International Review of Financial Analysis 109 (2026) 104748

Contents lists available at ScienceDirect

International Review of Financial Analysis

journal homepage: www.elsevier.com/locate/irfa

Generative AI-powered venture screening: Can large language models help venture capitalists?☆

Silvio Vismara[a,b,*], Gresa Latifi[c], Leonard Meinzinger[c], Alexander Pass[d]

[a] University of Bergamo, Department of Management, Via dei Caniana 2, 24127 Bergamo, Italy
[b] IMT School for Advanced Studies Lucca, Italy
[c] TUM School of Management, Technical University of Munich, Arcisstraße 21, 80333 Munich, Germany
[d] Freigeist Capital, Joseph-Schumpeter-Allee 25, 5327 Bonn, Germany

ARTICLE INFO

Keywords:
Venture capital
VCs
Large language models
LLMs
Generative artificial intelligence
GenAI
Investments

ABSTRACT

This paper examines the potential of Large Language Models (LLMs) to enhance the selection process of venture capitalists (VCs). We employ an LLM agent on a VC database comprising 61,814 early-stage ventures to evaluate the efficiency and categorization quality of the venture screening process. Our findings indicate that LLM agents outperform humans, operating 537 times faster than a human VC analyst without sacrificing categorization quality. Specifically, the LLM agent performs comparably to a human analyst in forming distinct clusters (as indicated by Silhouette scores of 0.35 and 0.37, respectively) while exceeding human performance in cluster separation and compactness, as evidenced by a 70 % increase in the Calinski-Harabasz Index. These results suggest a transformative shift in VC practices, highlighting the potential of LLMs as tools for structuring and organizing deal flow.

1. Introduction

Venture capital (VC) firms play a pivotal role in fostering innovation by providing equity financing to promising early-stage ventures. A substantial body of literature identifies VC as one of the most suitable financing mechanisms for entrepreneurial firms in their formative stages (e.g., Gompers & Lerner, 2004). This effectiveness is typically attributed to both selection (picking winners) and treatment (building winners) effects, with selection often regarded as the most critical and, simultaneously, the most challenging task (e.g., Sørensen, 2007). The magnitude of this challenge is illustrated by the fact that more than 20,000 firms received VC funding in the United States in 2023 (U.S. Bureau of Labor Statistics, 2023). Given that VCs usually finance less than 1 % of applicants, this implies that investment decisions are drawn from a pool exceeding 2 million potential ventures. The difficulty of identifying the most promising firms is thus amplified not only by the sheer volume of opportunities but also by the limited availability and reliability of early-stage data. To address these challenges, VCs are increasingly adopting a variety of "quantitative sourcing" tools to support venture screening (Gompers et al., 2020, p. 15).

☆ Silvio Vismara received financial support from the European Union–NextGenerationEU, in the framework of the GRINS—Growing Resilient, INclusive and Sustainable project (GRINS PE00000018–CUP F83C22001720001). Silvio Vismara also received financial support from the European Union–Next- GenerationEU, under the National Recovery and Resilience Plan (NRRP), Mission 4, Component 2, Investment 1.1, funding call PRIN 2022 D.D. 104 published on 2.2.2022 by the Italian Ministry of University and Research PRIN programme "Digitization and the inclusiveness of entrepreneurial finance"—PRIN 2022Y3AWN5_02, financed by the European Union–Next- GenerationEU, M4C2 1.1 CUP: F53D23003190006.

* Corresponding author at: University of Bergamo, Department of Management, Via dei Caniana 2, 24127 Bergamo, Italy.
E-mail addresses: silvio.vismara@unibg.it, silvio.vismara@imtlucca.it (S. Vismara), gresa.latifi@tum.de (G. Latifi), le.meinzinger@tum.de (L. Meinzinger), alexander@freigeist.com (A. Pass).

https://doi.org/10.1016/j.irfa.2025.104748
Received 27 June 2025; Received in revised form 3 October 2025; Accepted 10 November 2025
Available online 12 November 2025

In this context, the advent of artificial intelligence (AI) is poised to emerge "not just as a tool, but as a transformative force, redefining the parameters of investment strategies and operational efficiency."[1] Given these developments' substantial financial and systemic impact, it is crucial to comprehend AI's potential to enhance the VC screening process, i.e., can AI help VCs?

This question is not trivial. While the impact of AI—and, more broadly, digitalization—on transactional efficiency is significant in reducing transaction time and costs (Momtaz, 2024), the role of AI—particularly Generative AI (GenAI)— in improving search efficiency (the process of finding a suitable transaction counterparty) remains uncertain. Search friction and market granularity require that agents (VCs) thoroughly screen deep markets to avoid resource misallocations. Anecdotal evidence suggests that VC fund managers often rely confidently on their experience, with some (9 %) even reporting that data play only a small role in their investment decisions, instead depending on their intuition and "gut feelings" (Pitchbook, 2018).[2] Thus, the nature of entrepreneurial finance poses peculiar challenges to adopting AI compared with other finance markets. First, gathering substantial *historical* data is challenging and costly for early-stage ventures. Second, entrepreneurial finance requires selecting high-potential ventures *before* others, yet learning from past decisions tends to bias VCs toward ventures with previously successful traits. Third, successful exits from a few ventures typically generate returns for the entire VC fund, but AI has limited effectiveness in forecasting *infrequent* events (De Prado, 2018), reducing the likelihood that VCs can identify early-stage ventures poised for exceptional success.

Despite these challenges, integrating GenAI, defined as "computational techniques that are capable of generating seemingly new, meaningful content such as text, images or audio from training data" (Feuerriegel et al., 2024, p. 111) presents significant potential to enhance VC decision-making processes. First, AI excels at processing unstructured textual data—a crucial capability, as most venture-related information typically comes from company presentations, websites, and similar sources that generally are accessible in text form (or can be made accessible, e.g., through transcripts). Given access to this material, large language models (LLMs)—the most widely used form of GenAI -go beyond traditional AI methods by not only analyzing and structuring vast amounts of unstructured data, but also substantially supporting early-stage decision-making. This capability, which traditionally would require considerable human resources and time, enables LLMs to process, synthesize, and extract relevant information from diverse sources efficiently, rendering them valuable tools for initial venture screening. However, while LLMs have proven effective in processing and interpreting financial texts (e.g., Martin & Nagel, 2022), their application in early-stage venture identification, though promising, remains a largely unexamined area. Although AI-driven venture selection has gained some attention in practitioner discussions (e.g., IngestAI., 2023; MLQ, 2023), these narratives primarily highlight broad applications rather than providing systematic, empirical validation. Moreover, proprietary GenAI-based approaches, such as Moonfire's use of LLMs, limit accessibility and make it difficult for others to learn from or replicate their approach. This study aims to bridge that gap by offering a structured evaluation of GenAI, specifically LLMs, within a real-world venture capital context, providing new empirical insights into their practical application.

This paper investigates the transformative role of large language models (LLMs) in the venture selection process of venture capital (VC) firms. The VC investment process unfolds as a multistage sequence that begins with the screening of early-stage ventures for deal origination (Sahlman, 1990). Our analysis focuses on whether, and in what ways, LLMs can support VC fund managers in streamlining the initial stages of managing and analyzing large deal flows. This stage is particularly critical, as the majority of ventures are eliminated during preliminary screening, yet it has received considerably less scholarly attention than subsequent stages such as due diligence or valuation. Adhering to the "quantitative sourcing" approach defined by Gompers et al. (2020), we conceptualize venture selection as a topic-driven process that identifies investment targets by mapping future developments of an industry niche or by analyzing the interplay between thematic and quantitative factors that influence ventures' success. This search methodology aligns with industry practices (e.g., Fried & Hisrich, 1994) and is consistent with the use of "investment memoranda" described by Kaplan and Strömberg (2001).

We compare the efficiency of agent-based LLM screening—measured in terms of speed and cost—as well as its categorization quality with that of traditional human-based analysis, assessing their respective implications for VCs' early-stage screening decisions. In examining speed, we assess whether the agent effectively can address a fundamental challenge in VC—the overwhelming volume of newly established ventures—which complicates optimal decision-making for human investment managers. Evaluating categorization quality poses a greater challenge. We do not measure venture success due to endogeneity, as receiving VC investments, everything else being equal, would increase the probability of success (Hellmann & Puri, 2000; Kaplan & Strömberg, 2001). Instead, we assess categorization quality using well-established clustering metrics to evaluate the quality of categorization. First, we measure the Silhouette score, which evaluates cluster cohesion (venture similarity within the same investment category) and cluster separation (ventures' distinctiveness across different categories). A high Silhouette score indicates that the LLM agent successfully groups ventures with similar attributes while maintaining clear boundaries between diverse categories. Second, we measure the Calinski-Harabasz Index, which assesses ventures' global variance within clusters by calculating the ratio of between-cluster variance to within-cluster variance. A higher score reflects better-defined and well-separated clusters, indicating categorization robustness.

Empirically, we base our analysis on a real-world application of LLMs to screening investment opportunities, conducted by Freigeist Capital, a German early-stage VC firm that focuses on high-tech and deep-tech investments. Our findings, based on their private dataset of 61,814 early-stage ventures, indicate that LLM agents significantly outperform human VC analysts in venture selection. LLM agents perform nearly as well as human analysts in creating distinct clusters, with Silhouette scores of 0.35 and 0.37, respectively. However, they surpass human performance in terms of cluster separation and compactness, demonstrated by a 70 % increase in the Calinski-Harabasz Index. Furthermore, LLM agents identify suitable ventures 537 times faster than human VC analysts. This estimate assumes a conservative two-hour review period per early-stage venture by a VC analyst. Notably, a survey by Gompers

[1] This quotation is from the 2024 Forbes article "Venture Capital's New Era: AI's Journey From Enhancing Operational Efficiency To Alpha Generation": https://www.forbes.com/sites/josipamajic/2024/01/16/venture-capitals-new-era-ais-journey-from-enhancing-operational-efficiency-to-alpha-generation/. Expectations of disruptive effects from artificial intelligence in finance, including entrepreneurial finance and VC investments, increasingly are being debated in news media and among practitioners. For example, datadrivenvc.io provides insights into VCs' use of AI and interviews with VCs. According to a recent Pitchbook survey, 22 % of VCs use machine learning technology to make investment decisions (Pitchbook, 2018). Gartner predicts that AI will be involved in 75 % of VC investment decisions by 2025: https://www.gartner.com/en/newsroom/press-releases/2021-03-10-gartner-says-tech-investors-will-prioritize-data-science-and-artificial-intelligence-above-gut-feel-for-investment-decisions-by-20250

[2] VC fund managers often advertise their past entrepreneurial successes through their profile pages on LinkedIn, AngelList, and CrunchBase (Venugopal & Yerramilli, 2022). To exemplify, we include a quotation from a *Forbes* article, "Get experience under your belt as an investor; it'll help you avoid common mistakes" (https://www.forbes.com/sites/mirunagirtu/2022/10/04/from-founder-to-funder-insights-from-angel-investors-with-an-entrepreneurial-background/).

et al. (2020) found that VCs spend an average of 118 h on screening and due diligence per venture.

Our findings contribute to the literature in at least two significant ways. First, we build on the rapidly growing body of research on the application of AI in finance (Ardekani et al., 2024; Aziz et al., 2022; Beckmann & Hark, 2024; Dowling & Lucey, 2023; Goodell et al., 2023; Ko & Lee, 2024; Pelster & Val, 2024), including private equity (Braun et al., 2023). While prior research has examined the role of generative AI in financial modeling, decision support, and advisory applications, our study advances this literature by offering a structured evaluation of LLMs in venture capital screening—an area where empirical evidence remains scarce. By comparing the LLM-based venture selection process with traditional manual screening through a series of venture-clustering experiments, we demonstrate that GenAI offers significant efficiency gains, particularly in processing unstructured investment data and supporting early-stage decision-making. This supports the broader thesis that AI-driven solutions can facilitate more scalable and sustainable business practices in entrepreneurial finance, in which the ability to assess potential investments quickly and affordably can impact overall success and operational efficiency (Zhang et al., 2024). This finding also builds on prior studies that have focused on digital transformation in finance (e.g., Bellardini et al., 2022; Collevecchio et al., 2024; Siddik et al., 2025), highlighting how specific GenAI technologies can alleviate operational bottlenecks in entrepreneurial financing. Potentially, our results suggest that LLMs may help democratize entrepreneurial finance (e.g., Cumming et al., 2021) by lowering screening costs, thereby reducing the barrier to entry into private equity for small investors.

Second, we contribute to the ongoing debate challenging existing perceptions about automated systems' limits in creative and analytical roles (Boussioux et al., 2023; Feuerriegel et al., 2024). Our documentation that LLM agents can match or even exceed human performance in selecting early-stage ventures bolsters the argument for GenAI's successful integration in decision-making processes within sectors that rely heavily on qualitative assessments and complex data analysis, as pointed out by management and entrepreneurship studies (Antretter et al., 2020; Cappa et al., 2021; Matthews et al., 2025; Obschonka et al., 2025; Obschonka & Fisch, 2022; Shepherd & Majchrzak, 2022). Unlike human analysts, who may be overwhelmed by large volumes of financial information (Hirshleifer et al., 2009), LLMs can distill relevant information more efficiently. These findings broaden the applicability of AI in the VC context, particularly in the initial stages of managing and analyzing large deal flows. Beyond this specific application, the study also contributes methodologically: its design is readily reproducible in other contexts, for instance by leveraging venture data collected by different VC firms. This enhances the external validity of our approach and underscores its potential to generalize across diverse investment settings.

The remainder of the paper is structured as follows. Section 2 connects our study to the literature on AI-assisted human judgment. Section 3 introduces LLM agents. Section 4 outlines our methodological approach, and Section 5 presents the results. Section 6 discusses the ecological validity of our study, while Section 7 highlights the practical implications. Section 8 concludes.

2. AI-assisted human judgment in finance and entrepreneurship

For most of history, judgment across domains was exclusively human-centered. From everyday business decisions to choices about retirement investments, individuals primarily relied on their own knowledge, seeking expert advice only when challenges exceeded their expertise (Klein, 1998). In finance, in particular, the dominant view held that sound evaluation required both extensive information and skilled interpretation (Slovic, 1972). With advances in information technology, however, the concern shifted from information scarcity to the ways in which abundant data could be meaningfully used. Decision-making is not purely statistical or logical; it is also shaped by heuristics. While heuristics enable faster judgments by drawing on familiar mental prototypes (Kahneman & Frederick, 2002), they also introduce systematic biases and errors.

The integration of AI into domains characterized by high task complexity represents a major shift in this landscape. As a decision-support tool, AI is increasingly prominent in entrepreneurship and finance—fields long dominated by human expertise and judgment. For instance, in the entrepreneurial context, AI has been shown to enhance creative problem-solving. In a large-scale crowdsourcing challenge on the circular economy, business idea generation achieved higher levels of strategic viability, environmental and financial value, and overall quality when humans and AI collaborated, underscoring the transformative potential of generative AI (Boussioux et al., 2023).

Recent research further examines AI's role not merely as a signal enhancer but also as an independent evaluator of venture quality. Csaszar et al. (2024) compare LLM-based assessments with those of experienced venture capitalists and angel investors in a startup competition at a leading business school, reporting a strong alignment between AI and investor judgments (correlation = 0.52). Similarly, Blohm et al. (2022) demonstrate that AI can replicate many aspects of expert evaluation, providing a scalable mechanism for early-stage screening. However, they also find that highly experienced business angels retain an advantage over both AI and less experienced investors.

Taken together, these studies highlight an ongoing reconfiguration of judgment in finance and entrepreneurship. AI is no longer limited to augmenting human decision-making but is increasingly assuming roles once reserved for experts. From helping entrepreneurs refine pitches to matching investor evaluations in structured contests (Csaszar et al., 2024) and even outperforming average business angels in returns (Blohm et al., 2022), AI demonstrates the potential to democratize access and scale evaluative expertise. Nonetheless, the persistent advantage of seasoned experts indicates that judgment in highly uncertain and ambiguous contexts may remain resistant to full automation. The challenge ahead lies in designing hybrid systems that combine the scalability and consistency of algorithms with the contextual sensitivity and intuition of experienced decision-makers—ultimately reshaping the division of labor between humans and intelligent systems.

3. The driving force behind GenAI: LLM agents

LLMs excel at translating unstructured data into structured formats. Given access to unstructured information, LLMs can process and organize the information quickly and, thus, substantially support the early decision-making process. However, understanding the limitations and challenges associated with LLMs in the VC context requires recognizing the complexity and diversity of data sources for early venture screening. These data sources consist of web content, PDF documents, databases, Excel files, and presentations, which must be queried in different orders and with various conditions and restrictions. Trying to achieve this with LLMs alone, even when paired with traditional data engineering techniques, is hardly feasible.

To address LLMs' shortcomings, we propose the use of LLM agents. LLMs' ability to formulate actionable plans was first posited by Huang et al. (2022), who prompted an LLM to develop a plan to interact with a simple and limited environment. In their original experiment, the authors tasked the LLM *GPT-3* with devising step-by-step plans to complete day-to-day challenges verbalized in natural language. The researchers formalized these challenges and tasked the LLM with devising a sequence of steps comparable to "walk to the fridge," "open fridge," "take out milk," and "close fridge," then creating a prompt, such as "make breakfast." In this way, they moved toward the concept of LLM agents by examining how LLMs can generate actionable outputs.

As for the structuring process of actionable steps, the language model possesses an awareness of the third component of the agent, which encompasses accessible tools. These tools serve as the agent's core functionality, facilitating the language model's interaction with its environment. For example, in the context of the "making breakfast" task, these tools might encompass an application programming interface (API) linked to a cookbook, thereby furnishing the agent with reliable and current information on recipes that surpass the language model's internal knowledge (Petroni et al., 2019). By comprehending these tools' essential functions, the agent can make informed decisions regarding which tasks to tackle solely with a language model and those that require a traditional tool. This comprehension also includes an acknowledgment of its limitations, usually encoded into the central prompt executing the agent. Consequently, tasks can be addressed through a finite loop involving tool selection, tool utilization, and the LLM's interpretation of tool-generated outputs. Once the user's originally presented task has been resolved to the agent's satisfaction, the outcome is delivered (Yao et al., 2022a).[3]

It is also important to acknowledge that hallucinations remain a general risk of generative AI (GenAI) systems, which frequently produce outputs that appear plausible but are in fact inaccurate. When such content is uncritically adopted by humans, it gives rise to what Hannigan et al. (2024) term "botshit"—LLM-generated hallucinatory content that is incorporated into organizational tasks without verification. Hannigan et al. (2024) further distinguish between intrinsic botshit, which occurs when outputs contradict established data, and extrinsic botshit, which arises when outputs speculate beyond available evidence. Both forms are highly relevant in the context of venture screening, where unreliable classifications may distort evaluations or misdirect analyst attention. However, the risk of hallucination in our setting is limited. The data employed in our analysis is not generated by the LLM; rather, the LLM is tasked with translating and structuring existing venture data. Consequently, the model does not have the capacity to invent non-existent companies or attributes.

4. Methods

4.1. LLM agents on VC screening

The screening process of VCs can be facilitated through the deployment of LLM agents. We framed LLMs' role in the VC investment process as a *hypothesis-based* search. Gompers et al. (2020) first described the general idea of utilizing ventures' specific characteristics to realize elevated returns with the help of quantitative approaches, aligning with Kaplan and Strömberg's (2001, p. 5) 'investment memoranda.' This approach mirrors the broader role of certification and transparency mechanisms in financial markets, where transparent bookbuilding has been shown to reduce information asymmetries and enhance selection efficiency in IPOs (Khurshed et al., 2014). In particular, we conceptualized venture selection as a topic-driven process that aligns with the 'quantitative sourcing' methodology that Gompers et al. (2020) delineated. This approach involves identifying investment targets by assessing the future trajectories of specific industry niches through market mapping, that is, by examining the interplay between thematic and quantitative factors that predict startup success. This approach aligns with standard industry practice, as described, for example, by Fried and Hisrich (1994) where the goal is to screen ventures, filter out those with strong potential, conduct due diligence and eventually make investment decisions.

An example is Sequoia Capital's investment, alongside Freigeist, in the Munich-based robotics startup RobCo. Unlike conventional sourcing practices, where deals often originate from referrals or conferences (Teten & Farmer, 2010), Sequoia began by developing a focused perspective on the industrial production sector. The firm reasoned that modular robotics would enable scalable adoption across small- and medium-sized enterprises, and only after defining and validating this view did they search for ventures aligned with it, ultimately investing in RobCo with Freigeist. Other VC firms also adopt comparably structured approaches. For instance, Moonfire, mentioned earlier, uses explicit investment theses in its routines. In integrating LLMs into early-stage screening, the firm describes how it first sets out thematic directions regarding where an industry has been, where it stands, and where it is headed. It then drills down into verticals and tools, asking questions such as "What are the attributes of successful tools?" and "What future models may emerge?", before drafting concise descriptions of promising companies in the space (Moonfire, 2023). Both cases illustrate how structured, forward-looking reasoning guides venture selection. In our framework, these reasoned projections about how a niche, technology, or business model may evolve into an investable opportunity represent fixed inputs defined by VC teams.

Coherently, our approach begins by formulating and considering hypotheses related to a particular market segment. The search process often involves analyzing macro- and microeconomic trends, identifying promising first movers, and monitoring scientific advances as signals of future technological breakthroughs. In the VC literature, this practice is closely related to market mapping (Teten & Farmer, 2010). Such search activities are highly automatable with LLM agents. After receiving initial input, the agent refines its understanding using internal knowledge and external tools (Petroni et al., 2019). It then performs an embedding-based search to generate a list of suitable ventures. With access to updated venture databases, the agent can cover a far broader opportunity set than a human analyst.

We developed and implemented an intelligent agent system based on the reasoning and acting (ReAct) framework introduced by Huang et al. (2022) and Yao et al. (2022a). Our approach combines advanced natural language processing with traditional financial analytical tools to enhance VC investment decisions. This methodological framework implements a sequential process for investment analysis. The system initiates with user-defined investment criteria that are then processed through an LLM agent. We provide a formal description of our model in Appendix 1.

We constructed the LLM agent using the LangChain Python library. To this extent, we developed an SQL Topic Tool capable of translating natural language requests into SQL queries for the database. To achieve this, it employs a combination of embedded searches (Levy & Goldberg, 2014) and traditional SQL querying, enabling targeted navigation of the language model within the venture database. The agent's operation follows a structured five-step process. First, it processes initial user inputs and then defines investment parameters. Second, it evaluates and decomposes these objectives into analyzable components. Third, it defines specific actionable steps based on available analytical tools. Fourth, it determines optimal task allocation between LLM processing and specialized tools. Finally, it executes the defined steps and generates investment recommendations.

A key innovation in our approach is the comprehensive integration of multiple data sources. Our implementation utilizes a zero-shot-react-description LLM agent architecture, which enables the system to process complex investment criteria without requiring extensive pre-training. The agent can analyze both microeconomic indicators at the firm level and macroeconomic trends affecting market conditions. Specifically, the agent employs three categories of analytical tools. First, the system connects to established VC databases, including Dealroom, Crunchbase, and Pitchbook. Second, it incorporates scientific and technological tracking through the ArXiv and PubMed databases, enabling early identification of potential breakthrough technologies that

[3] To illustrate this with the case of the breakfast example, this outcome might comprise a series of actionable steps for preparing a French toast breakfast under the direction of a recipe retrieved from the corresponding cookbook API. In a more advanced scenario, in which the toolkit incorporates access to robotic capabilities, enabling the agent to interact with physical space and a real kitchen, the outcome could extend to a fully prepared breakfast meal. This vision's viability, which initially may seem ambitious, is substantiated by Google DeepMind's research on generalized agents, as Reed et al. (2022) demonstrated.

may influence investment opportunities. Third, it includes Wikipedia API integration and web search functionality. For company registry access, we implemented REST API calls with rate limiting and automatic retries to ensure stable data collection. The Dealroom, Crunchbase, and Pitchbook integrations utilize OAuth 2.0 authentication protocols, with data retrieval orchestrated through asynchronous calls to minimize latency. These tools enable the system to maintain current market awareness and technological understanding.

4.2. Sample

We based our analysis on a real-world application of LLMs to screen out investment opportunities, conducted by Freigeist Capital (formerly known as e42 Ventures), a German early-stage VC firm focused on investing in high tech and deep tech, targeting sectors such as AI, robotics, biotech, and quantum computing. Freigeist's typical investment range is €1 M—to €3 M, supporting ventures primarily in Germany to scale from seed stages through Series A and beyond. Notable companies in its portfolio include MyTaxi (acquired by Daimler) and Kraftblock (sustainable energy storage systems). Freigeist Capital has distinguished itself as an influential player within Europe's VC ecosystem.

Our dataset comprises observations from the Freigeist Capital's deal flow records from 2018 to 2023, collected as part of the firm's initiative to integrate data-driven methodologies into investment decision-making. It includes all 61,814 ventures evaluated by Freigeist during this period, not only those ultimately funded. This aligns with recent research underscoring the importance of analyzing the entire set of sourced ventures when studying VC decision-making (Jang & Kaplan, 2025). The dataset contains both structured covariates—such as company and founder attributes, geographic location, and funding stage—and unstructured textual descriptions reflecting each venture's thematic and technological orientation. It also records secondary variables that enable more advanced screening, including funding amounts (equity or debt) and employment size. These features allow both human analysts and LLM agents to move beyond simple thematic clustering and incorporate richer, multidimensional decision criteria.

The scale of the dataset supports a highly granular analysis, particularly important given the unsupervised methods applied in our study. The data reflect Freigeist's geographic focus on Germany and the broader European Union, which is relevant for three main reasons. First, Germany constitutes one of the largest and most dynamic startup ecosystems in Europe, ranking alongside France as the EU leader with USD 4.3 billion in VC investment (Dealroom, 2024; Pitchbook, 2024). Second, Germany's entrepreneurial environment is unusually data-rich, benefiting from comprehensive public registries (e.g., Handelsregister and the Federal Gazette), statistical business registers (Destatis), and curated startup directories (e.g., Startup Detector). Such heterogeneous datasets are critical for LLM-based venture screening, as the type and structure of data strongly influence clustering and categorization outcomes. Third, the dataset reflects Freigeist's strategic orientation: with the investment team based in Germany, ventures are concentrated in geographies that facilitate effective monitoring and advising. This local orientation reflects the well-documented "home bias" in VC investing, which reduces portfolio-monitoring costs and enhances post-investment support (Bernstein et al., 2016; Chan et al., 2005; Lyonnet & Stern, 2024). While the EU focus limits the generalizability of our findings, it aligns with ex ante selection practices and ex post monitoring needs in the VC industry. Table 1 reports the country-level distribution of ventures in the dataset.

4.3. Examples

In this section, we explain our approach via examples. The primary variable collected for the analysis was *description,* i.e., an automatically generated description of the venture, its market, technology, and maturity. The text was formulated based on the venture's website. Its content was scraped first and then summarized by OpenAI's *GPT-3.5-turbo*. The model was chosen because of its superior inference speed and low cost per token. To derive an effective summary, a specific prompt (Liu et al., 2023) was developed to calibrate the LLM toward capturing key information for a quick understanding of the venture through a VC lens. The resulting summary is a text of about five to seven sentences formulated in natural language.

Table 1
Geographic distribution of the sample.

Country	No.	%
Germany	19,488	31.53
Great Britain	6500	10.52
France	4771	7.72
Switzerland	4153	6.72
United States	4072	6.59
Netherlands	3759	6.08
Sweden	2115	3.42
Estonia	1948	3.15
Denmark	1561	2.53
Finland	1382	2.24
Belgium	1163	1.88
Austria	1151	1.86
Poland	1041	1.68
Italy	1020	1.65
Other Countries (<1,000)	7690	12.44
Total	61,814	%100

As is typically done in natural language processing (Levy & Goldberg, 2014), the *embedding* variable was designed to create a vector representation of the *description* variable, in which a 1536-dimension numeric vector equivalently represents the description. The selection of 1536 dimensional vectors for our embedding framework followed established conventions in natural language processing, particularly in applications utilizing GPT architecture (Bai et al., 2024). This dimensionality represents an encoding process, rather than a reduction, in which the LLM transforms raw text into high-dimensional vectors that capture semantic relationships and contextual information. Our methodology employed a two-stage approach to dimensionality: First, the text-to-vector transformation creates 1536 dimensional embeddings that preserve comprehensive linguistic features, followed by a principal component analysis (PCA) that reduces these vectors to their two most significant components for visualization purposes. While this reduction necessarily involves some information loss, our empirical results demonstrate that the reduced representations maintain sufficient discriminatory power to differentiate between technology and non-technology ventures effectively. This retained separability post-PCA provides robust evidence of the effectiveness of our initial 1536-dimension vector representations. The methodology's success in maintaining interpretable distinctions even after significant dimensionality reduction underscores our embedding approach's robustness and its applicability in financial analysis contexts, particularly in VC investment screening.

While conducting venture selection and prioritization of a venture based on such a short thematic overview of a venture may seem counterintuitive, VCs similarly often conduct fast initial assessments. In this regard, Hall and Hofer (1993) described how many early rejections of funding for ventures are finalized within less than two minutes by an experienced human investment manager. Thus, initial decisions on whether a venture is worthy of extensive analysis have also been conducted with minimal information at hand in the case of traditional investment approaches. This is solidified further when considering the illustrations in Fig. 1, depicting a confrontation between the first two principal components of a matrix of venture embeddings. The companies in this plot were selected manually to fit a profile of either deep-tech ventures or traditional small and medium-sized enterprises (SMEs). The short LLM-based *embedding* variable, even when substantially reduced in dimensionality through PCA, holds enough information to

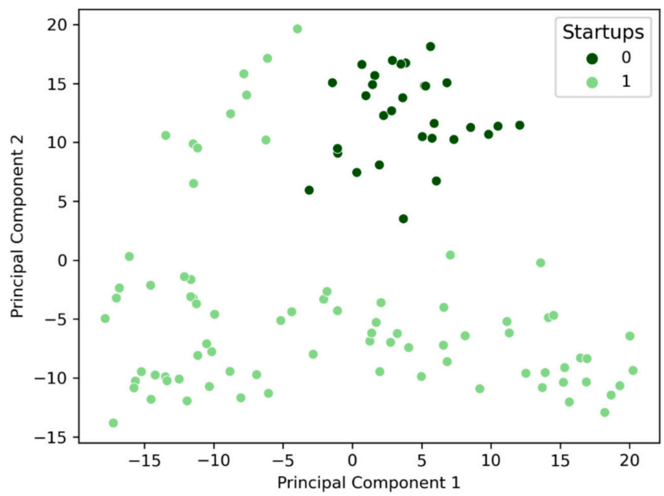

Fig. 1. Scatterplot of first two principal components of the embedding of LLM-based company descriptions; differentiated between Deeptech (1) vs. traditional (0) ventures.

separate SMEs from deep-tech ventures linearly. This hints at a solid internal validity of the analysis conducted, as the evidence suggests the existence of sufficient information within the *description* for the segmentation of ventures. Moreover, the random selection of ventures from a larger existing database provided by Freigeist Capital induces confidence that selection bias will be limited.

Finally, Appendix 2 provides some examples of hypothesis searches conducted using the LLM agent. The varying queries indicate a wide range of possible formulations of hypotheses that the agent can execute. For example, a VC fund might observe the potential of ventures that have bootstrapped for the initial years since founding the company and demonstrate evidence of some success by hiring a few employees. This process may hint at already-proven product market fit and suggest some experience in managing profitable growth. To this extent, the second query searches for companies with these characteristics, and the LLM agent produces suitable ventures from the database within a few seconds. However, the final query follows a similar approach, as observed by Sequoia Capital (see Footnote 4), and searches for companies in a specific niche. Similarly, the agent produces suitable companies that otherwise would have required intensive research by a human VC analyst.

4.4. *Agent evaluation metrics*

We compared agent-based LLM screening's efficiency—measured in terms of speed and cost—and categorization quality with traditional human-based analysis, assessing their impact on VC decision-making. As for speed, we examined whether the agent could address a fundamental VC challenge effectively—the overwhelming volume of newly established ventures—which makes it difficult for human investment managers to reach optimal decisions. Simple metrics, such as arithmetic mean, accompanied by an analysis of distribution, provide sufficient insight into these performance indicators' characteristics. Focusing on screening quality presents a significantly greater challenge. We deliberately avoided using venture success as a performance metric, not only because our approach originates from a real-world application—in which we cannot determine *ex ante* the success of ventures under consideration—but more importantly to mitigate potential endogeneity bias. Receiving VC investment significantly increases a venture's likelihood of success (Hellmann & Puri, 2000). Yet, comparatively little is known about the early stages of managing and analyzing large deal flows, despite these initial steps being highly consequential for the overall screening process.

Instead, we focus on categorization quality through established clustering metrics that assess venture categorization quality. Specifically, we employed the Silhouette score to measure both cluster cohesion (i.e., how similar ventures within the same investment category are) and separation (i.e., how distinct ventures are from other investment categories). This metric provides insights into whether the LLM agent consistently identifies ventures with similar characteristics and distinguishes between different venture types. Furthermore, we utilized the Calinski-Harabasz Index to evaluate global variance in ventures within clusters, which helps assess overall categorization quality by measuring the ratio of between-cluster variance to within-cluster variance. We provide details on these metrics in Appendix 3.

These clustering metrics are particularly suitable for evaluating LLM agent performance because they align with fundamental VC screening objectives: identifying ventures that share characteristics with successful portfolio companies (cohesion) while maintaining a diversified portfolio (separation). Furthermore, these metrics provide quantitative benchmarks that are independent of eventual venture outcomes, allowing for immediate assessment of screening quality without waiting for long-term performance data. The combination of these metrics enables us to confront the LLM agent with human analysts.

4.5. *Human analysts*

To provide a comparative baseline for the LLM agent's performance, we engaged four analysts from Freigeist Capital. This selection aimed to capture a spectrum of experience levels and domain expertise relevant to the venture screening task. The analyst cohort consisted of German nationals, with ages ranging from 29 to 52 years, reflecting a diverse age distribution within the professional setting. We categorized the analysts into two groups: early-career analysts and experienced investment professionals. This categorization allowed us to assess the impact of professional experience on the clustering task. The early-career analysts group comprises two individuals with distinct academic backgrounds: one participant possesses a technical background with 16 months of professional experience or equivalent training, while the second has a business-oriented educational foundation with 14 months of industry experience or formal training. The experienced investment professionals group consists of two senior members: an investment manager with technical expertise and four years of professional experience in the field, and a partner with academic credentials in business and economics complemented by eight years of investment management experience.

These analysts collectively possess expertise across key technologies such as deeptech and AI-driven innovation. Their diverse educational backgrounds and professional trajectories equip them with the capacity to identify and group similar investment opportunities while maintaining meaningful distinctions between different venture categories. All analysts were presented with the same information and were tasked with clustering ventures based on their perceived similarities and differences. While allowing for individual analytical approaches, we provided general guidelines regarding the key factors to consider when clustering ventures, such as industry, technology, and business model. The resulting cluster assignments from each analyst were recorded for subsequent comparative analysis with the LLM agent's output. The results of the inter-rater reliability analysis revealed a high level of agreement within the two categories, early-career analysts and experienced investment professionals.

5. Results

We assess the LLM agent's performance in terms of speed, cost, and categorization quality and, thus, compare with a baseline of random clustering and analytics conducted by a human VC investor. Hall and Hofer (1993) analyzed a human investment manager's speed when reviewing proposals from potential venture targets. They observed that the fastest time that firm rejections were being made was about two minutes; thus, an initial analysis for a given venture at hand can be conducted in a concise time frame. However, to fit ventures to a certain hypothesis, a VC analyst must conduct several such short analyses and rank them by relevance. Considering the literature and industry practices, a VC analyst might spend around two hours conducting this initial ranking analysis. This conservative estimate aligns with the due diligence process documented by Gompers et al. (2020), who report that the average deal requires approximately 83 days to close, with VCs spending about 118 h contacting around ten references. For early-stage ventures, where less extensive due diligence is performed, the average time commitment remains substantial, at roughly 81 h. While examining this, and also breaking down the decision-making process (see Elango et al., 1995), we find that apart from initial screening, which Hall and Hofer (1993) found would take 10 min, the preliminary analysis involving assessments of business model, market potential, team, and alignment with investment thesis (Hall & Hofer, 1993) would take roughly 90 min in our assessment. Then there would be 20 more minutes allocated to multiple quick assessments, ensuring that the most prominent opportunities are filtered for further due diligence, totaling roughly two hours (120 min). This is a rather sharp contrast to the time that the LLM agent needs to conduct the search corresponding to a proposed hypothesis.

Fig. 2 illustrates the distribution of the time required. The observed duration strongly depends on the quota of loops of thought and actions that the agent deems necessary to identify ventures suitable to a hypothesis. More specifically, it depends on the need to access additional knowledge through the APIs provided. The mean search time in 67 conducted executions was calculated to be 13.426 s. When referenced to the estimated time it takes a human VC analyst to complete the same task, this is an improvement in speed by a factor of over 500. Specifically, for a task that a human VC analyst would need two hours to complete (7200 s), an LLM agent would need only 13.4 s; therefore, an LLM agent would be 537 times faster than a human VC analyst. In terms of expenditures, a VC analyst in Germany would command an annual salary of approximately $73,000,[4] corresponding to an hourly compensation of approximately $35. Factoring in the average duration for manual, hypothesis-driven searches, this translates to a total cost of $70 for the entire undertaking.

Fig. 3 illustrates an LLM agent's use of input and output tokens. Corresponding costs subsequently depend on the LLM chosen,

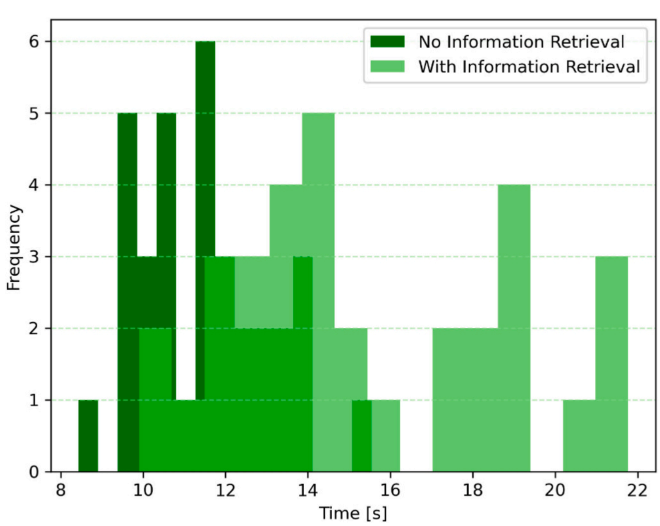

Fig. 2. Distribution of runtime of the agent clustered by the need or absence of additional information retrieval from external APIs.

[4] https://sifted.eu/articles/vc-salary-germany-europe

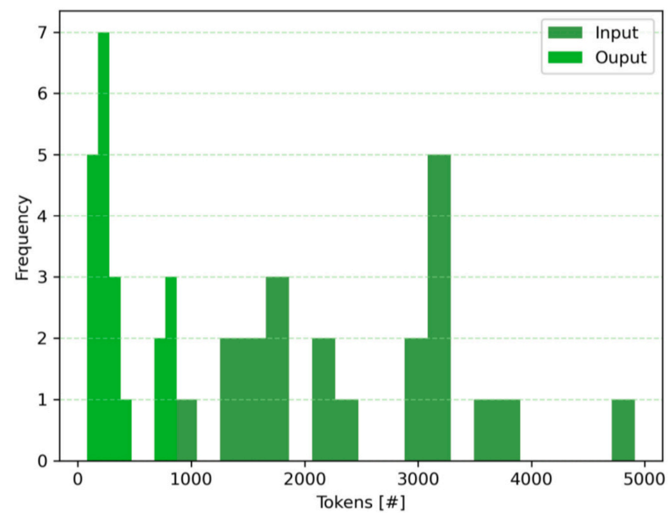

Fig. 3. Distribution of input and output tokens to process a given query. Output queries represent generated content by LLM; input represents utilized tokens in agent and tool prompts.

particularly whether the model is accessed via a commercial API or an open-source provider. The cost is determined mainly by the number of input tokens, which can be explained by the complex prompting involved when setting up an LLM agent (Yao et al., 2022a) and considering that the returned output mainly comprises a display of database entries in which no additional text generation is required. The strongly observed variance in the number of input tokens consumed can be explained by the varying number of input tools used for different prompts to the agent. This also justifies the segregated appearance of prompt tokens in disjointed clusters, mainly driven by the number and kind of tools used to answer a given query.

In the case at hand, we used OpenAI's *GPT-3.5-turbo* and *GPT-4*, at a cost of 6.9 cents per request on average. Compared with a human analysis by a VC, this is an improvement by a factor of 1000. While these are impressive findings, they are of little worth if the identified ventures' quality leaves anything to be desired. Therefore, we investigated the LLM agent's performance when executing a hypothesis-based search concerning the Silhouette metric and Calinski-Harabasz Index

Table 2
Evaluation of LLM agent, human VC analysts, and a random baseline sorting based on the unsupervised learning metrics of Silhouette metric and Calinski-Harabasz Index.

	No. of Clusters	Silhouette Score	Calinski-Harabasz
Baseline Estimators			
Random Sorting	6	−0.032	1.01
Human VC Analyst	3 disjointed	0.370	8.43
Early-career analysts	–	0.367	8.35
Experienced investment professionals	–	0.373	8.51
LLM Agents			
Using Mean Aggregation	6	0.142	15.48
Using Max Aggregation	6	0.223	12.57
Using Mean Aggregation	3 disjointed	0.350	14.32

The early-career analysts group comprises two individuals with distinct academic backgrounds: one participant possesses a technical background with 16 months of professional experience or equivalent training, while the second has a business-oriented educational foundation with 14 months of industry experience or formal training. The experienced investment professionals group consists of two senior members: an investment manager with technical expertise and four years of professional experience in the field, and a partner with academic credentials in business and economics complemented by eight years of investment management experience.

introduced earlier.

In terms of categorization quality, Table 2 presents the evaluation of the LLM agent, human VC analyst, and a random baseline sorting based on the unsupervised learning metrics of the Silhouette metric and the Calinski-Harabasz Index. As a baseline, we randomly selected ventures uniformly and, without replacement from the database, distributed them into six buckets, then henceforth applied the two metrics of cluster quality. *Random sorting* served as a baseline to demonstrate expected performance without any intelligent clustering. Typically, it results in poor clustering quality, as captured by the low Silhouette scores and Calinski-Harabsz Index, thereby illustrating the minimal structure expected without undertaking more sophisticated analysis. As expected, a Silhouette score of −0.0327 using cosine distance and a Calinski-Harabasz Index of 1.01 indicated poor clustering quality with unclear decision boundaries between the clusters (Caliński & Harabasz, 1974).

To ensure the agent's validity, we compared the findings on the LLM agent with an analysis conducted by human VC analysts, which represents the performance of the domain experts in clustering ventures. These scores provide a benchmark to compare against the LLM agent. As two human analysts are expected to differ in their results regarding such an inquiry, we did not expect a perfect match between the LLM agent and human analyses. However, achieving a certain level of overlap would serve as an additional indicator of the agent's trustworthiness. The LLM agent's performance was evaluated using various aggregation methods, such as mean and max aggregation. Considering that different aggregation methods can lead to different clustering outcomes, this evaluation helps identify the best-performing method. By comparing mean and max aggregation methods, we can determine the approach that yields higher-quality clusters based on the Silhouette metric and Calinski-Harabasz Index.

To interpret the scores, we compare the automated clustering (also tested using max aggregation) with human analysts and the random baseline. The Silhouette metric measures the cohesion and separation of clusters (in which case, higher scores indicate better-defined clusters), while the Calinski-Harabasz Index assesses the ratio of the sum of between-cluster dispersion to within-cluster dispersion. For the latter, higher values indicate better clustering performance. By tasking the LLM agent with conducting the corresponding hypothesis-based searches, utilizing the entire dataset of 61,814 ventures, we observe a Silhouette metric of 0.142 and a Calinski-Harabasz Index of 15.48. The findings reveal an enhanced Silhouette score of 0.223 when employing the maximum aggregation method, while a minor decline is observed in the CH Index, which registers 12.57. One possible interpretation of these outcomes is that utilization of the maximum method optimizes the resultant clusters in terms of internal cohesiveness and distinctiveness, yet it does not fully exploit the variance and distribution of the entire dataset. Conversely, the mean aggregation method takes a more comprehensive perspective of the data into consideration, resulting in less well-defined and more disjointed clusters. Nevertheless, these clusters do a better job of capturing the overall data distribution.

We compare these results with the selection of human VC analysts, who are asked to identify suitable ventures from the same industries and, henceforth, to form distinct clusters. The resulting buckets show a slightly improved Silhouette score of 0.37, but a significantly reduced Calinski-Harabasz Index of 8.43. Indeed, this result reflects one's expectations about human performance in such a task. While the high Silhouette score indicates the analyst's ability to group ventures into distinct and well-formed clusters, the dataset's large size impedes the analyst from getting a comprehensive overview of the dataset. Thus, the spread of data points within the clusters is insufficient when compared with the variance in the entire ground truth, indicated by an increase of 70 % in the Calinski-Harabasz Index, from 8.43 for humans to 14.32 for LLM agents.

Finally, we compare early-career analysts vs experienced investment professionals. The Silhouette Scores (0.373 vs 0.367) and Calinski-Harabasz Index values (8.51 vs 8.35) are slightly higher experienced professionals relative to early-career analysts. While these differences are modest and our limited sample of four analysts precludes robust statistical inference, the results nonetheless point to experience contributing positively to categorization quality. This finding may potentially be interpreted in two ways: it may validate the enduring value of professional experience in investment assessment, or alternatively, indicate that generative AI technologies, which mainly helps to perform high-skill tasks (Bloom et al., 2025), may exert a more pronounced impact on roles requiring less specialized experience—a hypothesis consistent with recent academic literature (Brynjolfsson et al., 2025; Noy & Zhang, 2023).

While this analysis indicates the LLM agent's competence to create sensible and well-formed clusters, further data collection is needed to determine whether the resulting ventures correspond to the initially formulated hypothesis. For this reason, we defined 10 industry sections and tasked VC analysts to identify 10 suitable ventures from the dataset for each case. We then tasked the agent with finding suitable ventures that correspond to the industry sections using different modes of aggregation and cluster sizes. While certainly no complete match is expected (as tasking another VC analyst with the same task would not produce such a match), a partial overlap among selected ventures would indicate correspondence between the initial hypothesis and selected ventures.

6. Ecological validity

6.1. Generalizability through real-world application

A critical aspect of assessing the effectiveness of GenAI versus human VC analyst clustering is ensuring that our findings are applicable beyond a specific dataset or controlled conditions. Unlike experimental approaches or simulations, which may lack external validity, our analysis is firmly grounded in a real-world context. This is made possible not only through privileged access to a private dataset from an actual VC firm—allowing us to work with real ventures—but also through direct collaboration throughout the process. Importantly, the VC fund continues to rely on this approach, further strengthening its real-world applicability and generalizability. In a retrospective discussion, a senior VC analyst noted:

> "The main goal was increasing the efficiency of our very early deal flow pipeline. We were focusing on that because this was a process that used to engage a small team to actually look at a large number of ventures. If LLMs do this, we save time for the later due diligence and reduce the number of wrong decisions later."

This perspective highlights how the integration of LLMs, beyond efficiency enhancements, provides a scalable solution for venture screening, supporting the argument for generalizability.

6.2. Validation

In our analysis, we deliberately avoided using venture success as a performance metric, not only because our approach originates from a real-world application—where we cannot determine *ex ante* the success of ventures under consideration—but, more importantly, to mitigate potential endogeneity bias (Hellmann & Puri, 2000; Kaplan & Strömberg, 2001). Our approach, therefore, departs from Becker-like tests that interpret *ex post* realized performance as *ex ante* patterns of error in VCs' decisions. Nevertheless, it would be interesting to evaluate whether the clusters formed by LLMs lead to better investment decisions or higher returns compared to the screening performed by human analysts. This is empirically challenging, as we are dealing with early-stage ventures of potential interest to VC firms, some of which never raised funds and are invisible to data providers such as Pitchbook or Crunchbase. Moreover, the data does not indicate whether, or which, ventures actually sought funding and were denied.

Although data on screening decisions for early-stage ventures do not allow for the same type of risk-return analysis typically used with publicly listed firms or other alternative investments, we attempt to assess whether ventures selected by human analysts or by the LLM agent are more likely to survive or raise funds. We run logistic regressions on the (*ex post*) probability of venture survival (to proxy negative performance) and funding (to proxy positive performance), using a sample of 9911 ventures screened by the VC fund Freigeist Capital in October–November 2023.

Risk assessment in early-stage ventures relies on the quantification of survival profiles. We defined *Survival* as a dummy variable equal to 1 if the venture is classified as active as of April 12, 2025. To determine survival, we use a proxy based on website operability (Arora et al., 2016). The absence of an active website can be interpreted as a signal of firm inactivity, consistent with findings by Signori and Vismara (2018), who show that discontinued web presence is often linked to venture failure. Specifically, we developed an automated script in Python, run in a Jupyter Notebook, that sends HTTP HEAD requests to each venture's website. We interpret a response code of 200 (OK) as evidence of an active and maintained web presence—assumed to correspond to an operational website. All other responses (including timeouts, unresolved redirects, or server errors) are treated as indicators of inactivity.

We cannot measure the return on a potential investment in the early-stage ventures screened by the VC fund, as some of these ventures never raised external funding, and those that have not yet provided an exit opportunity for external investors. However, we can further distinguish among the surviving ventures between those that did not receive any funding and those that secured a round of financing after the screening conducted in October–November 2023. We define *Funding* as a dummy variable equal to 1 if the venture received external funding between the Freigeist screening in October–November 2023 and April 12, 2025. Although this measure does not directly capture returns, we argue that it serves as a meaningful proxy for positive performance, as successful fundraising represents a critical milestone in enabling growth potential.

The explanatory variables include Human VC, a dummy equal to 1 if the venture was selected by at least one of the Freigeist human analysts, and LLM Agent, a dummy equal to 1 if the venture was selected by the LLM agent. We control for venture size, measured as the log number of employees at the time of Freigeist's screening—a commonly used proxy for firm size in corporate finance research. We also include industry and country fixed effects to account for unobserved heterogeneity across sectors and geographies.

Results reported in Table 3 show that ventures selected by either human analysts or the LLM agent are more likely to raise funding (Model 3). This pattern suggests that the LLM reflects alternative indicators of quality—signals that human analysts may deprioritize. The predictive value of the LLM in cases without human involvement highlights its potential to surface promising ventures that may otherwise be overlooked in traditional screening. Ventures selected by the LLM agent are also more likely to survive (Model 1: $\beta = 0.638$; $p < 0.01$), whereas the selection by human analysts is only weakly significant ($\beta = 0.199$; $p < 0.10$). The interaction between human analysts and the LLM agent is insignificant for funding (Model 4) and positive—though only weakly significant—for survival (Model 2: $\beta = 1.510$; $p < 0.10$).

Taken together, these results suggest that LLMs and human decision-making follow complementary evaluative patterns. While we acknowledge the limitations of our analysis of the post-screening performance of early-stage ventures (for example, firm size remains an easily measurable and strong predictor of performance), we remark that it is intended solely as a validation of the potential of using LLMs for VC screening decisions. Overall, these findings reinforce the broader conclusion of this work: LLMs can contribute to venture screening—particularly in high-volume or resource-constrained environments—while supporting earlier evidence of their strong clustering performance.

Table 3
Validation.

	(1)	(2)	(3)	(4)
	Survival		Funding	
Human VC Analyst	0.199*	0.203	0.697***	0.899***
	(0.111)	(0.132)	(0.082)	(0.102)
LLM Agent	0.638***	0.612**	0.403**	0.792***
	(0.172)	(0.279)	(0.159)	(0.265)
Human VC x LLM Agent	–	1.510*	–	−0.283
	–	(0.794)	–	(0.610)
Size	0.503***	0.522***	0.678***	0.680***
	(0.026)	(0.027)	(0.023)	(0.023)
Industry fixed effects	Yes	Yes	Yes	Yes
Country fixed effects	Yes	Yes	Yes	Yes
Constant	0.527***	1.066	−2.017***	−2.178***
	(0.043)	(1.102)	(0.048)	(0.770)
Pseudo R-squared	0.048	0.061	0.101	0.135

Logistic regression on the probability of venture survival (Models 1 and 2) and funding (Models 3 and 4). The analysis is based on a sample of 9911 ventures, screened by VC fund Freigeist Capital in October–November 2023. The dependent variable in Models 1 and 2 (Survival) is a dummy variable equal to 1 if the venture is classified as active (as of April 12, 2025) and 0 otherwise. To determine survival, we use a proxy based on website operability (Arora et al., 2016). The absence of an active website can therefore be interpreted as a signal of firm inactivity, consistent with findings by Signori and Vismara (2018), who show that discontinued web presence is often linked to failed ventures. Specifically, we developed an automated script in Python run in a Jupyter Notebook that sends HTTP HEAD requests to each venture's website. We interpret a response code of 200 (OK) as evidence of an active and maintained web presence—assumed to be an operational website. All other responses (including timeouts, redirects without resolution, or server errors) are treated as indicators of inactivity. The dependent variable in Models 3 and 4 (Funding) is a dummy variable equal to 1 if the venture received funding between the screening process performed by Freigeist in October–November 2023 and April 12, 2025. Human VC is a dummy variable equal to 1 if the venture was selected by at least one of the Freigeist human analysts, while LLM Agent is equal to 1 if the venture was selected by the LLM Agent. Size is measured as the log number of employees at the time of Freigeist's screening and serves as commonly used proxy for venture size in corporate finance research. We also include industry and country fixed effects to account for unobserved heterogeneity across sectors and geographies. See Appendix 4 for the descriptive statistics. Standard errors are presented in parentheses. Significance levels are denoted as follows: *** $P < 0.01$, ** $P < 0.05$, * $P < 0.10$.

7. Implications

Beyond technical performance metrics, LLM implementation in VC workflows has operational implications. The tool could significantly enhance the day-to-day operations of a venture capital firm by streamlining several critical workflows. It can assist in deal sourcing and screening by rapidly classifying and clustering large volumes of incoming pitch decks and company descriptions, helping VCs prioritize high-potential opportunities without manually reviewing each submission. During due diligence, the tool can extract key insights from unstructured data, such as founders' backgrounds or market trends, and flag promising attributes or potential risks based on historical patterns. Its clustering functionality can map ventures against the competitive landscape, identifying gaps, overlaps, or emerging clusters in specific sectors, enabling more informed investment decisions.

As one VC professional involved in the study explained:

> "We wanted to make our very early deal flow pipeline more efficient. We used to hire interns to filter out startups for us, which meant critical early decisions were made by less-experienced individuals. With LLMs, we now conduct these screenings ourselves and have eliminated the need for interns, reducing wrong early-stage decisions and filtering out 70% of startups that do not fit our thesis."

This strengthens the argument about how LLMs streamline early-stage hypothesis-driven venture selection, shifting the screening process from subjective-manual-decision-making to a more systematic one. Importantly, VCs highlighted that LLMs reduce bias in early-stage filtering, whereby subjective factors such as website design and founder presentation drive decisions. By providing data-driven insights, the tool supports collaborative decision-making, acting as a neutral reference point to validate or challenge subjective opinions.

These practitioners' insights indicate that LLMs not only accelerate the early screening process, but they also successfully introduce a much more structured mechanism. While LLMs have proven valuable for early-stage filtering, their effectiveness diminishes as the investment process moves into deeper due diligence.

As one VC noted:

> "When it comes to due diligence, we also use LLMs—but not in such a systematic manner. Instead, we use them as a research tool, rather than for structured analysis."

This suggests that while LLMs excel in broad screening, they do not replace the deeper, qualitative aspects of venture evaluation, such as market fit, founder potential, and financial modeling. Instead, VCs see LLMs as a complementary tool—useful for synthesizing information, but not yet a replacement for rigorous human analysis at later investment stages.

At a practical level, our findings highlight several key takeaways for VC firms adopting LLMs. First, LLMs are most effective in early-stage venture screening, where they can replace human-driven filtering. This may democratize entrepreneurial finance (Cumming et al., 2021) by reducing barriers to entry into private equity for small investors through lowered screening costs, and by mitigating cognitive and cultural biases inherent in human decision-making. VCs, despite their professional incentives, are subject to human cognitive limitations when making decisions under conditions of uncertainty. This decision-making context leads to the formation of stereotypical heuristics that disproportionately emphasize certain traits perceived to correlate with entrepreneurial success. Such cognitive shortcuts ultimately contribute to the documented homogeneity within venture capital funding patterns (Bates & Bradford, 2008; Bengtsson & Hsu, 2015; Butticè et al., 2024; Howell & Nanda, 2024; Lerner & Nanda, 2020).

Second, LLMs should be used for hypothesis-driven categorization – as they prove to work best when trained on a VC firm's investment thesis, ensuring alignment between automated selection and strategic goals. Although our benchmark includes experienced VC analysts, it is important to note that the task of initial venture screening is typically performed by early-career analysts or interns. Thus, our insights speak of GenAI being more likely to impact lower-experience roles, as sometimes anticipated in literature (Brynjolfsson et al., 2025; Noy & Zhang, 2023) or the press.[5] However, while LLMs excel at information synthesis, they do not replace human judgment in later investment stages.

Complicating this picture is the phenomenon of "mechanized convergence" (Sarkar, 2023) in AI-assisted knowledge workflows, where users leveraging generative AI tools produce less diverse outcomes for identical tasks compared to those working without such assistance (Lee et al., 2025a, b). While LLMs enhance efficiency in venture screening processes, they may simultaneously reduce critical engagement, particularly in routine early-stage filtering tasks where VC professionals might simply defer to AI recommendations. This raises significant concerns about long-term reliance and the potential erosion of independent analytical capabilities within investment teams. As one senior VC professional remarked, "I think that LLMs help remove this bias by focusing on whether a venture matches our thesis or not," indicating a shift toward algorithmic dependence that may standardize decision-making across the industry. This could ultimately lead to a transition in the VC industry from direct task execution to oversight roles when employing generative AI tools.

Our findings yield important recommendations for policymakers. From a regulatory perspective, generative AI tools should be recognized as a means to lower entry barriers for smaller funds and emerging VC ecosystems by reducing the resource intensity of early-stage screening. Policymakers aiming to foster innovation financing should actively encourage the responsible adoption of such tools—for instance, by establishing standardized data-sharing infrastructures that reduce biases and limit the risks of hallucination. Moreover, regulation should explicitly promote human-in-the-loop practices, ensuring that AI complements rather than substitutes human expertise in financial decision-making. Embedding these principles into policy frameworks would align with broader agendas on trustworthy AI, reinforcing commitments to transparency, accountability, and safeguards against overreliance.

8. Conclusions

This paper investigates whether LLMs can assist VCs in their screening decisions. Our results regarding cost and speed demonstrate a substantial improvement in efficiency when selecting suitable ventures through hypothesis-driven searches. These findings suggest that GenAI will pose significant implications for VCs, reflecting the broader transformative power of GenAI in venture investing. For VC firms, the strategic use of LLM technology has the potential to surpass the performance of larger and more reputable firms that do not employ similar technology. Consequently, data collection, algorithms, and server capacity will become as crucial as traditional success factors, such as reputation (Nahata, 2008).

However, these larger firms will face increasing competition from smaller LLM-enhanced VCs. The implications extend to early-stage ventures and their entrepreneurs. Agent-based screening has the potential of making VC funding more accessible to ventures outside of traditional ecosystems and networks. Historically, VC investment outcomes have favored certain geographic regions and industry networks significantly (Sorenson & Stuart, 2001), thereby selectively benefiting these established circles' economic development (Florida & Kenney, 1988). By reducing personal connections' importance in obtaining VC financing, the adoption of LLM agents will benefit ventures from remote areas (Fisch et al., 2022), potentially contributing to the democratization of fintech and digital finance (e.g., Butticè & Vismara, 2022; Cumming et al., 2021).

Despite these promising implications, our study has limitations that suggest avenues for future research. First, outcomes depend on the scope and quality of venture data; richer longitudinal datasets would allow more refined analyses, including supervised approaches linking clustering to realized outcomes. Second, our results reflect one model version. With rapid LLM advancements—such as database connectivity and improved reasoning—future iterations are likely to enhance the robustness, efficiency, and scalability of LLM-assisted venture screening.

Appendix 1. Model

An agent with the overall goal of finding suitable ventures for a given hypothesis operates in the following way (Yao et al., 2022a): Let $\mathscr{A}$ be the set

[5] The Economist. (2025, February 13). "How AI will divide the best from the rest." Retrieved from https://www.economist.com/briefing/2025/02/13/how-ai-will-divide-the-best-from-the-rest.

of the agent's feasible actions and T the set of tools, then $T \subset \mathscr{A}$. Let $\mathscr{L}$ be the set of language, in which we defined $\widehat{\mathscr{A}} = \mathscr{A} \cup \mathscr{L}$ as the augmented set of actions and $\widehat{a}_t \in \mathscr{L}$ as a thought. Thus, we defined the generation of natural language as being analogous to the choice and execution of a tool. Let o_1 be the initial observation of the agent, i.e., a user prompt.

Now let

$$\|\cdot\|_a : \mathscr{H} \subset \mathscr{L} \rightarrow R_0^+$$

be a measure of the *actionability* of a hypothesis in written language, with $\mathscr{H}$ the space of possible hypotheses.[6] In this respect, actionability refers to the degree of feasibility and practicability of a hypothesis. Thus, a very general hypothesis with little specificity would score rather poorly and first would need structurization and compartmentalization. To clarify this, using an example, a theoretical query tasking the agent to identify ventures that aim to improve users' happiness could be viewed as rather unspecific. However, if the agent would access the external knowledge at its disposal and consult scientific studies investigating sources of unhappiness as a starting point for potential improvements, the resulting search would be much more concrete and promising. While such a measure is hardly computable in practice, it serves as a basis for a concise definition of the LLM agent and to clarify its potential and purpose.

Finally, let S be a set of venture companies and

$$d : \mathscr{H} \times S \rightarrow R_0^+$$

be a metric describing the distance between a hypothesis and a suitable venture fulfilling the hypothesis. The agent's work then is the result of the following optimization problem:

$$\min_{s \in S} d\left(\operatorname{argmax}_{\widehat{a}_T \circ \ldots \circ \widehat{a}_1 \in \widehat{\mathscr{A}}^{\times T}} ||\widehat{a_T} \circ \ldots \circ \widehat{a_1}(o_1)||_a, s\right)$$

for $T \leq T_{max}$. Here, T_{max} is a hyperparameter in need of specification before running the model.

To evaluate our approach's success, a suitable and practically feasible definition of *d* is critical. However, given the matter's subjective and vague nature, it has proven to be a non-trivial endeavor.

Appendix 2. Examples

Exemplary queries executable by the LLM agent pair with the top three returned answers. Returned company names are anonymized.

Description	Country	Funding	Employees
Prompt: Find ventures from the area of carbon accounting from the DACH region with less than $100,000 in funding			
Company 1 is a data and services provider that helps organizations measure and manage chemicals and plastics' environmental impacts through their LCA database, carbon footprint data, and consulting services.	DE	0	25
Company 2 offers a carbon dashboard and allows users to calculate and compensate for their greenhouse gas emissions, including a one-off purchase or a subscription-based model for carbon dioxide removal.	CH	0	2
Company 3 provides a transparent carbon registry and marketplace for data-driven, consumer-generated carbon credits in Europe to incentivize adoption of energy-efficient green technologies, reduce carbon emissions, and accelerate progress toward global climate targets.	DE	0	2
Prompt: Find ventures with no funding, but more than 25 employees			
Company 4 is a new-generation construction management software provider that optimizes the administration, planning, and project management of construction companies.	NL	0	26
Company 5's Digital Experience Platform pools all marketing and communications content within one information-centric system and a consistent user interface. The best-of-breed applications for collaboration—DAM, PIM, Omnichannel CMS, and MRM—create relevant digital experiences from real-time information for the first time.	DE	0	202
Company 6 is a young and dynamic company with the single goal of accelerating clinical trials, i.e., it wants to speed up drug development. Its young and highly motivated team has all the necessary skills and knowledge to support our customers. Company 6 combines expertise from clinical trials and information technology sectors to bring a new breed of networked services to the pharmaceutical industry.	DE	0	115
Prompt: Access recent publications on ArXiv to find new technologies in solid-state batteries and find corresponding ventures with more than $500,000 in funding.			
Company 7 produces 3D solid-state batteries that offer higher performance, safety, and sustainability, while also being scalable and using sustainable materials. They are retrofittable into existing commercial formats.	NL	5.6 m	16
Company 8 is developing a low-cost and sustainable redox flow battery for energy storage on a multi-day basis. The technology is fully compliant with circular economy principles and allows for penetration of renewable energy up to 90 %. The company uses a sulfur byproduct of the petrochemical industry, which opens new market opportunities and favors energy communities' independence.	IT	1.8 m	5

(*continued on next page*)

[6] Please note that in our framework, "actionability" refers to LLM-generated outputs' ability to be as relevant and useful as possible for subsequent querying of the database. Specifically, the descriptions produced by the LLM should align with the language and "vibe" of the startup descriptions in the database, facilitating effective querying without relying on simple keyword matching or traditional methods. For example, if the input requests learning-related startups, the LLM is prompted to provide descriptions that lead to relevant results in the venture database, even if these startups do not explicitly mention terms such as "learning" or "education" in their descriptions. The actionability step is an implicit part of the agent's reasoning process and cannot be directly measured. Instead, it reflects the LLM's capacity to produce descriptions that correspond well with the database.

(*continued*)

Description	Country	Funding	Employees
Company 9 develops high-power, mechanical flywheel energy storage for short-duration and high-power applications, such as ultra-rapid EV charging that would quickly degrade batteries. The company replaces toxic battery chemicals with more convenient and less-carbon-embedded materials, such as steel and copper, which are more affordable and easily recycled at end-of-life.	GB	3.37 m	13
Prompt: Find companies corresponding to modular, low-cost robotic product offerings			
Company 10 offers a marketplace and modular system for automation with robots, allowing small and medium-sized businesses to implement robot solutions for their individual needs and budgets. The company facilitates selection of compatible products for cost-effective implementation and provides expert advice and verified vendors.	DE	0	5
Company 11 designs and sells intelligent autonomous robots, including ready-to-use models and customization options for any service robot application. It specializes in both industrial and service robotics, including general-purpose robotic vehicles, indoor localization systems, and advanced robotic components. Furthermore, it offers components for mobile robots and software solutions, as well as development and support services.	FR	0	1
Company 12 provides AI-based modular software solutions for complex motion and handling processes in industrial production environments. Its solutions enable efficient execution of typical tasks and reduce robot implementation times, such as autonomous object recognition and autonomous planning of robot arm movements in handling processes.	DE	0	6

Appendix 3. Agent evaluation metrics

As commonly done in natural language processing (Kenny et al., 2013; Sanh et al., 2019), we employed the cosine distance to compare the distance between two embeddings of textual information. For two n-dimensional vectors x, y (i.e., embeddings in this case), it is given by

$$\mathrm{d}_{\cos}(x,y) = \frac{\sum_{i=1}^{n} x_i y_i}{\sqrt{\sum_{i=1}^{n} x_i^2 \cdot \sum_{i=1}^{n} y_i^2}}$$

Based on this, we measure the Silhouette score to identify how well the agent has chosen the ventures based on a certain hypothesis. The measure is designed to quantify the similarity of a vector within a cluster compared with the other clusters at hand.

Let $i \in C_I$ be an element of a cluster C_I. Then let.

$a(i) = \frac{1}{|C_I|-1}\sum_{j\in C_I, i\neq j} d_{cos}(i,j)$ and $b(i) = \min_{J\neq I}\frac{1}{|C_J|}\sum_{j\in C_J} d_{cos}(i,j)$.

Then the Silhouette score of a dataset D is given as

$$s = \max_{i\in D}\frac{b(i) - a(i)}{\max(a(i), b(i))}$$

with $-1 \le s \le 1$. A value of $s = 1$ signals an optimal separation of the clusters with small intra-cluster distances between points, $s = 0$ suggests a strong overlap of clusters, whereas $s = -1$ indicates an incorrect assignment of an element.

As an addition to the Silhouette score, we employed the Calinski-Harabasz Score (Caliński & Harabasz, 1974), which is a measure of the ratio of between-cluster dispersion and within-cluster dispersion. It is written as ch, with $ch \ge 0$. Higher values of the measure indicate enhanced cluster constellations, as the inter-cluster variance is expected to be much higher than the dispersion measured within a well-defined cluster.

Let for dataset D and cluster $C_i, i = 1, \ldots, K$.

$c = \frac{1}{|D|}\sum_{x\in D} x$ and $c_i = \frac{1}{|C_i|}\sum_{x\in C_i} x$

The Calinski-Harabasz Score then is given by

$$ch = (\mathrm{n}-k)\frac{\sum_i |C_i|\cdot||c_i - c||^2}{(n-1)\sum_i\sum_{x\in C_i}||x - c_i||^2}$$

Appendix 4. Descriptive statistics

Descriptive statistics and correlation matrix for the validation analysis (see Table 3).

Variables	Mean	S. D.	(1)	(2)	(3)	(4)	(5)
(1) Survival	0.784	0.411	1.000				
(2) Funding	0.324	0.468	0.083*	1.000			
(3) Human VC Analyst	0.075	0.263	0.054*	0.138*	1.000		
(4) LLM Agent	0.075	0.263	0.039*	0.024*	0.094*	1.000	
(5) Size	1.776	1.180	0.340*	0.205*	0.138*	−0.016	1.000

* Indicates significance at the 0.05 level.

Data availability

The authors do not have permission to share data.

References

Zhang, D., Wang, C., He, Y., & Vigne, S. A. (2024). Does FinTech efficiently hamper manipulating ESG data behavior? *British Accounting Review, 101494*, 1–23.

Antretter, T., Blohm, I., Sirén, C., Grichnik, D., Malmström, M., & Wincent, J. (2020). Do algorithms make better-and fairer-investments than angel investors? *Harvard Business Review, 2-6*. Digital Articles.

Ardekani, A. M., Bertz, J., Bryce, C., Dowling, M., & Long, S. C. (2024). FinSentGPT: A universal financial sentiment engine? *International Review of Financial Analysis, 94*, Article 103291.

Arora, S. K., Li, Y., Youtie, J., & Shapira, P. (2016). Using the Wayback machine to mine websites in the social sciences: A methodological resource. *Journal of the Association for Information Science and Technology, 67*(8), 1904–1915. https://doi.org/10.1002/asi.23503

Aziz, S., Dowling, M., Hammami, H., & Piepenbrink, A. (2022). Machine learning in finance: A topic modeling approach. *European Financial Management, 28*(3), 744–770.

Bai, Y., Ying, J., Cao, Y., Lv, X., He, Y., Wang, X., & Hou, L. (2024). Benchmarking foundation models with language-model-as-an-examiner. *Advances in Neural Information Processing Systems, 36*.

Bates, T., & Bradford, W. D. (2008). Venture-capital investment in minority business. *Journal of Money, Credit and Banking, 40*(2–3), 489–504.

Beckmann, L., & Hark, P. F. (2024). ChatGPT and the banking business: Insights from the US stock market on potential implications for banks. *Finance Research Letters, 63*, Article 105237.

Bellardini, L., Del Gaudio, B. L., Previtali, D., & Verdoliva, V. (2022). How do banks invest in fintechs? Evidence from advanced economies. *Journal of International Financial Markets, Institutions and Money, 77*, Article 101498.

Bengtsson, O., & Hsu, D. H. (2015). Ethnic matching in the US venture capital market. *Journal of Business Venturing, 30*(2), 338–354.

Bernstein, S., Giroud, X., & Townsend, R. R. (2016). The impact of venture capital monitoring. *Journal of Finance, 71*(4), 1591–1622.

Blohm, I., Antretter, T., Sirén, C., Grichnik, D., & Wincent, J. (2022). It'sa peoples game, isn't it?! A comparison between the investment returns of business angels and machine learning algorithms. *Entrepreneurship Theory and Practice, 46*(4), 1054–1091.

Bloom, D. E., Prettner, K., Saadaoui, J., & Veruete, M. (2025). Artificial intelligence and the skill premium. *Finance Research Letters, 81*, 1–7, 107401.

Boussioux, L., Lane, J. N., Zhang, M., Jacimovic, V., & Lakhani, K. R. (2023). *The crowdless future? How generative AI is shaping the future of human crowdsourcing.* Harvard Business School.

Braun, R., Fernández Tamayo, B., Lopez de Silanes, F., Phalippou, L., & Sigrist, N. (2023). *Limited partners versus unlimited machines; artificial intelligence and the performance of private equity funds. SSRN working paper no. 4490991*.

Brynjolfsson, E., Li, D., & Raymond, L. (2025). Generative AI at work. *The Quarterly Journal of Economics, 140*(2), 889–942.

Butticè, V., Colombo, M. G., & Rovelli, P. (2024). Venture capital and the delegation of decision authority in startups: An exploratory study. *Journal of Industrial and Business Economics, 51*(4), 893–923.

Butticè, V., & Vismara, S. (2022). Inclusive digital finance: The industry of equity crowdfunding. *Journal of Technology Transfer, 47*(4), 1224–1241.

Caliński, T., & Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics-Theory and Methods, 3*(1), 1–27.

Cappa, F., Oriani, R., Peruffo, E., & McCarthy, I. (2021). Big data, a valuable resource in the digitalized environment? Unpacking the effects of volume, variety and veracity on firm performance. *Journal of Product Innovation Management, 38*(1), 49–67.

Chan, K., Covrig, V., & Ng, L. (2005). What determines domestic bias and foreign bias? Evidence from mutual fund equity allocations worldwide. *Journal of Finance, 60*, 1495–1534.

Collevecchio, F., Cappa, F., Peruffo, E., & Oriani, R. (2024). When do M&As with fintech firms benefit traditional banks? *British Journal of Management, 35*(1), 192–209.

Csaszar, F. A., Ketkar, H., & Kim, H. (2024). Artificial intelligence and strategic decision-making: Evidence from entrepreneurs and investors. *Strategy Science, 9*(4), 322–345.

Cumming, D., Meoli, M., & Vismara, S. (2021). Does equity crowdfunding democratize entrepreneurial finance? *Small Business Economics, 56*(2), 533–552.

De Prado, M. L. (2018). *Advances in financial machine learning*. John Wiley & Sons.

Dealroom. (2024). Europe: Guide to startups and venture capital. https://dealroom.co/guides/europe.

Dowling, M., & Lucey, B. (2023). ChatGPT for (finance) research: The Bananarama conjecture. *Finance Research Letters, 53*, Article 103662.

Elango, B., Fried, V. H., Hisrich, R. D., & Polonchek, A. (1995). How venture capital firms differ. *Journal of Business Venturing, 10*(2), 157–179.

Feuerriegel, S., Hartmann, J., Janiesch, C., & Zschech, P. (2024). Generative ai. *Business & Information Systems Engineering, 66*(1), 111–126.

Fisch, C., Meoli, M., & Vismara, S. (2022). Does blockchain technology democratize entrepreneurial finance? An empirical comparison of ICOs, venture capital, and REITs. *Economics of Innovation and New Technology, 31*(1–2), 70–89.

Florida, R., & Kenney, M. (1988). Venture capital and high technology entrepreneurship. *Journal of Business Venturing, 3*(4), 301–319.

Fried, V. H., & Hisrich, R. D. (1994). Toward a model of venture capital investment decision making. *Financial Management, 23*(3), 28–37.

Gompers, P., Gornall, W., Kaplan, S. N., & Strebulaev, I. A. (2020). How do venture capitalists make decisions? *Journal of Financial Economics, 135*(1), 169–190.

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

Goodell, J. W., Jabeur, S. B., Saâdaoui, F., & Nasir, M. A. (2023). Explainable artificial intelligence modeling to forecast bitcoin prices. *International Review of Financial Analysis, 88*, Article 102702.

Hall, J., & Hofer, C. W. (1993). Venture capitalists' decision criteria in new venture evaluation. *Journal of Business Venturing, 8*(1), 25–42.

Hannigan, T. R., McCarthy, I. P., & Spicer, A. (2024). Beware of botshit: How to manage the epistemic risks of generative chatbots. *Business Horizons, 67*(5), 471–486.

Hellmann, T., & Puri, M. (2000). The interaction between product market and financing strategy: The role of venture capital. *Review of Financial Studies, 13*(4), 959–984.

Hirshleifer, D., Lim, S. S., & Teoh, S. H. (2009). Driven to distraction: Extraneous events and underreaction to earnings news. *Journal of Finance, 64*(5), 2289–2325.

Howell, S. T., & Nanda, R. (2024). Networking frictions in venture capital, and the gender gap in entrepreneurship. *Journal of Financial and Quantitative Analysis, 59*(6), 2733–2761.

Huang, W., Abbeel, P., Pathak, D., & Mordatch, I. (2022). Language models as zero-shot dplanners: Extracting actionable knowledge for embodied agents. In *International conference on machine learning* (pp. 9118–9147). PMLR.

IngestAI.. (2023). *AI tools for venture capital in 2024*. IngestAI. Retrieved from https://ingestai.io/blog/ai-tools-for-venture-capital.

Jang, Y. S., & Kaplan, S. N. (2025). *Venture capital start-up selection (NBER working paper No. 33483)*. National Bureau of Economic Research.

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. *Heuristics and Biases: The Psychology of Intuitive Judgment, 49*(49–81), 74.

Kaplan, S. N., & Strömberg, P. (2001). Venture capitalists as principals: Contracting, screening, and monitoring. *American Economic Review, 91*(2), 426–430.

Kenny, P., Stafylakis, T., Ouellet, P., Alam, M. J., & Dumouchel, P. (2013). PLDA for speaker verification with utterances of arbitrary duration. In *2013 IEEE international conference on acoustics, speech and signal processing* (pp. 7649–7653).

Khurshed, A., Paleari, S., Pande, A., & Vismara, S. (2014). Transparent bookbuilding, certification and initial public offerings. *Journal of Financial Markets, 19*, 154–169.

Klein, G. A. (1998). *Sources of power: How people make decisions*. Cambridge, MA: MIT Press.

Ko, H., & Lee, J. (2024). Can ChatGPT improve investment decisions? From a portfolio management perspective. *Finance Research Letters, 64*(105433), 1–10.

Lee, H. P. H., Sarkar, A., Tankelevitch, L., Drosos, I., Rintel, S., Banks, R., & Wilson, N. (2025). *The impact of generative AI on critical thinking: Self-reported reductions in cognitive effort and confidence effects from a survey of knowledge workers*.

Lerner, J., & Nanda, R. (2020). Venture capital's role in financing innovation: What we know and how much we still need to learn. *Journal of Economic Perspectives, 34*(3), 237–261.

Levy, O., & Goldberg, Y. (2014). Dependency-based word embeddings. In *Proceedings of the 52nd annual meeting of the Association for Computational Linguistics (volume 2: Short papers)*.

Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H., & Neubig, G. (2023). Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *ACM Computing Surveys, 55*(9), 1–35.

Lyonnet, V., & Stern, L. H. (2024). Machine learning about venture capital choices. Texas A&M University working paper. https://mays.tamu.edu/wp-content/uploads/2025/03/288-Machine-learning-about-VC-choices.pdf.

Martin, I. W., & Nagel, S. (2022). Market efficiency in the age of big data. *Journal of Financial Economics, 145*(1), 154–177.

Matthews, M. J., Su, R., Yonish, L., McClean, S., Koopman, J., & Yam, K. C. (2025). A review of artificial intelligence, algorithms, and robots through the lens of stakeholder theory. *Journal of Management, 51*(6), 2627–2676.

MLQ. (2023). *Large language models (LLMs) in venture capital - This week in VC*. MLQ.ai. Retrieved from https://blog.mlq.ai/large-language-models-venture-capital-this-week-in-vc/.

Momtaz, P. P. (2024). Decentralized finance (DeFi) markets for startups: Search frictions, intermediation, and the efficiency of the ICO market. *Small Business Economics, 63*(4), 1415–1447.

Moonfire. (2023, February). In *Language Models in Venture Capital. Moonfire Blog*. https://www.moonfire.com/blog/how-we-use-large-language-models-in-venture-capital.

Nahata, R. (2008). Venture capital reputation and investment performance. *Journal of Financial Economics, 90*(2), 127–151.

Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Science, 381*(6654), 187–192.

Obschonka, M., & Fisch, C. (2022). Artificial intelligence and entrepreneurship research. In *Oxford research encyclopedia of business and management*.

Obschonka, M., Gregoire, D. A., Nikolaev, B., Ooms, F., Lévesque, M., Pollack, J. M., & Behrend, T. S. (2025). Artificial intelligence and entrepreneurship: A call for research to prospect and establish the scholarly AI frontiers. *Entrepreneurship Theory and Practice, 49*(3), 620–641.

Pelster, M., & Val, J. (2024). Can ChatGPT assist in picking stocks? *Finance Research Letters, 59*, Article 104786.

Petroni, F., Rocktäschel, T., Lewis, P., Bakhtin, A., Wu, Y., Miller, A. H., & Riedel, S. (2019). *Language models as knowledge bases? ArXiv Preprint ArXiv:1909.01066*.

Pitchbook. (2018). 2018 VC data usage survey. https://pitchbook.brightspotcdn.com/dc/89/f04cb710458ea799b628cd9a60c6/vc-data-usage-survey-v3-1.pdf.

PitchBook. (2024). Q3 2024 European venture report. https://pitchbook.com/news/reports/q3-2024-european-venture-report.

Reed, S., Zolna, K., Parisotto, E., Colmenarejo, S. G., Novikov, A., Barth-Maron, G., … Springenberg, J. T. (2022). *A generalist agent. Transactions on machine learning research*.

Sahlman, W. A. (1990). The structure and governance of venture-capital organizations. *Journal of Financial Economics, 27*(2), 473–521.

Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2019). *DistilBERT, a distilled version of BERT: Smaller, faster, cheaper and lighter. ArXiv preprint ArXiv:1910.01108*.

Sarkar, A. (2023). Exploring perspectives on the impact of artificial intelligence on the creativity of knowledge work: Beyond mechanised plagiarism and stochastic parrots.

In *Proceedings of the 2nd annual meeting of the symposium on human-computer interaction for work*.
Shepherd, D. A., & Majchrzak, A. (2022). Machines augmenting entrepreneurs: Opportunities (and threats) at the nexus of artificial intelligence and entrepreneurship. *Journal of Business Venturing, 37*(4), Article 106227.
Siddik, A. B., Yong, L., Du, A. M., Vigne, S. A., & Sharif, A. (2025). Harnessing artificial intelligence for enhanced environmental sustainability in China's banking sector: A mixed-methods approach. *British Journal of Management, 36*(3), 1256–1273.
Signori, A., & Vismara, S. (2018). Does success bring success? The post-offering lives of equity-crowdfunded firms. *Journal of Corporate Finance, 50*, 575–591.
Slovic, P. (1972). Psychological study of human judgment: Implications for investment decision making. *Journal of Finance, 27*(4), 779–799.
Sørensen, M. (2007). How smart is smart money? A two-sided matching model of venture capital. *Journal of Finance, 62*(6), 2725–2762.
Sorenson, O., & Stuart, T. E. (2001). Syndication networks and the spatial distribution of venture capital investments. *American Journal of Sociology, 106*(6), 1546–1588.
Teten, D., & Farmer, C. (2010). Where are the deals? Private equity and venture capital funds' best practices in sourcing new investments. *Journal of Private Equity, 14*(1), 32–52.
U.S. Bureau of Labor Statistics. (2023). *Quarterly business starts United States* (p. 2023). Statista. https://www.statista.com/statistics/277505/venture-capital-number-of-deals-in-the-united-states-since-1995/.
Venugopal, B., & Yerramilli, V. (2022). Seed-stage success and growth of angel co-investment networks. *Review of Corporate Finance Studies, 11*(1), 169–210.
Yao, S., Zhao, J., Yu, D., Du, N., Shafran, I., Narasimhan, K., & Cao, Y. (2022). React: Synergizing reasoning and acting in language models. *ArXiv preprint ArXiv: 2210.03629*, 1–23.