be a bit of a game changer for the investment team. Because we exclude all the deals that we know that they are not interested in. And the investment team is focused on a subset, which is most likely very interesting deals for them. So they are going to spend much less time on, wasting less time on reading all the deals. They are speeding up the investment process. They are speeding up the due diligence. They have much more time to do research or to focus on specific companies and work on those.

IP07	C6: Business Value	So to help cut out the noise from all the rubbish companies that we don't want to look at and just go straight into find the ones that will fit [FUND NAME] and ones that we want to look at. So I guess in that respect, we're more so creating these models to aid and kind of make time of the analysts and the associates more valuable so they don't spend a lot of time trolling through.
IP07	C6: Business Value	In addition, we want to provide companies that human research probably wouldn't get right, because there is too much data, too much deals happening, too many companies.
IP07	C6: Business Value	But we are trying to pre-select them and actually take them to their intention because otherwise they would be lost or they would be missing opportunities and stuff like that.
IP07	C6: Business Value	So we join basically what we have internally and what we get from external parties. We join the data together and we give a full picture to the investment team, so they don't need to jump between different platforms from different services. They just have everything in one view, and they can easily do CRM operations directly from that platform. That's going to speed up. The goal is to speed up the process and to let them focus only on the companies they are interested in. Because maybe two analysts they don't work on the same thesis, on the same countries, on the same company sets. One analyst works in UK and focuses on UK companies. Another one works in Berlin, and she only focuses on Berlin companies. One is only FinTech. One is only DeepTech, so they don't want to see all the companies in general. They want to be able to only see what they care about. So we pre-filter that, then they can see the status of those selected companies in our CRM, how they are doing what data we have about them, and that you can just change the deal flow status and attributes so that we can reach even more the data.
IP07	C7: Challenges	And also state of mind of the investment team because they are not tech guys, most of the times, so they usually don't work with that perspective. We are working on trying to enforce with technology, solutions but also with persuasion that technology and data-driven solutions are going to help them. So that they can fill in more data and give us more quality data.
IP07	C7.4: Technology & Data	As you know, the VC industry is still quite behind on the data-driven and AI & ML engineering site. They were relying a lot on external party services, data providers and stuff like that.
IP07	C7.4: Technology & Data	For example, we have an additional machine learning proprietary tool which is trying to translate what PitchBook provides us with a [FUND NAME] language. So of course, we have an internal language in describing, in classifying companies, while PitchBook and CrunchBase and other data providers speak in another language. Because they just have fields. And for example, they classify the industry of a company based on some criteria. But for us, those criteria might be different. Right. So we are trying to translate what we get from the external site to internal language. For example, the thesis of a company. PitchBook doesn't provide really a thesis.
IP07	C7.4: Technology & Data	So the hardest part for us is to actually avoid parse data and filling all the data points for any deal that we have, for any opportunity that we look at. And data quality is the biggest problem that I guess every VC is facing. Because if in the past, when you started to have a proper data strategy, data structure, then after it becomes very hard to switch.
IP07	C7.6: Organization & Culture	And also state of mind of the investment team because they are not tech guys most of the times, so they usually don't work with that perspective. We are working on trying to enforce with technology, solutions but also with persuasion that technology and data-driven solutions is going to help them. So that they can fill in more data and give us more quality data.
IP07	C7.7: Resources	As you know, the VC industry is still quite behind on the data-driven and AI & ML engineering site. They were relying a lot on external party services, data providers and stuff like that.
IP07	C8.1: Data	And then yes, our plan. Our strategy in the future is to get as many as external data providers as possible and join all the data together somehow so that we can give the investment team a massive broader view. And they can then decide what they want to look at. So the goal is to get as much data points as you can, pull them internally. Do ML so enrich them even more with the [FUND NAME] view. And then provide all these data points to the investment team, and they are going to be able to filter by whatever field they want.
IP07	C8.5: Technology	There are two different strategies and efforts, but we are going in parallel with them because, of course, we want to learn because we are going to collect data points and KPIs on our portfolio companies. So we want to keep track of that historical data, financial data, how they are doing, how [FUND NAME] is supporting them, they want to use them in the future, also for sourcing purposes.

IP07	C8.7: Post-Investment	Because we can use those data points, that amount of data to understand what went good, what went bad on those companies and for some useful data that we can use for the sourcing efforts. But we don't have like, of course, mainly we're sourcing efforts, but we are also constantly working on the portfolio management, KPI management and stuff like that.

IP08 – 12.2021

IP08	C1.1: Challenges	How am I with a team of six people going to nurture them in a way that I can build a relationship to them and be a good option as an investor? Because it's not enough for me just to know that they are good or they might be good. I need to maintain that relationship because I don't know which and I don't have time to meet 200. Maybe in two years. I have time. But if it is just a matter of six months, one year, it's going to be difficult.
IP08	C1.3: Criteria	We look into metrics. So we look into hard data. We look into if the companies have 1 million in revenue to start, and if their growth has been at least two X, it's kind of the minimum for us to be interested because that's, based on all the analysis that we've done, that's where the company will be at the stage of development, where the risk that we want to take is appropriate for our investment strategy. And I think with time we will spend less we can extend our criteria and complement these with soft data. I'm sure that funds all over Europe that are in their fourth funds, fifth fund. They already have seen and invested because it's not just about seeing, but you have to invest. You have to make the decisions to know what works in terms of soft data and not just hard data. But we don't have that history. And so we pay attention to that data to start looking into an investment.
IP08	C2: Rationale & Trigger	Well, it was related mostly because of my experience at Techstars. We had a lot of data, and we tracked a lot of data about our decisions. And so one of the first things that I shared, we always make team based decisions. But one of the things that I thought would be a great learning to have is to use technology to document our decisions.
IP08	C3: Requirements	So I think efficiency would be important, is always important, but will be more important. Two years from now, we can put hands working now and will cover almost everything. No one never covers everything, but I think it's more the fear of missing out, that you might miss something. So I would not call it augmented decision because it's not about covering more. It's not about bias, it's more about making sure that we cover everything. And by covering everything, we found a leading indicator, that of someone that might become successful in the future. But we are not worried with efficiency at this moment. The tool will not be about again, not about efficiency. I think it will be about making sure we don't miss anything. I don't know how to label it.
IP08	C4: Data	We already covering the accelerators, the incubators, et cetera, and those ones.
IP08	C4.1: Co-Investor Data	They have this super powerful technologies thats proprietary that they develop, and they get a lot of data where they capture a lot of deals. They share that with us.
IP08	C4.3: Data Platforms	Are people looking into CrunchBase as well to understand what's happening. I've seen that more happening in terms of knowing that someone was funded so that you keep tracking or keep track of funding. That's also important we've been able to do, as you mentioned, more manually, where we go to PitchBook with someone and do these kind of one-time analysis about historically what happened in the last six months, what happened in last year in terms of funding.
IP08	C4.4: Talent Data	But I do think there are things like definitely paying attention to the LinkedIn, the job titles. There the changes in the job posts, the number of employees growing, and AngelList, as well that work around talents. I think it's for sure it is helpful.
IP08	C5.1: Market Situation	What we have done, though, that we invested, we are LPs in other funds that do that. And so we are just learning to see what others are doing, and we cannot build everything at once. So one day we will get there. But we are not there yet.
IP08	C5.1: Market Situation	But maybe part of it is influenced by the stage where we are. And then internationally, again, we have invested in funds, [VC Fund 1], [VC Fund 2], [VC Fund 3], [VC Fund 4], because we also have access to their deal flow. We are able to understand what's happening in a market that is not just our resources. And we're seeing best practice that people are doing. And once we have our foundation, we'll just incrementally build on top of our fund, they should pick up best practices of other firms.

IP08	C5.2.2: CRM	We're not using the technology right at this moment. But we are trying to get the companies at the pre-seed, trying to meet everyone, even though we just invest in Series A, and then what we are doing, I think well, is we are using affinity to place all these companies. And we have a lot of granularity in the information that we collect. And what we've been doing is just like a machine learning model that learns with more data. You put that in more iterations. That's what we do. We have regular weekly meetings about the pipeline. And then we are looking back to see what's working, what's not working. And so I think that for the sourcing, that's the main thing that we're doing.
IP08	C5.3: Overall Usage	We are not using and just to be super clear direct about it, we're not using tools at this moment, like other firms are using where they're tracking the LinkedIn jobs to see the startups that have LinkedIn jobs growing and web scraping that type of stuff. So we don't do that yet.
IP08	C6: Business Value	The tool will not be about efficiency. I think it will be about making sure we don't miss anything.
IP08	C7.4: Technology & Data	It's a good question, but I think it's like building an AI model where you need good data to train the model. We don't have data. And even if we consider ourselves not AI model, but ourselves as investors, it's a young team. There's not a lot of I've been an investor for three years, so even ourselves, we need to train ourselves as investors. I made 25 investment decisions.
IP08	C8: Future of VC Investments Data	The best founders are trained to contact the investors, talk to investors, build relationships with the investors, so the VC that are able to build really great brands. Of course they still need to do all of that. I think a majority of the founders that are building great companies, go to them and interact with them and they will interact with them at the right timing. I think those things are still going to continue to happen in the future.
IP08	C8.3: Commodity	Yeah, I think it's going to be more of a commodity. I hope it is something that everyone needs to have. But at this moment, I cannot see how that can be different for you to pick the unicorn, the deck cards of the next decade.

IP09 – 12.2021

IP09	C1.1: Challenges	So that's of course very basic, but quantity is not the problem, but quality is of course always the problem. And then as soon as you have something qualitative on the table, to get that through as quickly as possible, because it's already kind of right now in a landscape, where there's a lot of money on the market and there's not a lot of good deals as usual. So we're getting more good deals now as well. But it's always difficult for me personally to somehow get into the German deals with the London VCs, because of course in case of doubt, they prefer to work together with a [FUND NAME] that is local etc. That is, they are mostly companies that somehow have a focus in the UK. Whether that is somehow a company that has a founder in the UK or a founder that somehow has UK exposure. Or if it's more business model side that want to expand to London or in the UK.
IP09	C1.3: Criteria	That's a no-brainer, so it's called "team" and then of course it's kind of the market size. That's the competitive landscape. How do you differentiate yourself from the other players. So number one of course, team. And only then if the team is good, I do on number 2 and 3. In terms of "is this really a problem?" and "is this really the right solution for the problem?" But as I said, point 2 and 3 fly out a little bit like that because the team is really good. And then it's the hygiene factors along the lines of market size, competitive landscape, and the points that I just touched on.
IP09	C3: Requirements	So, very basic is simply data aggregation. Data aggregation and insights really, to have context. So data aggregation from various sources, from which insights are then generated that give context to a particular deal.
IP09	C3: Requirements	Track record outside of LinkedIn. Reputation. What are all the soft things. Everything outside of LinkedIn. So you don't get to experience that often, but you do get to talk to founders for a while and everything looks good, the company is cool. And you make reference calls and then that's an absolute armaggedon. But yes, because by the time I start with my reference calls, I've really put a lot of work into it. So maybe also a learning for me, but of course it's always difficult to ask for reference checks at different places at an early stage. So if you have a loose connection in that direction. That would be cool, of course, but to what extent is that realistic? But well, it is about requirements and not about realistic requirements. Som standardization, that too. Let's calculate the market sizes, using a mechanism given to me. And give me comparability across deals. I think that would be interesting again too. Standardize info from text, throw something out to me.

IP09	C5.2.2: CRM	Exactly for CRM, we use Affinity. That is also actually quite well maintained. So simply, according to stages. And then all communication is copied in there. The decks are then inserted and are then discussed in the deal flow meetings. But there we're actually relatively happy to say "Okay, that doesn't fit." (...) but also simply things, if then one, two concerns with team or also a concern with other topics, which also comes from a person in the round then, then such a deal also falls fast very, very fast out. And yes, it's often not just founder skills that matter, but also how well the founder fits in with us. And do we think this is a good human?
IP09	C5.3: Overall Usage	That's the only place where I could imagine now personally with [FUND NAME], so in the very, very early phase.
IP09	C5.3.2: Screening	Deck comes in and then just not an intern or an analyst that sits there looking at the inbox, but it runs through XY that then throws out at the end: "You look at it and that's definitely still worth any of your time."
IP09	C6: Business Value	Predictable training. For me. That you can give feedback. So I mean, if you have seen 50000 companies now, then you have a certain opinion, which is composed of factors that are completely invisible. And that can also kind of be in the day's opinion or according to the day's form. Exactly, standardization. On the one hand is predictive and on the other hand brings the standardization across deals, which just sometimes are not really right.
IP09	C7.2: Automated Denial	So the biggest fear is always that you overlook something. Which is then in 7 to 10 years the hot stuff, or in half a year the hot stuff. And if, in the end, something like that is thrown out by some algorithm, the feedback loops are too short for you to be able to say, "hh, we missed one. Okay, we have to say again very briefly that we have to pay attention to XY in the future." That's probably the only set back.
IP09	C8.8: Future requirements	But of course, now and that's of course the exciting question at the end of the day so what are the deals that fall out there? How far can you pre-qualify them. And how close do you end up getting to the Tinder swipe right or left logic? Because that's a bit, from my perspective, the dream, that you somehow look at the "Deal Tinder" every day and swipe right and left. And then ideally, that's then a bit of an investment strategy, that then in the background when you swipe right, ultimately the investment is already triggered.

IP10 – 12.2021

IP10	C2: Rationale & Trigger	Yeah, the reevaluation process is just a matter of that. We've done one investment using the tool. We used it for a year. We made one investment which already justifies having made the investment into the tool. So might as well just continue with it. As soon as you do one investment through a platform like that, it is already worth having the platform. At the time. We had twelve months to try it and we didn't make any investment, so we decided to stop it. But now that in fact, we have done an investment coming from there too. I think it could be worth using it again.
IP10	C2: Rationale & Trigger	We are investors at the pre-seed stages. So like, super early stage, and therefore they're like seeing talent that is about to start a company is actually a very good way of looking at it. I think it's actually a positive thing. And so I would put it that way. I think it's interesting. What I mean is that it's interesting to find talent that is about to start a new company, which is something that I think it's a bit bespoke to the way we look into companies.
IP10	C3: Requirement s	Now, what I'm going to say is very basic, but like a very strong and good user experience really makes a difference, because I remember this one of the pain points that you had to open it, filter it, and then the filters when the columns of the table are too wide that you need to scroll horizontally.
IP10	C3: Requirement s	But then I think the core feature for us is that there is way for us to classify what we would want, because a lot of these alerts are not material for the mandate that we have. So, for instance, if we invest into B2B SaaS, it would be helpful if you could find a way to filter for founders around that space, as opposed to filtering. It makes no point for me to be looking into founders that are doing like a B2C when I'm mostly investing B2B, and I think that customization towards an investor preference is something that I'd say is important.
IP10	C4.4: Talent Data	What I mean is that it's interesting to find talent that is about to start a new company, which is something that I think it's a bit bespoke to the way we look into companies.

IP10	C5.2.1: Third-Party Tools	We looked into tools like Specter.AI you might have heard of them when it comes to scraping talent on the web. We used them for about a year.
IP10	C5.2.1: Third-Party Tools	But then turns out that after having made that investment, after having made that decision, we actually ended up investing in one company and would have hoped to invest in another one that was flagged by that tool. So we actually realized that maybe it could have been a good investment to stay active on the platform as a fund.
IP10	C5.2.1: Third-Party Tools	I think it's supporting us in sourcing mostly. Then I don't think we would be in a position to apply AI. We looked into another company. We looked into a company called Koble, [https://www.koble.ai/]. Which is trying to help funds on the diligence elements, like filtering up the universe of companies we see, so that instead of the associate or the analyst to do the job of making that initial filter, you would be like getting a piece of AI that would map out the revenue figures of the market size, or the market space and would filter immediately. We were buyers of the first company.
IP10	C5.2: Systems	So it was not plugged to the CRM system. We had to manually upload it. The way it works is that every week they had a data dump of all the most exciting people that were starting a new company with the different links to their LinkedIn and Twitter profiles. And what we would do is that each week one of us within the investment team would go through the platform, just checking for those founders, checking if anyone within that founding group would be of interest for us to keep an eye on.
IP10	C5.2.2: CRM	I put them in within my CRM. I've got, like, a separate section that is, like, "not ready" where basically, I've got all the ventures there that I know I passed, but I'm keen to stay in touch.
IP10	C6: Business Value	The most rival for us to use it is for us not to miss, as in from a scraping standpoint, is for us not to miss an opportunity.
IP10	C7.6: Organization & Culture	I think it takes just a moment to adapt. I wouldn't say that it is a big issue to adapt to it. There is a bit of work to be done there, and everyone gets used to the tool. The tool when we use it was very early stage, so the product was not there yet, which is a bit painful to use.
IP10	C7.6: Organization & Culture	This is too strong of a statement, but our view is that making an investment in this at the moment might not be the right timing to market, because when you think, for instance, like scraping. Yes, I think it's fine, but the amount of money that we'll spend in scraping that information is probably worth just subscribing to Specter. And then if we decide, okay, then not Specter, to do something more like Koble.AI that is like helping you on the filtering process. And I'm sorry that I don't know many other players in that space. I think on the filtering elements of it.
IP10	C8: Future of VC Investments Data	It will be more around automating the investment process. It'll be a bit more around, like getting each team member up to speed on a certain investment opportunity. Almost like I have an investment opportunity. And I'd like the partner to slowly, slowly getting acquainted about the opportunity. Almost like flashcards or like a Q&A about company like, here is what this company is about. Like, kind of almost promoting because a lot of the time, what's happening is that there's a big information gap between us, the investment lead, and the rest of the team. Or between the founder and the investment lead and investors. And I think there's a bit of an opportunity to automate parts of this process so that we can dedicate more time to real discussions, as opposed to just descriptive type of work. Like describing what the company does, what the company does not do.