sequel

The Founder Files

Largest quantitative analysis of pitch decks

17,500 decks across 121 countries

Abstract

Pitch decks are central to early-stage fundraising, yet there has been limited large-scale, systematic analysis of their characteristics and outcomes. This study presents a quantitative investigation of 17,546 pitch decks submitted via Pitchleague.ai between April 2023 and June 2025, covering startups from 121 countries. The dataset is predominantly early-stage, with 96% of ventures at the Pre-Seed or Seed stage. The purpose of this paper is to identify patterns that differentiate funded from non-funded startups, examine common deck features, and assess the extent to which standardized scoring can capture investor-relevant signals. The analysis employs a combination of natural language processing (NLP) for information extraction, data cleaning to address inconsistencies and missing values, and enrichment with verified funding records from Harmonic. Two benchmarks guide the study: Pitchleague's proprietary scoring system, reflecting startup and deck quality, and externally validated fundraising outcomes. The study demonstrates that systematic analysis of pitch decks offers valuable insights into early-stage fundraising dynamics and underscores the potential of data-driven benchmarks to complement investor judgment. Importantly, all results in this study are derived exclusively from the PitchLeague.ai dataset; findings should therefore be interpreted as insights into this dataset rather than as a comprehensive representation of the global startup population.

1. Introduction

1.1 Sequel

Sequel (https://www.sequel.co/) is a digital family office built for the world's best athletes, with headquarters in Miami and offices in London. It was founded in 2022 with a clear mission: to help athletes invest in startups that are shaping the future and in doing so, build a lasting legacy beyond sport. The platform runs through an iOS and Android app where athletes can learn the ins and outs of startup investing, with easy-to-follow educational content and masterclasses from some of the top names in startup investing. Sequel also gives its members the chance to invest directly into vetted startups through regular investment "drops." So far, there have been 25 drops, and the platform has grown to a paying community of around 200 athlete-investors.

Sequel's model is built around three forms of modern leverage: code, media, and capital. Its technology enables scale, its content builds trust and education, and its curated deal flow unlocks access to opportunities traditionally out of reach. Together, these elements empower athletes not just to invest but to become influential players in the innovation economy.

1.2 Pitchleague.ai

Sequel built PitchLeague.ai (https://www.pitchleague.ai/) during a three-day hackathon in April 2023 with the mission of supporting early-stage founders in improving their fundraising pitch decks. The platform provides AI-generated, slide-by-slide feedback on both design and content within 60 seconds, helping founders communicate their startup's potential more effectively and increase their chances of securing an initial meeting with venture capital investors. Each deck is scored on a scale from 0 to 100, based on both startup fundamentals such as team, market, and traction as well as deck quality which includes clarity, structure, length and grammar (Figure 1). User engagement is driven by a live public leaderboard, which introduces a gamified, competitive dynamic that motivates founders to iterate on their decks. This approach, reinforced by the platform's visibility on Product Hunt, significantly increased both submission volume and startup diversity, resulting in a rich dataset covering a broad range of sectors and geographies.

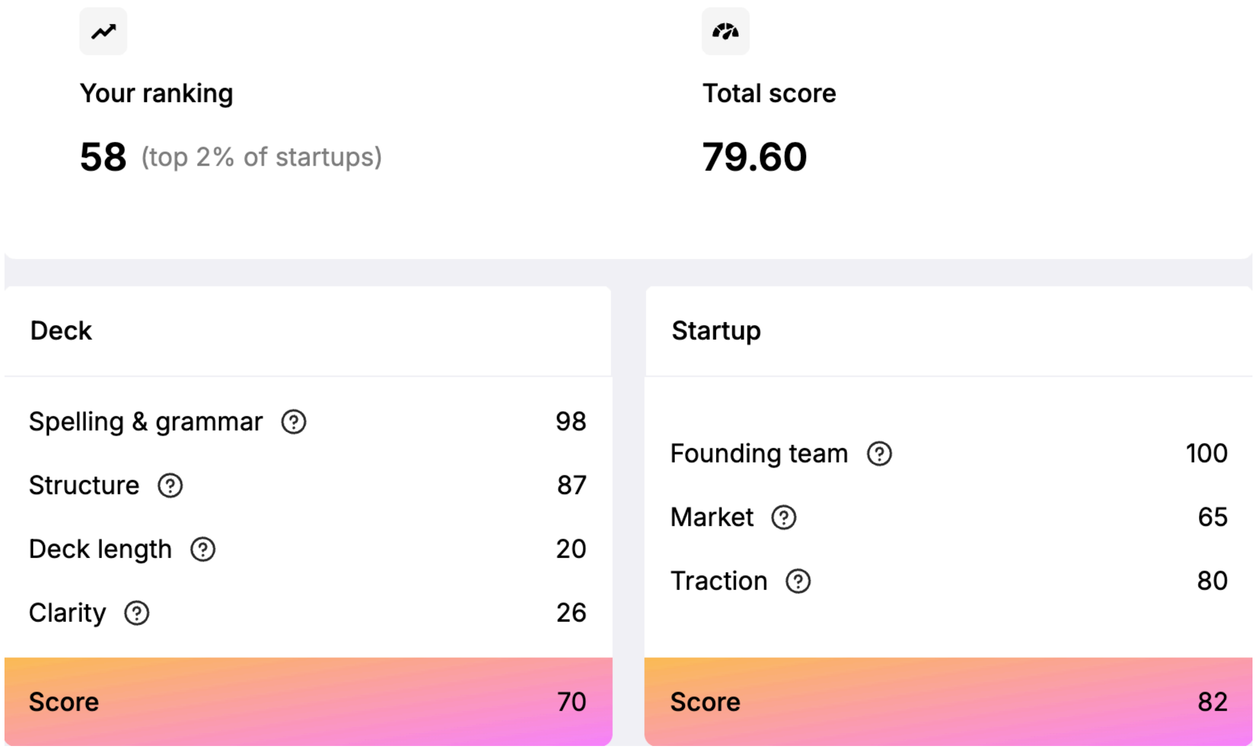

Figure 1: Scoring System (Pitchleague, 2025)

1.3 Motivations for this study

Why Study Pitch Decks?

For decades, pitch decks have been the go-to tool for early-stage startups to share their vision, team credentials, traction, and growth plans into just a few slides. Yet, with investors typically spending between two and five minutes per deck (Armstrong, 2025), every word, number, and design choice matters. Despite this high-stakes reality, there's surprisingly little rigorous research on what truly makes a pitch deck effective. This study addresses that gap by analyzing thousands of real-world decks, moving beyond anecdotal advice to uncover evidence-based patterns linked to fundraising success.

Why Focus on Differences, Not Predictions?

Pitch decks are key for analyzing how startups present themselves to investors, but they are not designed to predict long-term success. Startup outcomes are shaped by far more than what appears in a slide deck, including market shifts, timing, execution, networks and luck. Machine learning models, while powerful, also struggle with the incomplete and uneven nature of deck data and risk embedding historical biases if used for forecasting. Recognizing this, we do not attempt to predict which startups will succeed. Instead, we use pitch-deck data to (i) identify systematic differences between funded and non-funded ventures, (ii) map broader patterns in founder backgrounds, deck design, and traction signals, and (iii) validate whether standardized scoring captures the same signals investors respond to

Challenges in Early-Stage Venture Capital

Previous research on early-stage venture capital highlights three consistent challenges:

1. **Information asymmetry**: Early-stage investors often make funding decisions with minimal hard data, typically relying on a pitch deck and a short founder meeting (Ewens et al., 2022). Many startups are pre-revenue and lack standardized performance metrics, making objective evaluation difficult.

2. **Reliance on heuristics**: In the absence of reliable traction data, investors often lean on proxies such as prior founder experience, elite educational backgrounds, accelerator participation, or endorsements from known investors (Gompers et al., 2016; Islam et al., 2018). While these can signal quality, they may also reinforce structural biases and overlook unconventional founders who could succeed.

3. **Power-law outcomes**: A small proportion of investments drive the majority of VC returns (Buffington, 2025). This increases the pressure to identify outlier

startups early, yet these are often the hardest to predict from limited information.

Recent advances in AI-assisted evaluation attempt to address these gaps by combining diverse data sources such as founder backgrounds, product traction and market trends into structured, scalable assessments (Wang and Ihlamur, 2024). However, prior studies also caution that algorithms trained on historical funding decisions can inherit existing biases unless actively mitigated (Sachs and Unbescheiden, 2024). The consensus emerging in the literature is that the most effective approach is a hybrid "human + AI" model, where algorithmic insights complement, rather than replace, investor judgment.

Against this backdrop, the aim of this study is not to predict individual startup outcomes, but to identify patterns and associations between measurable startup and pitch characteristics and two benchmarks:

- **Internal benchmark:** PitchLeague scores, reflecting startup and deck quality as assessed by a standardized scoring process.

- **External benchmark:** Verified funding outcomes, capturing tangible validation from the investment market.

2. Data & Methodology

2.1 Data Sources

The dataset used in this study consists of 17,546 fundraising pitch decks submitted to PitchLeague.ai between April 2023 and June 2025. This period coincides with the rise of generative AI, heightened scrutiny of investment decisions, and a sharp contraction in global venture funding, which fell to $319 billion in 2023, the lowest level since 2020 (Dealroom, 2023). The scale and diversity of submissions offer a unique window into the early-stage startup ecosystem, spanning multiple industries, geographies, and founder profiles. This temporal and contextual coverage provides a valuable lens for analyzing how market dynamics and technological shifts have shaped early-stage activity, emerging trends, and the pitch-deck characteristics associated with fundraising success.

2.2 Data Extraction

When a pitch deck is uploaded, it goes through an automated extraction pipeline that uses natural language processing (NLP) and AI to pull out both structured and unstructured information. The system first processes the text, checks the language, and then uses advanced NLP models to pick out things like company details, team info, market size, and financials, not just from the text, but from visuals and slide layouts too. This way, we are able to capture everything from straightforward numbers and names to more subtle signals in the slides. All of the extracted data is checked for quality, mapped to a consistent format, and then stored in the database as a JSON file for each individual deck. This makes it easy to track exactly which information comes from which deck, and to keep everything organized for later analysis. By combining both structured and unstructured data extraction, we make sure no important detail gets missed in our analysis.

2.3 Data Cleaning

2.3.1 Removing Failed Uploads

The cleaning process began by filtering out failed uploads, which represented 6.3% of all submissions. These failures were due primarily to unsupported languages (55%), decks exceeding the platform's page or token capacity limits (36%), and API processing errors (9%). Since PitchLeague supports only English-language decks and has practical limits on length and word count, removing these records was essential to ensure the dataset contained only valid, analyzable entries.

2.3.2 Identifying and Removing Duplicates

After this initial step, we addressed duplicates. Many founders resubmitted their decks after making revisions in response to PitchLeague's feedback, which often resulted in multiple near-identical versions of the same pitch. To identify duplicates, we used three key fields present in almost all submissions, file name, company name, and founder name(s) as matching identifiers. These fields were standardized by converting all text to lowercase, removing punctuation and extra spaces, and harmonizing spelling variations to ensure that small formatting differences did not prevent detection. When one of the fields was missing, the remaining fields served as fallbacks, allowing for accurate and consistent duplicate removal.

After removing duplicates, the dataset was reduced to 6,181 unique decks, representing 35.2% of the original uploads. By applying matching rules and using multiple identifiers, we ensured genuine duplicates were removed without merging distinct companies, resulting in a clean, representative dataset ready for analysis (Figure 2).

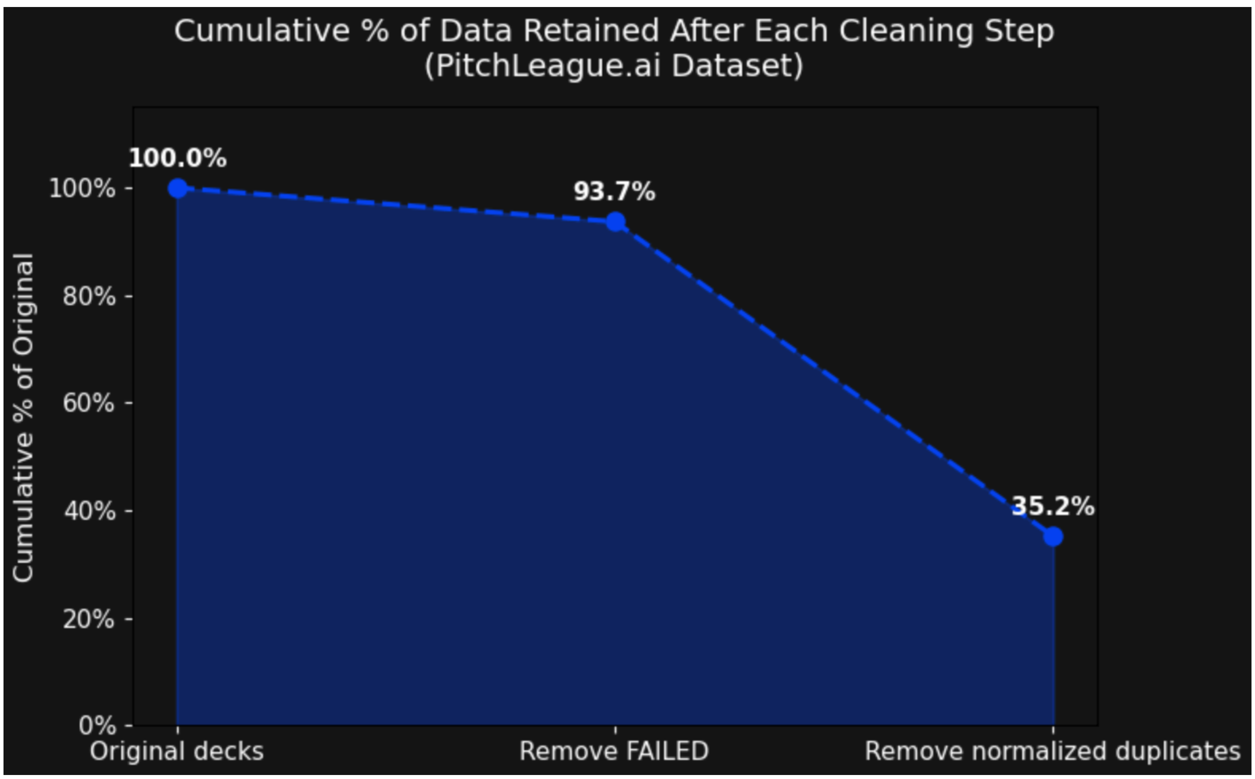

Figure 2: Data Retained After Each Cleaning Step

2.4 External Data Enrichment

Pitch decks offer a snapshot of a startup's story at a single moment in time, often omitting critical follow-up information such as subsequent fundraising. To address this gap, the dataset was enriched using the Harmonic platform, which provides longitudinal data on company funding history and stage progression.

The enrichment process matched startups from PitchLeague.ai to Harmonic's database using unique identifiers such as company websites and official names (Figure 3). While the process faced challenges, particularly with very early-stage startups lacking a web presence or public funding records, it successfully linked a significant subset of companies. This revealed that approximately 14% of startups in the dataset had raised funding.

This enriched funding status was then used as a binary benchmark throughout the analysis, enabling direct comparisons between funded and non-funded startups in terms of their characteristics, scoring profiles, and founder attributes.

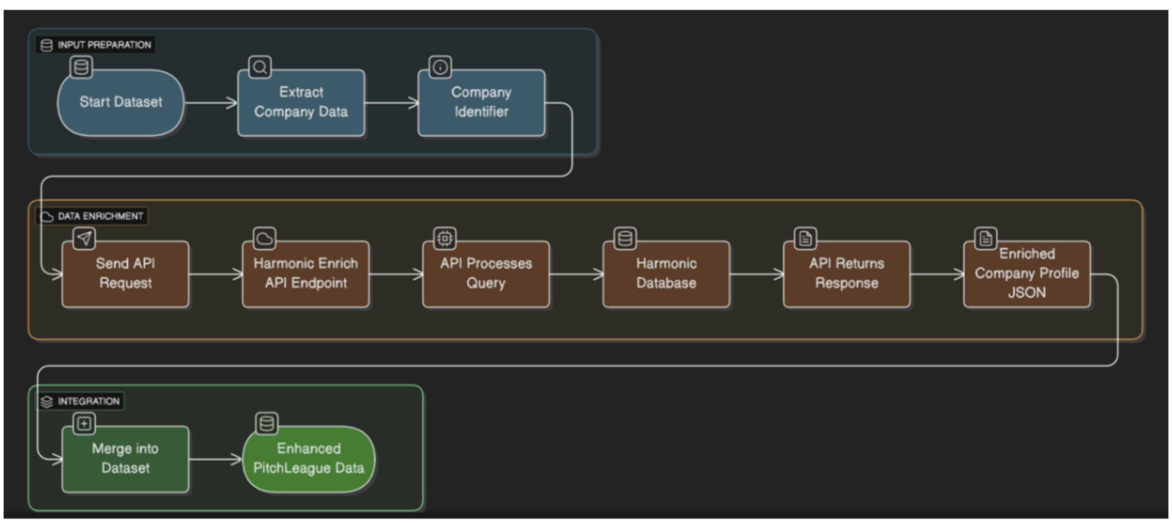

Figure 3: Data enrichment Harmonic API

2.5 Data Validation and Quality Assurance

For the final round of data validation, we wanted to be sure that our funding round information was as accurate as possible. To do this, we compared the funding round data extracted from Harmonic with the official records listed on Crunchbase and PitchBook. We checked that the round sizes, company stages, founder names, and timelines all lined up between our dataset and these trusted external sources. If we noticed any discrepancies, such as a funding round in our data that didn't show up in Crunchbase, or mismatched stages or dates, we dug deeper to figure out the source of the inconsistency and made corrections where needed. This hands-on approach ensured that, when we merged the funding round information into our final dataset, it closely reflected what actually happened in the real world. By doing these extra checks, we increased our confidence in the reliability of our data and the validity of any conclusions drawn from it.

2.6 Handling Missing Values in Early-Stage Startup Data

2.6.1 Continuous Numerical Variables: Traction and Market Metrics

One of the most consistent challenges in analyzing early-stage venture capital datasets is the high rate of missing values, particularly in core traction and financial metrics. This issue arises partly because data collection on PitchLeague.ai is voluntary, founders choose what to disclose and partly because many very early-stage companies are simply pre-revenue. Indeed, metrics like Monthly Recurring Revenue (MRR), Annual Recurring Revenue (ARR), Gross Margin, or Burn Rate require established tracking processes and financial maturity that many young startups have not yet reached.

Figure 4 highlights the scale of this problem. In which roughly 95% of startups in the dataset do not report MRR, and over 90% omit Gross Margin. Furthermore, burn rate measures (quarterly, yearly, monthly) are missing in nearly all cases.

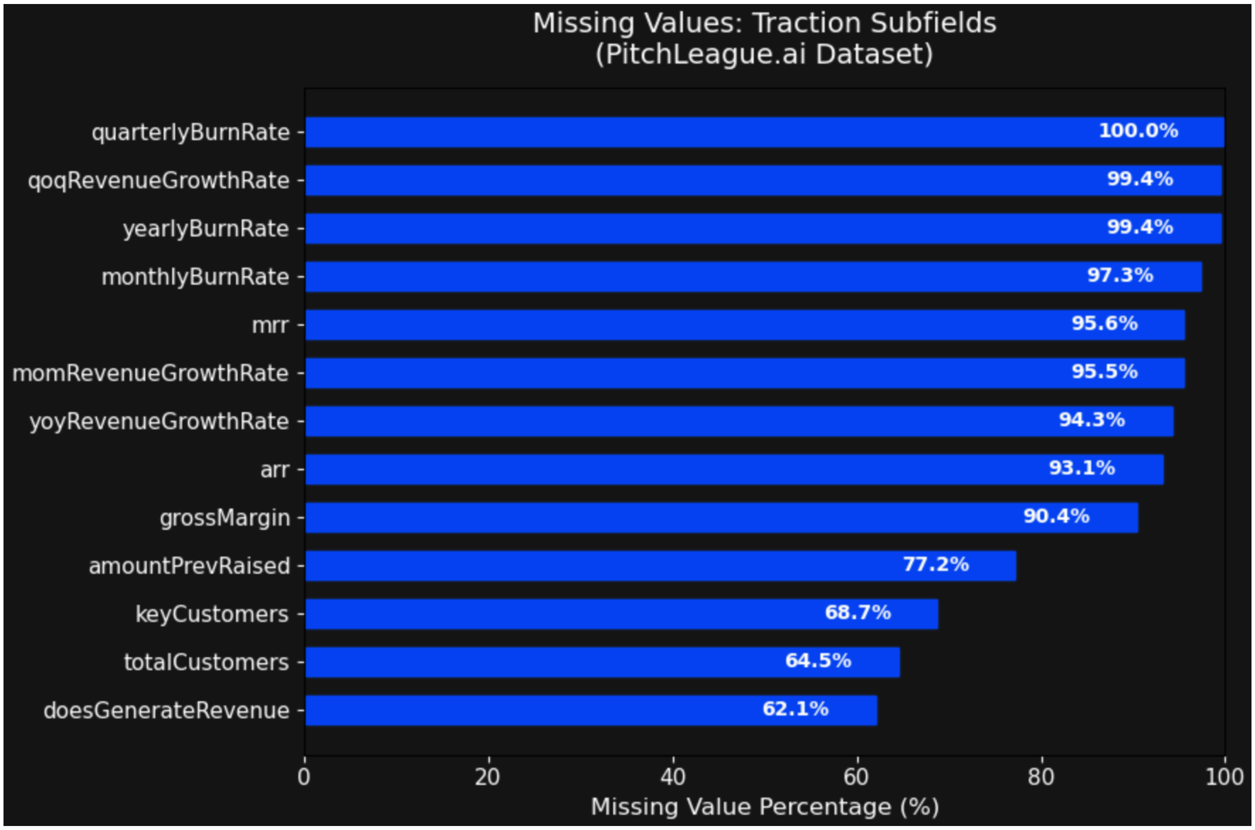

Figure 4: Missing Values Traction

To address the challenge of missing values, the analysis used pairwise deletion, meaning that each comparison includes all startups that have data for the variables being analyzed, rather than discarding an entire record if other fields are missing. For example, when studying the relationship between MRR and fundraising success, all startups with both MRR values and fundraising outcome data are included, even if they lack information on other metrics such as TAM or gross margin.

This approach preserves far more data while keeping comparisons valid. However, overlaps between certain variables remain very limited. As shown in Figure 5, while market size measures such as TAM have relatively high coverage (~85%), only 3.8% of startups report both TAM and MRR. Such small overlaps mean that any analysis combining these variables is based on a very small subset of the data, so findings should be interpreted with caution.

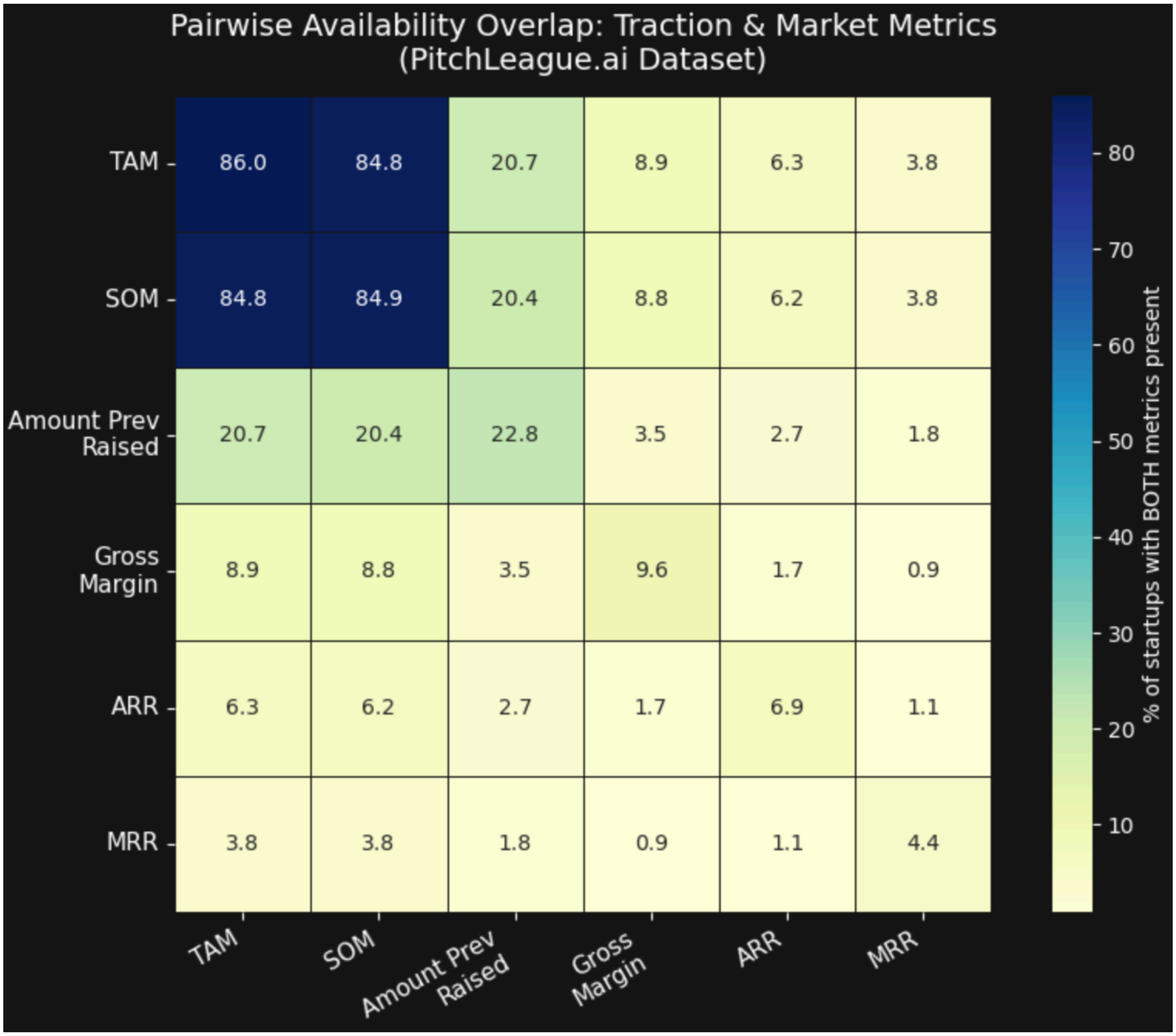

Figure 5: Pairwise Availability Overlap

2.6.2 Boolean Attributes: Simple, Consistent, and Complete

While traction metrics suffer from extensive missingness, certain fields are far easier to work with. Boolean variables, simple "Yes" or "No" responses, are consistently available across almost the entire dataset (N = 6,181). These include whether a startup's team has a technical background, complementary skills, previous founder experience, previous startup experience, or a previous exit.

Figure 6 shows that around 78% of teams report complementary skills, while roughly 70% have a technical background. In contrast, only about 12% have a previous exit, and just over a third have previous startup or founder experience. The key advantage of these variables is that they are immune to partial reporting issues, a trait either applies to the team or it does not. This makes them highly reliable for inclusion in comparative and predictive analyses without the risk of significant data loss.

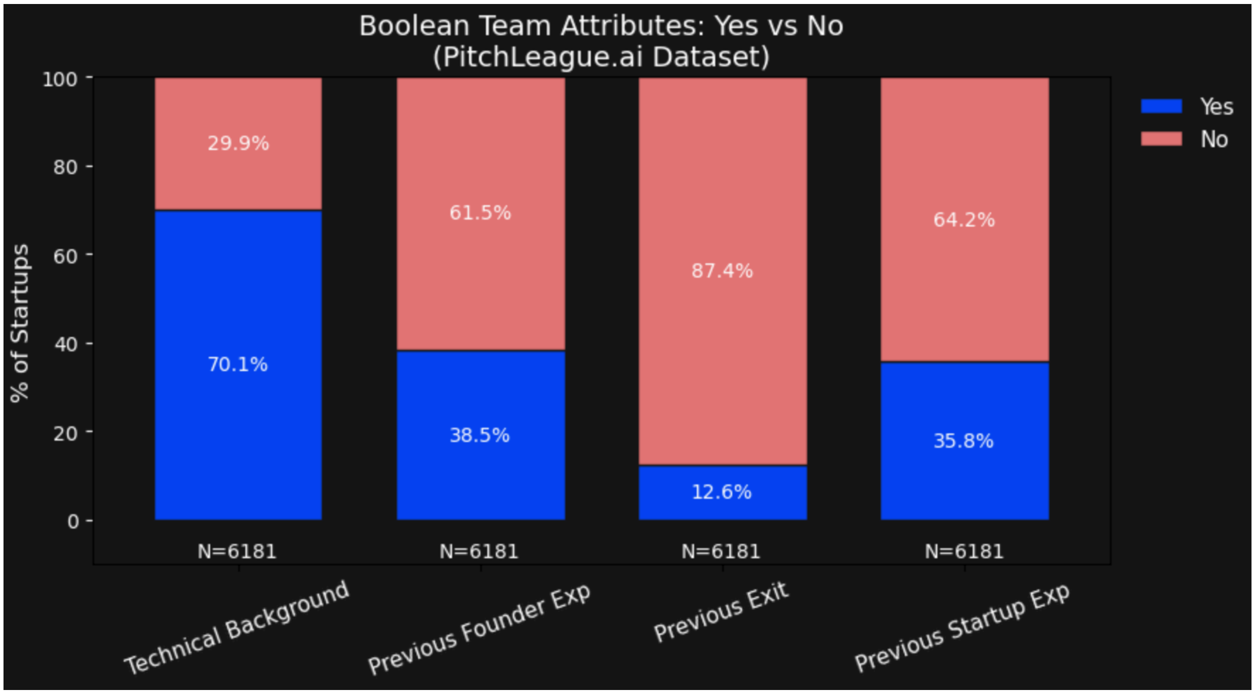

Figure 6: Boolean Team Attributes

2.6.3 Scoring Metrics: 100% Coverage

Some variables in the dataset have near-perfect or complete coverage, meaning they are available for every single startup. Alongside funding status (funded vs. not funded), three composite scoring metrics stand out:

- **overall**: the aggregate score combining all evaluation dimensions
- **startup.overall**: a measure of the startup's underlying business fundamentals
- **deck.overall**: a measure of the pitch deck's presentation quality

This level of completeness is rare in early-stage startup data, where many other variables suffer from partial reporting (Figure 7). Because these scores are standardized and consistently available, they form a reliable foundation for:

- **Benchmarking:** comparing startups against each other on the same scale
- **Correlational analysis:** testing whether higher scores align with funding success
- **Segment comparisons:** analyzing differences by stage, geography, or sector without the complications of missing data

In short, these fully populated variables act as a common reference frame for the analysis, enabling direct comparisons across the entire dataset.

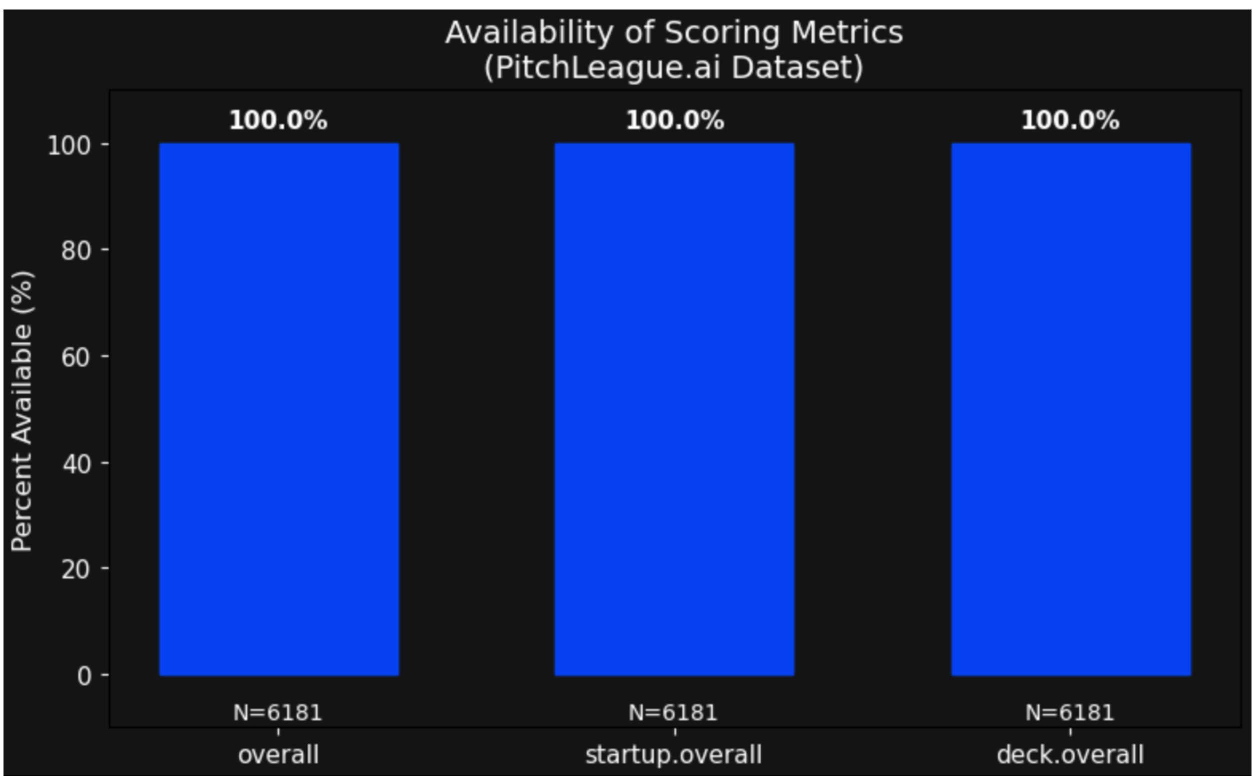

Figure 7: Availability of Scoring Metrics

2.6.4 Textual Continuous Variables: Cleaning for Accuracy

Furthermore, continuous textual attributes such as past companies and past universities, is generally well populated but faces a different challenge: inconsistency. Without cleaning, entries like “MIT,” “Massachusetts Institute of Technology,” and “M.I.T.” would be counted as three separate institutions. The same issue occurs with companies such as “Google,” “Google Inc.,” and “Google LLC” could all be present in the data.

To ensure accuracy and avoid undercounting, these variables underwent a thorough standardization process. All entries were converted to lowercase, punctuation and extra spaces were removed, and fuzzy matching was used to merge near-identical names. This step was critical for producing reliable counts and for correctly linking founder background data to fundraising performance. Without this cleaning process, the analysis could have missed important patterns in alumni networks or prior work experience, which are often key factors in investor decisions.

3. Analytical Framework for Comparative Analysis

3.1 The role of benchmarking

We evaluated startups using two benchmarks: (1) the PitchLeague scoring system, a multi-dimensional measure of startup and deck quality and (2) verified funding outcomes, representing real-world market validation. The scoring benchmark enabled consistent internal comparisons, helping to identify high performers and explore how traits like founder experience or deck design related to quality scores. The funding benchmark provided an external check, revealing whether characteristics associated with high scores also aligned with actual fundraising success.

3.2 Quarter-over-Quarter Analysis Method

To track changes in startup and pitch deck characteristics over time, each record in the dataset was assigned to a calendar quarter based on its timestamp. Where available, we used the specific *deck date* provided within the pitch deck metadata. If no deck date was available, we defaulted to the date the deck was submitted to PitchLeague. This ensured every record had a consistent time reference. Once timestamps were assigned, they were grouped into quarterly periods (e.g., Q2 2023, Q3 2023), enabling us to plot trends such as AI adoption, funding amounts, and key buzzword usage across time. This method allowed us to capture longitudinal patterns even when exact deck creation dates were missing, while maintaining temporal consistency across the dataset.

3.3 Use of Median for Skewed Distributions

In early-stage startup data, not only do we need to interpret core metrics like Total Addressable Market (TAM) and Annual Recurring Revenue (ARR) carefully, but we also need to apply the same caution to related measures such as the PitchLeague scoring system and reported funding amounts. These datasets are often highly skewed: a small number of startups report exceptionally large market sizes, revenues, or funding rounds, sometimes due to ambitious projections, aggressive rounding, or the presence of later-stage companies.

If we relied on the mean (arithmetic average), these extreme values would distort the picture, making the “typical” startup appear far larger, more mature, or higher-scoring than is realistic. This is why we use the median, the midpoint value when all observations are sorted, for all skew-sensitive variables, including market metrics, funding amounts, and even composite scores.

By focusing on medians, we avoid having our analysis pulled upward by a handful of outliers and instead provide a more representative benchmark for the dataset as a whole. For example, while the mean TAM might suggest a multi-billion-dollar opportunity, the median often reflects a more grounded and attainable market size. Likewise, median funding amounts give a truer sense of what most startups raise, rather than being inflated by rare mega-rounds. This approach ensures that comparisons, whether by stage, region, sector, or score percentile, are both statistically robust and practically meaningful.

4. Findings and Insights

4.1 Global Distribution and Trends in Startup Funding

4.1.1 Stage Maturity

The majority of startups in the PitchLeague dataset are concentrated at the Seed (62.1%) and Pre-Seed (33.7%) stages, with only a small minority (4.2%) reaching Series A. This distribution is ideal, as it closely matches PitchLeague's original mission: to serve as a feedback and benchmarking tool specifically for founders at the very earliest stages of their fundraising journey.

The platform's emphasis on Pre-Seed and Seed rounds ensures its feedback, scoring, and analytics align with the realities of founders who may not yet have access to institutional capital or formal investor networks. The strong representation of these ventures in the dataset shows that PitchLeague is successfully engaging its target audience at a critical stage (Figure 8).

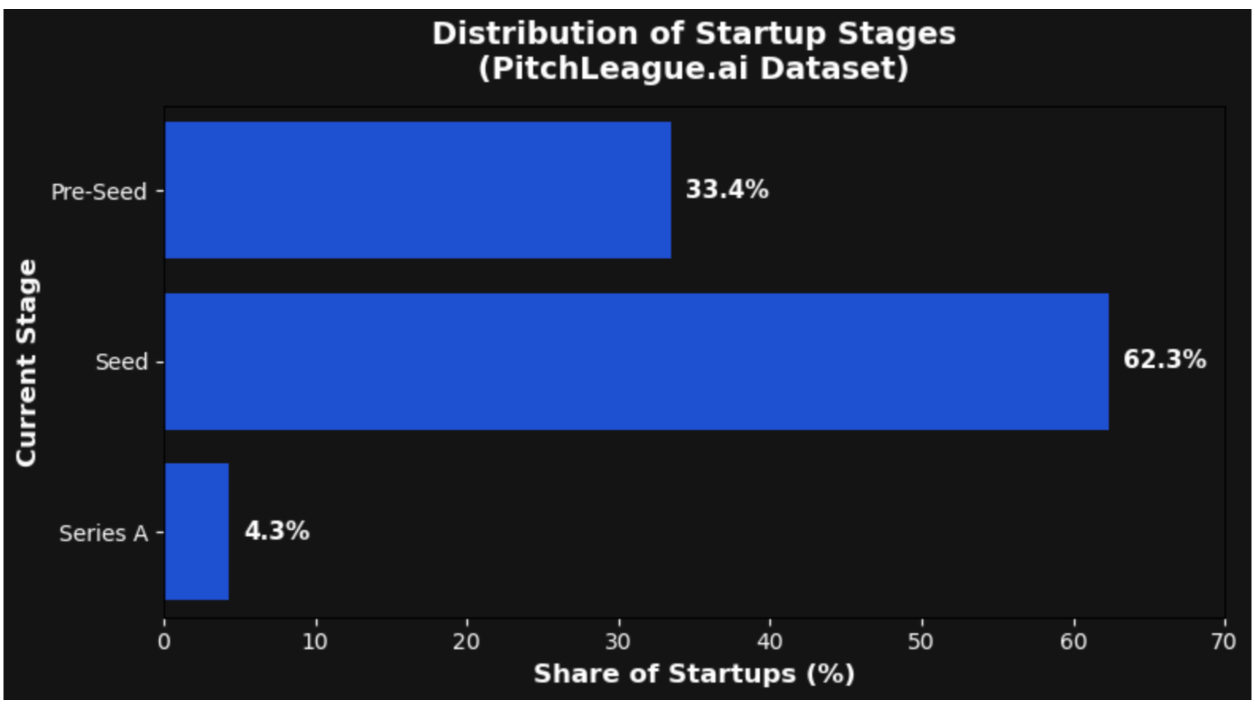

Figure 8: Distribution of Startup Stages

4.1.2 Startup Metrics by Stage

It is important to note that these benchmarks are based only on the subset of pitch decks that explicitly reported financials and market size figures. Within that group, amount raising, revenue, and market size ambitions all increase sharply as startups move through the fundraising stages, while margins show more fluctuation (Figure 9). The median raise ask doubles from $0.5M at Pre-Seed to $1.0M at Seed, before rising fivefold to $5.0M at Series A. Furthermore, revenue acceleration is even more pronounced: ARR grows 3.4× from Pre-Seed to Seed, and 5.2× from Seed to Series A, reflecting how traction becomes the central requirement for Series A readiness.

Gross margins follow a different pattern. They fell by around 7.5 percentage points between Pre-Seed and Seed (from 67.5% to 60.0%), before recovering modestly to 62.5% at Series A. At the same time, market size claims expanded substantially. TAM more than doubles (from $49B to $116B) and SOM grows even faster (from $0.66B to $2.50B, a 2.6× increase). This suggests that as startups mature, founders refine their narratives to focus on realistic market capture rather than broad opportunity sizing.

Taken together, these patterns show how early-stage ventures evolve from small, uncertain bets into businesses that demonstrate scale, traction, and credible market pathways.

Metric	Pre-Seed	Seed	Series A	Growth Multipliers
Median Raise	$0.50M	$1.00M	$5.00M	×2.0 → ×5.0
Median ARR	$84K (N=43)	$286K (N=156)	$1.50M (N=21)	×3.4 → ×5.2
Gross Margin	67.5% (N=62)	60.0% (N=223)	62.5% (N=26)	-7.5pp → +2.5pp
Median TAM	$49.1B (N=330)	$64.7B (N=302)	$116B (N=44)	×1.3 → ×1.8
Median SOM	$0.66B (N=325)	$0.95B (N=298)	$2.50B (N=44)	×1.4 → ×2.6

Figure 9: Startup Metrics by Stage

4.1.3 Fundraising Trends

To analyze year-on-year changes in startup funding, we used each company's funding round (as provided by Harmonic) and aggregated results for 2023 - 2025. Funding values, reported directly in USD, were included without currency conversion, and only startups with non-null funding data were considered. Two complementary measures were computed:

- Median round size, representing the central tendency of deal sizes.
- Total annual funding, summing all reported investments to capture the overall scale of capital deployed.

Within the PitchLeague.ai dataset, total global funding rose from $0.9B in 2023 to $1.9B in 2025, with the number of funded startups also increasing steadily (N=254 to N=328) (Figure 10). Median round size held steady between 2023 ($0.73M) and 2024 ($0.70M), before jumping to $1.10M in 2025 (Figure 11). These trends indicate that 2025 was not only a year of greater aggregate capital but also larger individual rounds, suggesting an expansion in both deal flow and deal size among early-stage startups captured in the dataset.

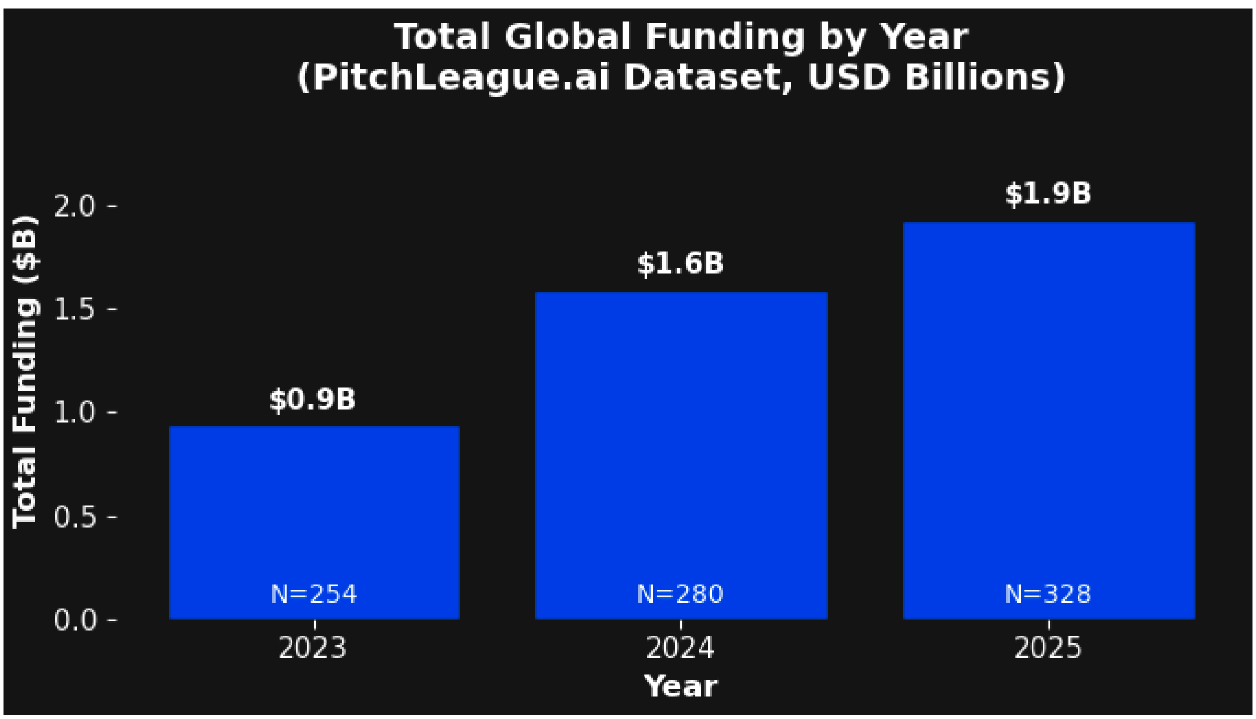

Figure 10: Total Startup Funding by Year

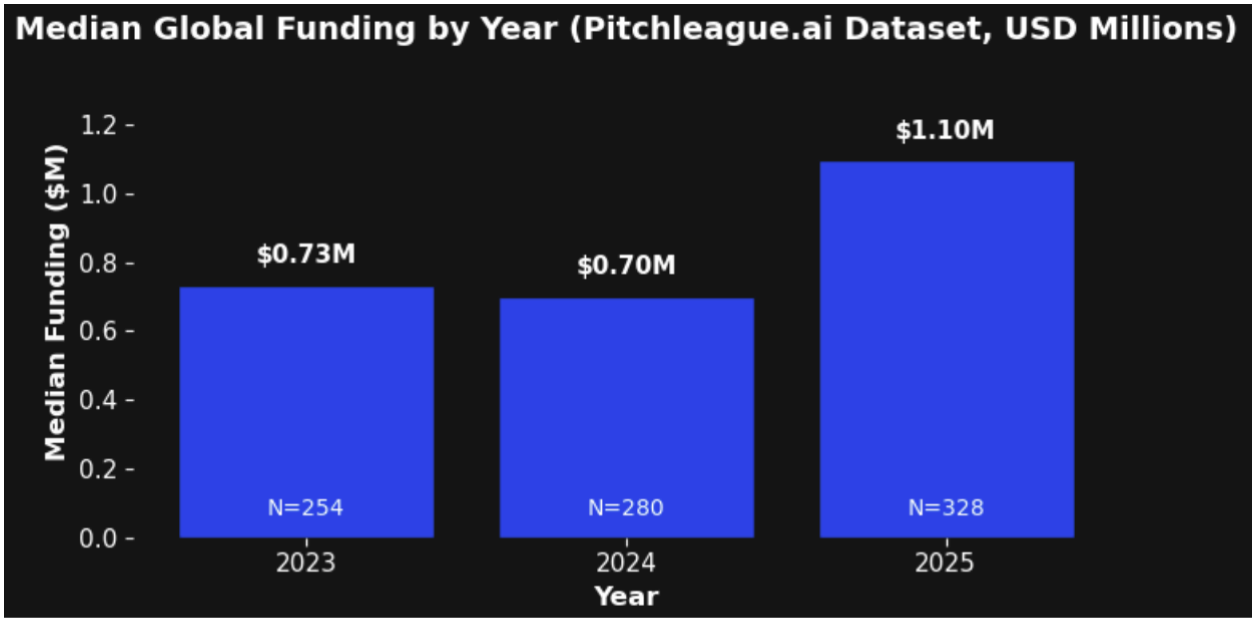

Figure 11: Median Global Funding by Year

4.1.4 Regional Concentration of Funding

While medians are used throughout this study to represent typical startup characteristics (e.g., market sizes, revenues, funding round sizes), Figures 12 present total funding raised. Here the goal is not to show the "average" startup but rather the aggregate concentration of capital across regions and sectors. Totals provide the most meaningful lens for highlighting geographic and sectoral dominance in global venture flows.

Funding is heavily concentrated in the United States ($1,592M) and the United Kingdom ($1,037M), which together account for the majority of early-stage capital. India ($460M) follows as a strong third, reflecting its growing startup ecosystem.

Beyond these three leaders, funding levels drop sharply. Spain ($126M), Brazil ($84M), Israel ($73M), and Nigeria ($70M) represent mid-tier hubs with notable but more contained activity.

Rounding out the top ten are the Netherlands ($62M), Germany ($57M), and Canada ($53M), which highlight the long tail of regions attracting smaller, though still meaningful, venture flows.

Overall, the distribution shows the dominance of U.S. and U.K. markets, the emergence of India, and a diverse set of secondary ecosystems competing for early-stage capital.

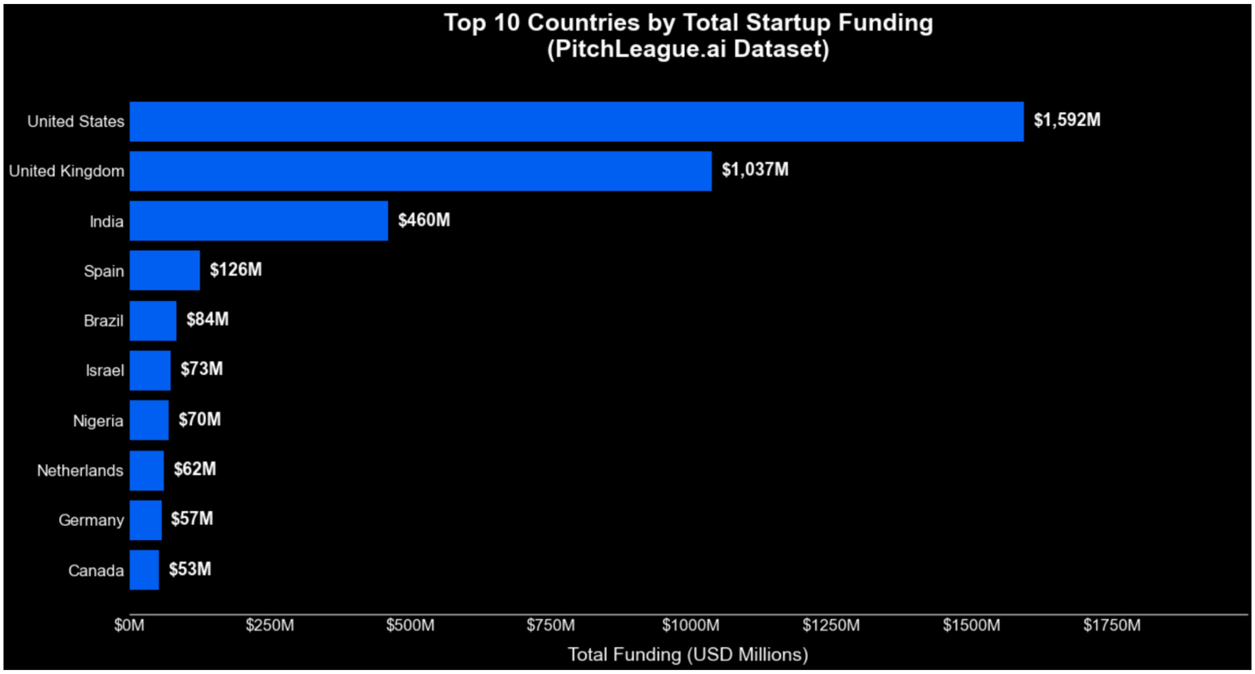

Figure 12: Regional Concentration

4.1.5 Sector Concentration of Funding

Looking at sector-level funding, Fintech leads decisively with $829M raised, followed by Enterprise Software ($717M) and Healthcare ($634M). Together, these three sectors capture nearly half of all capital in the dataset, underscoring investor preference for scalable, tech-driven markets with global reach. A second tier is formed by Consumer Goods ($256M) and Sportstech ($210M), which attracted meaningful interest but at a lower scale. Beyond these, Gaming ($97M), Education ($71M), and more niche categories such as Pet Care, Music Technology, and E-commerce drew smaller but still significant amounts. The distribution reflects a familiar venture pattern: heavy concentration in a handful of dominant industries, balanced by a longer tail of emerging and specialized verticals.

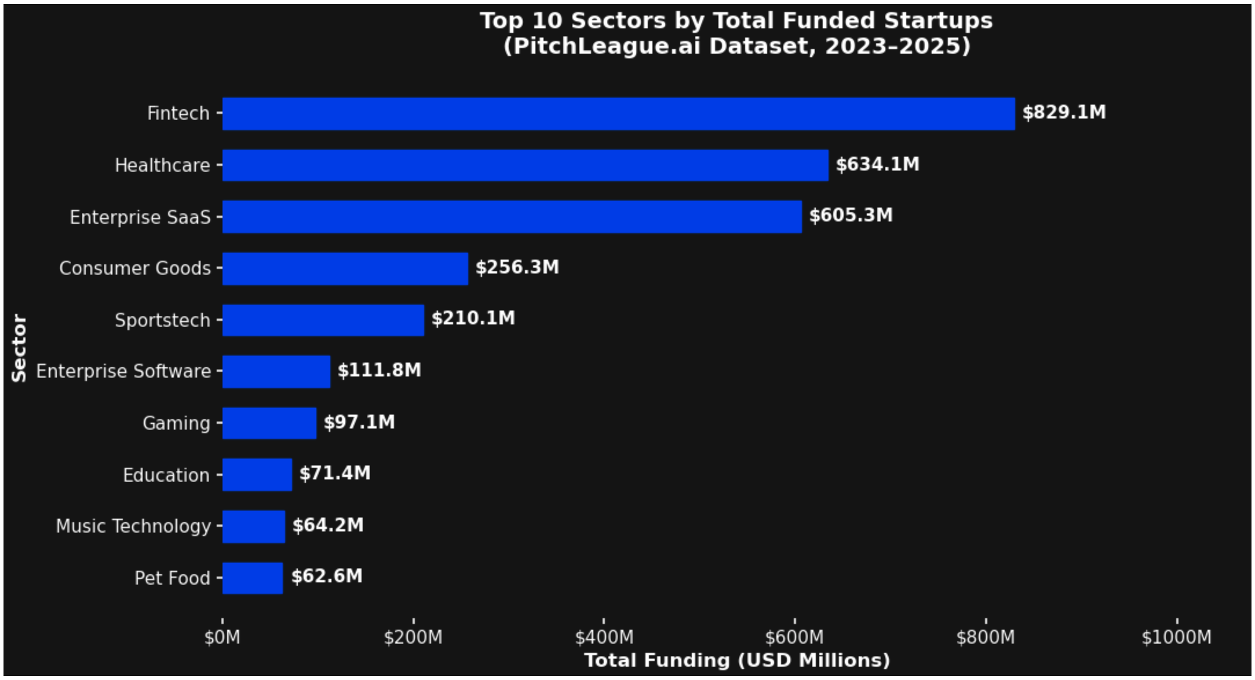

Figure 13: Sectorial Concentration

4.2 Team Characteristics and Funding Outcomes

Evaluating startups at the earliest stages is uniquely challenging due to the absence of standardized operating metrics. Many ventures are pre-revenue and have yet to demonstrate product–market fit, leaving investors with limited traction data on which to base decisions. As a result, subjective factors often carry disproportionate weight in the evaluation process. Among these, the founding team is consistently cited as the single most critical determinant of early-stage investment outcomes (Gompers et al., 2016).

We therefore had a closer look at team characteristics within the PitchLeague dataset, comparing funded and non-funded startups to identify which founder attributes and experiences appear most closely associated with fundraising success. This analysis compares funded startups (N = 864) with non-funded startups (N = 5,317) in the cleaned dataset. We examine five key dimensions of founder and team characteristics within the PitchLeague dataset: prior experience, team size, gender, previous employers, and institutional backgrounds.

4.2.1 Prior Entrepreneurial and Technical Experience

The comparison of founder experience (Figure 14) shows that prior entrepreneurial track record is more prevalent among funded startups. Nearly half (48.5%) of funded ventures included at least one founder with previous founding experience, versus 36.8% of non-funded startups. Prior exit experience, while rarer overall, follows the same pattern: 19.2% of funded startups had at least one founder with a past exit, compared to only 11.5% of those without funding. These gaps suggest that investors are more likely to back teams with demonstrated entrepreneurial histories, particularly successful exits.

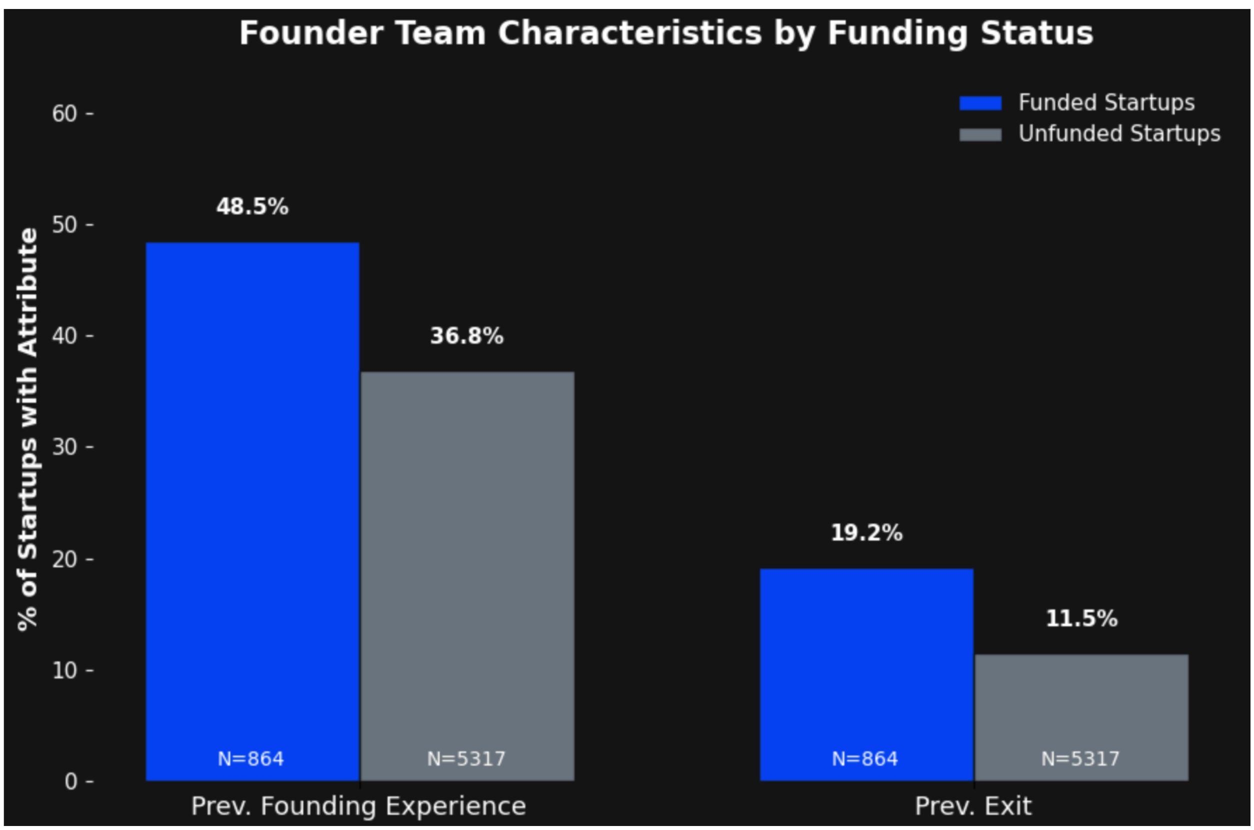

Figure 14: Founder Experience Funded vs Not Funded

Looking at broader team characteristics (Figure 15), technical expertise and prior startup exposure also show strong associations with funding outcomes. A striking 85.8% of funded startups had at least one founder with a technical background, compared to 67.6% of non-funded ventures. Similarly, 49.3% of funded startups reported previous startup experience among founders, compared to 33.6% of non-funded teams. These results highlight that both domain expertise and entrepreneurial familiarity may serve as credibility signals to investors, increasing the likelihood of securing capital.

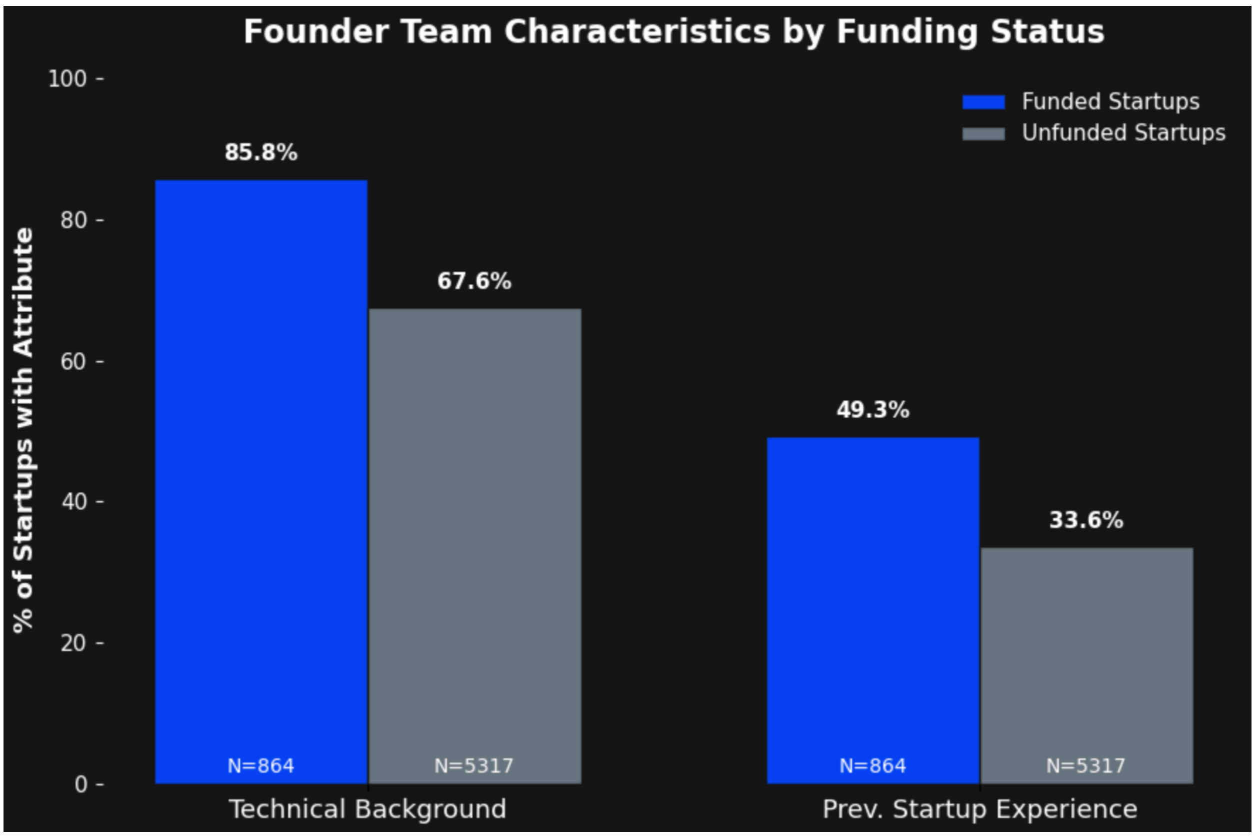

Figure 15: Founder Experience Funded vs Not Funded

4.2.2 Median Funding by Founder Experience

The results, shown in Figure 16, indicate a clear funding premium associated with prior experience. Startups with a founder who had a previous exit raised a median of \$1.13M, compared to \$0.80M for those without, an uplift of 41%. Similarly, founders with previous startup experience secured a median of $1.02M, versus $0.76M for those without, a 34% increase.

While these results do not establish causation, they suggest that investors may associate a founder's past track record, particularly prior exits, with reduced execution risk, leading to larger funding commitments.

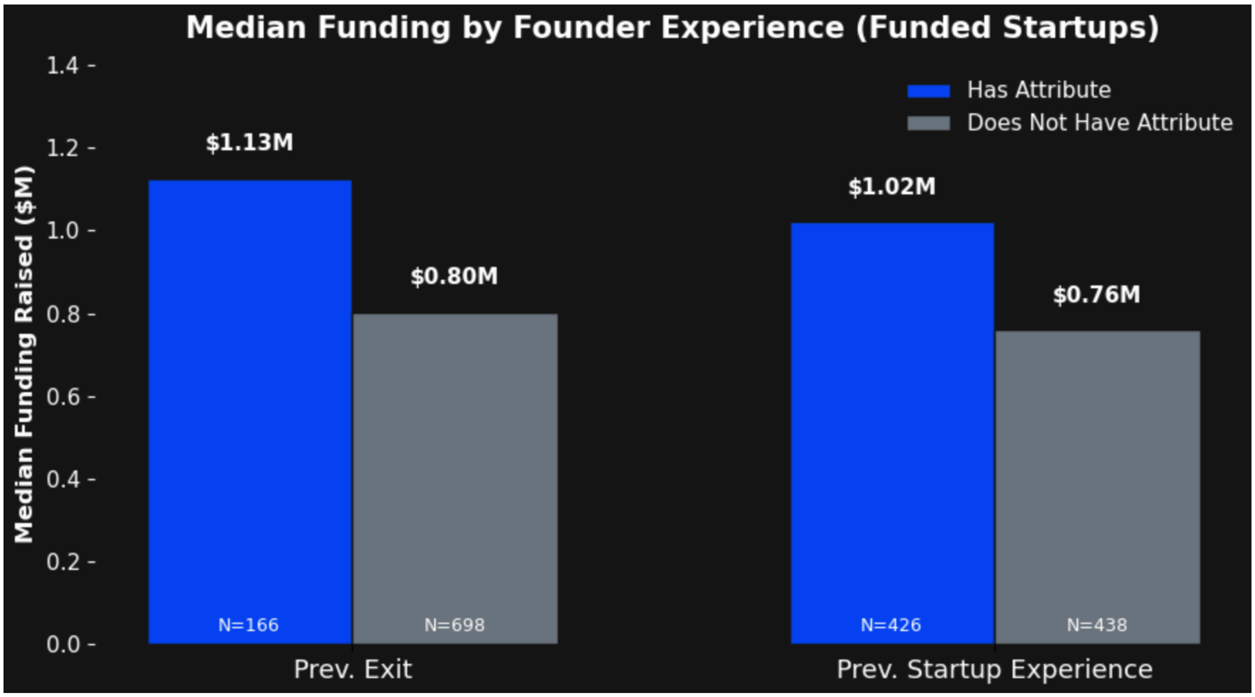

Figure 16: Median Funding by Founder Experience

4.2.3 Founder Team Size

4.2.3.1 Team Size Distribution

An analysis of team size reveals that larger founding teams are more common among funded startups. As shown in Figure 17, co-founder teams (44.9%) and three-founder teams (38.0%) make up a greater share of funded startups compared to their non-funded counterparts (+3.0% and +5.6%, respectively). In contrast, solo founders are less prevalent among funded startups (17.1% vs. 25.7%, -8.6%) (Figure 18). These differences suggest a positive association between team size and fundraising outcomes, with investors appearing more likely to back startups led by two or more founders.

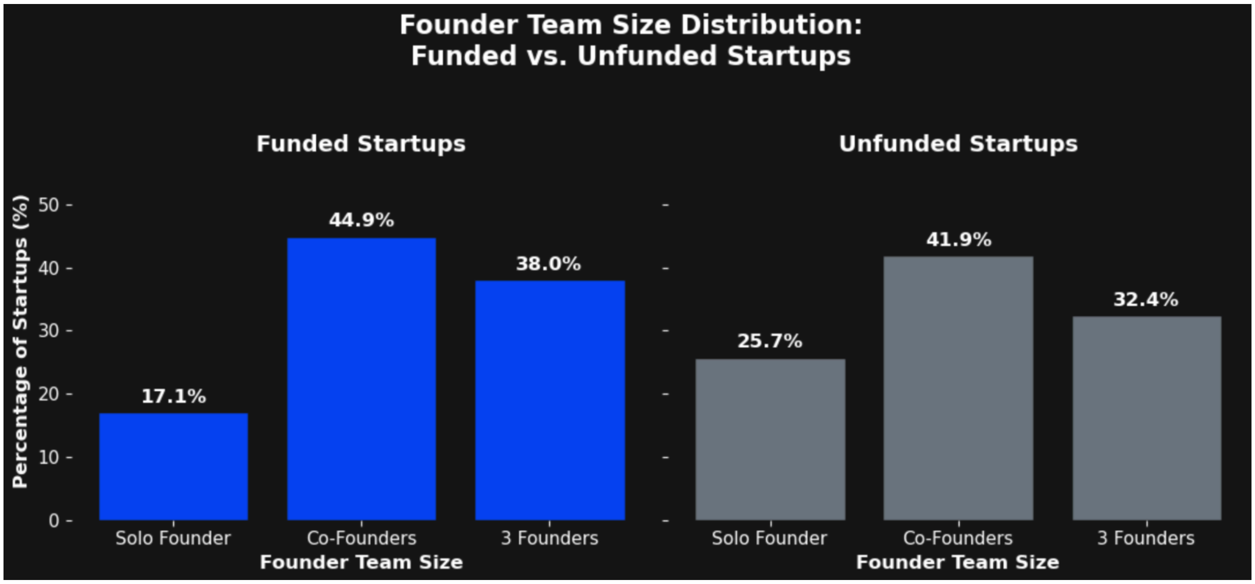

Figure 17: Founding Team size distribution

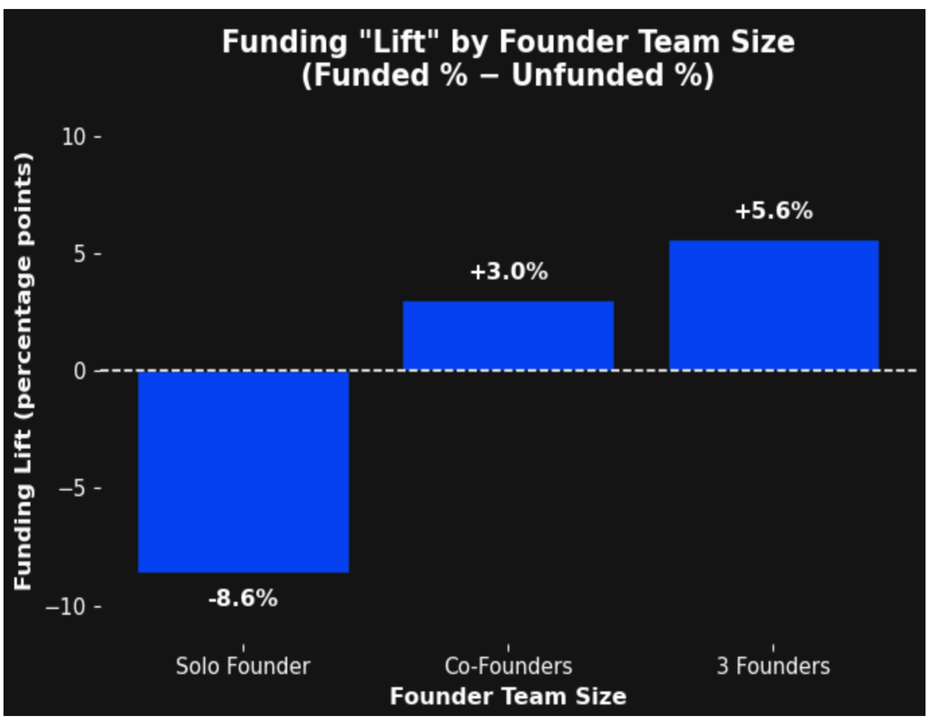

Figure 18: Funding Lift

4.2.3.2 Funding Outcomes by Team Size

Among funded startups, differences also emerge in the size of rounds secured (Figure 19). Solo founders, despite being less common overall, raised the highest median funding at $1.04M, followed closely by three-founder teams at $1.00M. By contrast, co-founder teams raised a lower median of $0.74M. These results suggest a nuanced picture: while larger teams are more likely to secure funding in the first place, solo founders who do succeed often raise disproportionately larger rounds. This duality underscores the complexity of investor perceptions, balancing the advantages of collaborative teams with the perceived decisiveness and focus of single-founder ventures.

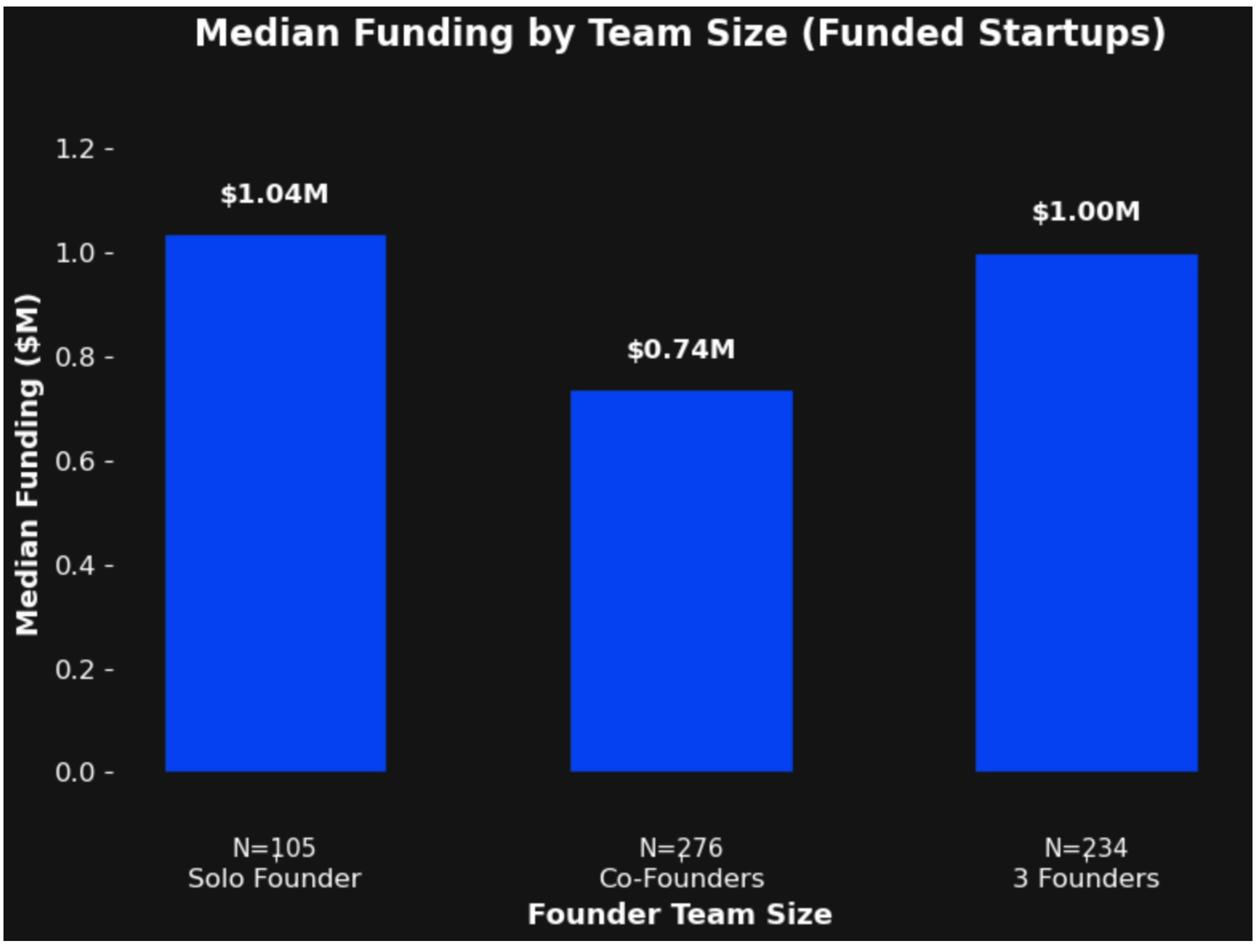

Figure 19: Median Funding by Team Size

4.2.3.3 Headcount Growth Across Funding Stages

Looking at broader team dynamics over the funding journey, a complementary picture emerges when examining overall startup headcount growth by stage (Figure 20). Median headcount rises from 6 employees at Pre-Seed to 8 at Seed, an increase of +2 people (+33%), reflecting modest growth during product validation. The sharpest expansion occurs between Seed and Series A, where median headcount more than doubles from 8 to 17 employees (+9 people, +112%). Overall, Series A startups employ nearly three times as many people as their Pre-Seed counterparts (+11 people, +183%). This trajectory underscores how investors expect startups to remain lean in early stages but demonstrate clear scaling capacity once institutional funding is secured.

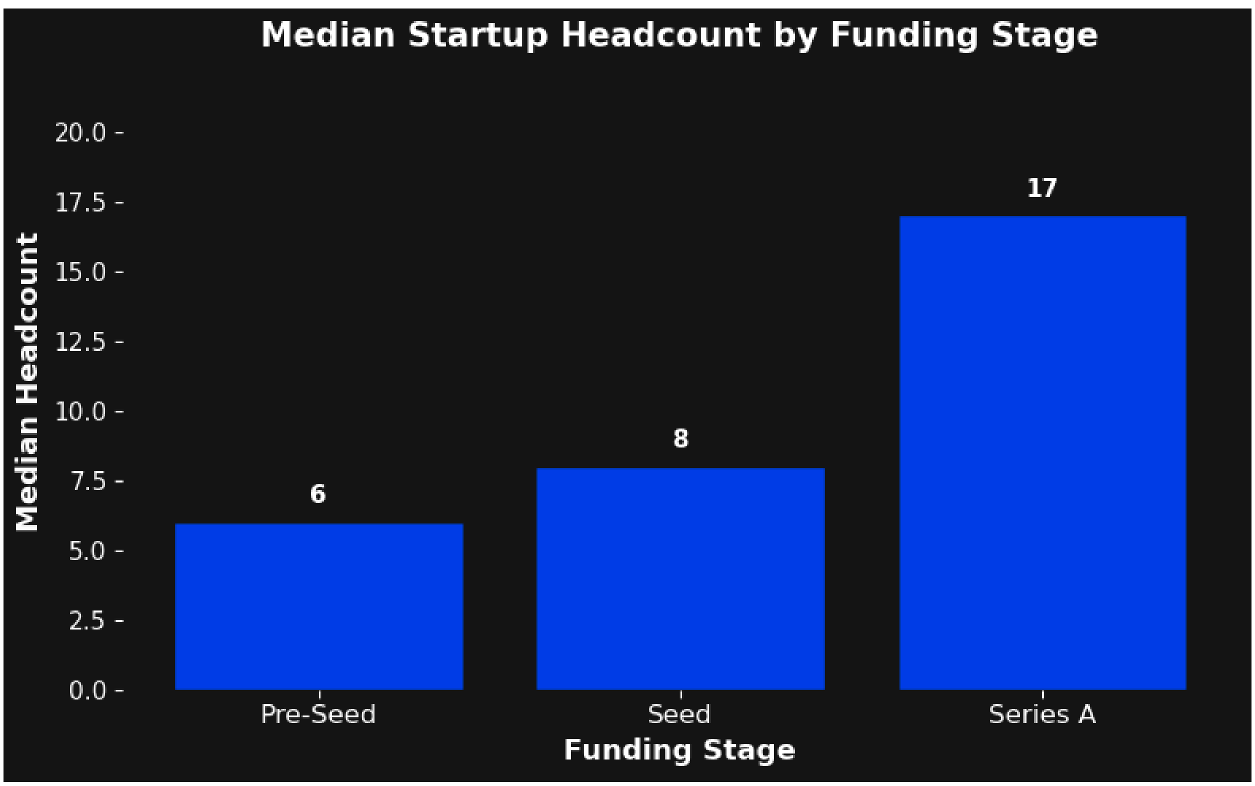

Figure 20: Median Startup Headcount by Funding Stage

4.2.4 Founder Gender Distribution

The gender distribution of founders reveals a substantial and persistent imbalance. Among funded startups, 81.8% of founders are male and only 18.2% are female, a near

4.5:1 ratio (Figure 21). This gap is almost identical in the non-funded sample (80.6% vs. 19.4%), suggesting that female underrepresentation is systemic rather than purely a function of funding outcomes. These findings are consistent with prior research showing that women face structural barriers in venture capital, including reduced access to investor networks, reliance on biased heuristics in decision-making, and greater scrutiny in pitch evaluation. The data highlights that despite growing awareness of gender diversity, female founders remain disproportionately excluded from both the entrepreneurial pipeline and venture financing, underscoring the persistence of gender bias within early-stage ecosystems.

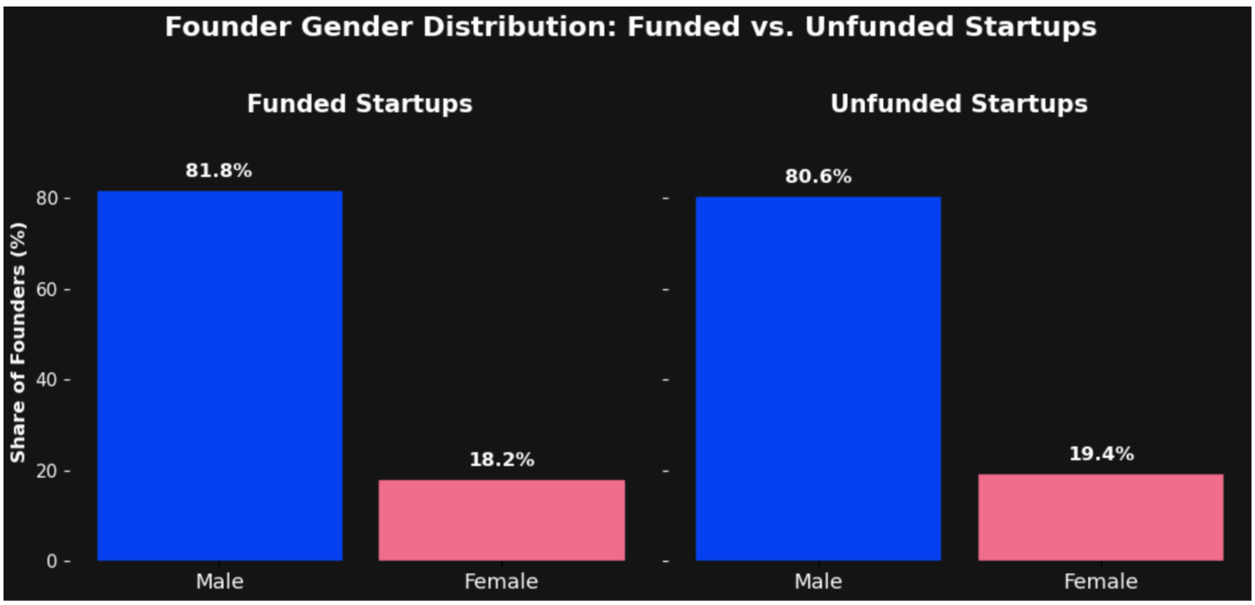

Figure 21: Founder Gender Distribution

4.2.5 Institutional Backgrounds

4.2.5.1 Previous Employers

The career histories of funded startup founders are dominated by Big Tech experience. Google (0.95%), Microsoft (0.90%), and Amazon (0.84%) emerge as the top three “feeder” companies, followed closely by IBM, Apple, and Meta (Figure 22). Furthermore, consulting and professional services firms also contribute meaningfully, with Deloitte, PwC, and Accenture collectively representing a notable segment, while financial institutions such as Goldman Sachs appear in the top 10 as well. These results suggest that entrepreneurial pipelines into funded startups are primarily shaped by Big Tech as the leading talent pool, but complemented by founders with backgrounds in consulting and finance.

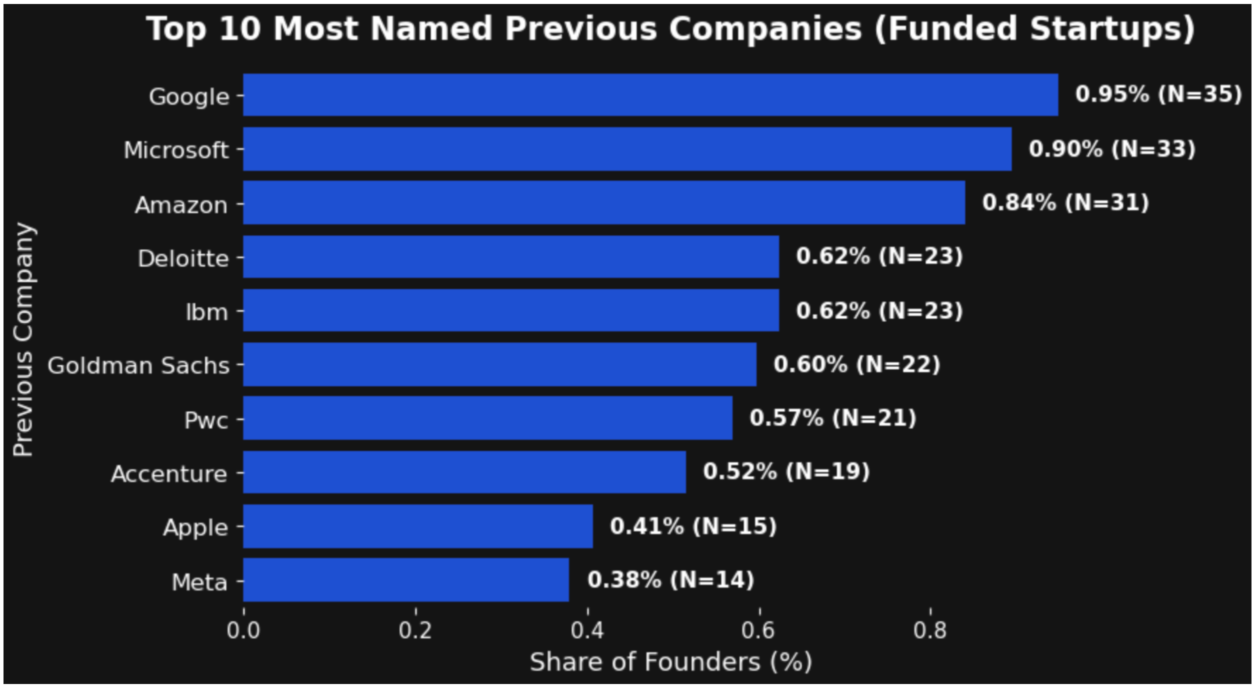

Figure 22: Most Common Previous Companies Among Funded Startups

4.2.5.2 Educational Backgrounds

Funded startup founders frequently come from well-established universities, with the University of Oxford accounting for the largest share (Figure 23). Other UK institutions, including Imperial College London, University of Cambridge, University College London, and King's College London, together represent a substantial share of the top educational pipelines. Prominent US universities such as Stanford, MIT, and Harvard also appear among the top feeders, reflecting their global influence in entrepreneurship. The inclusion of business schools such as INSEAD and London Business School demonstrates the role of both technical and management education in shaping startup leadership. While prestigious institutions account for a disproportionate share, the overall percentages remain modest, indicating that successful founders also emerge from a wide variety of educational backgrounds.

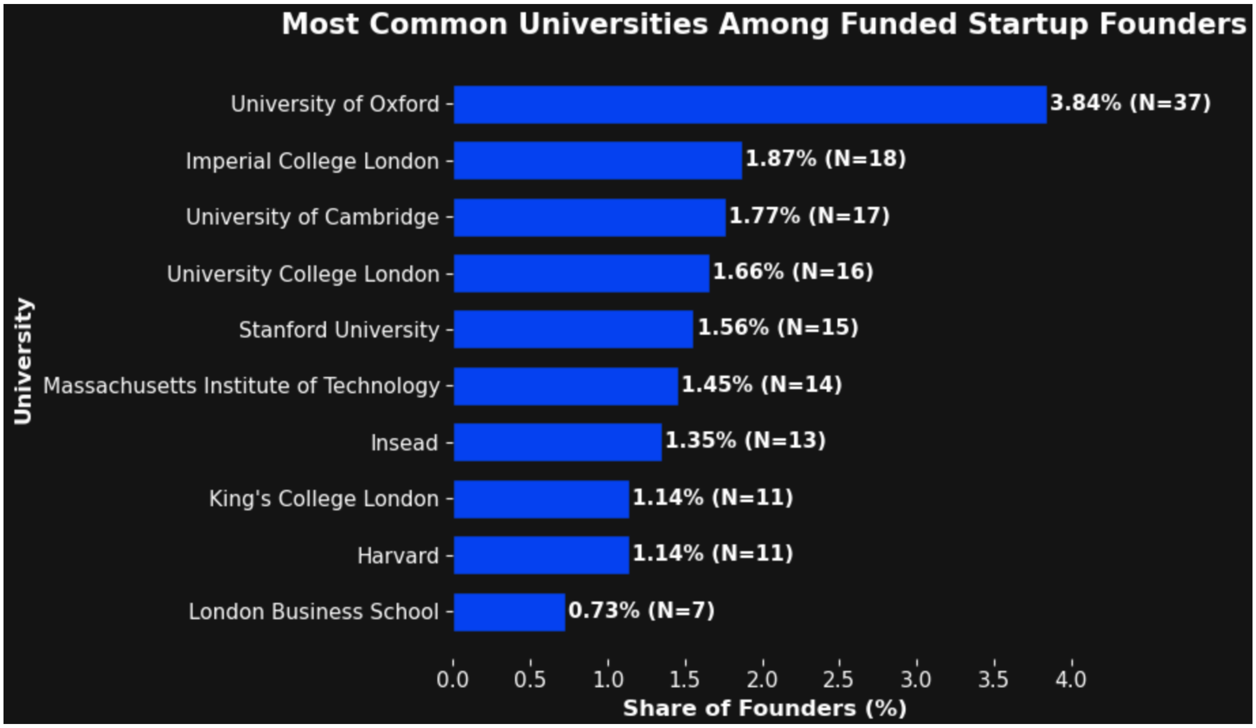

Figure 23: Most Common Universities Among Funded Startups

4.2.5.3 Highest Degree Attained

Beyond institutional background, the analysis also examined the highest degree obtained by founders and its relationship to funding outcomes. In this framework, only the most recent or terminal degree was considered, meaning that founders recorded as PhD holders will have also completed a Bachelor's (and often a Master's), but are represented only once under their highest level. This avoids double-counting across categories and ensures comparability across education groups.

The results, shown in Figure 24, are based on the difference in representation between founders of funded startups and those of non-funded startups. Clear patterns emerge: PhD holders are +34.7% more prevalent among funded ventures relative to their share in unfunded ones. MBA founders also show a positive association with funding outcomes (+12.3%), while Master's degrees exhibit a more modest uplift (+4.8%).

These findings suggest that advanced degrees, particularly MBAs and PhDs, provide important credibility signals to investors, whether through perceived technical expertise, managerial capability, or access to elite professional networks.

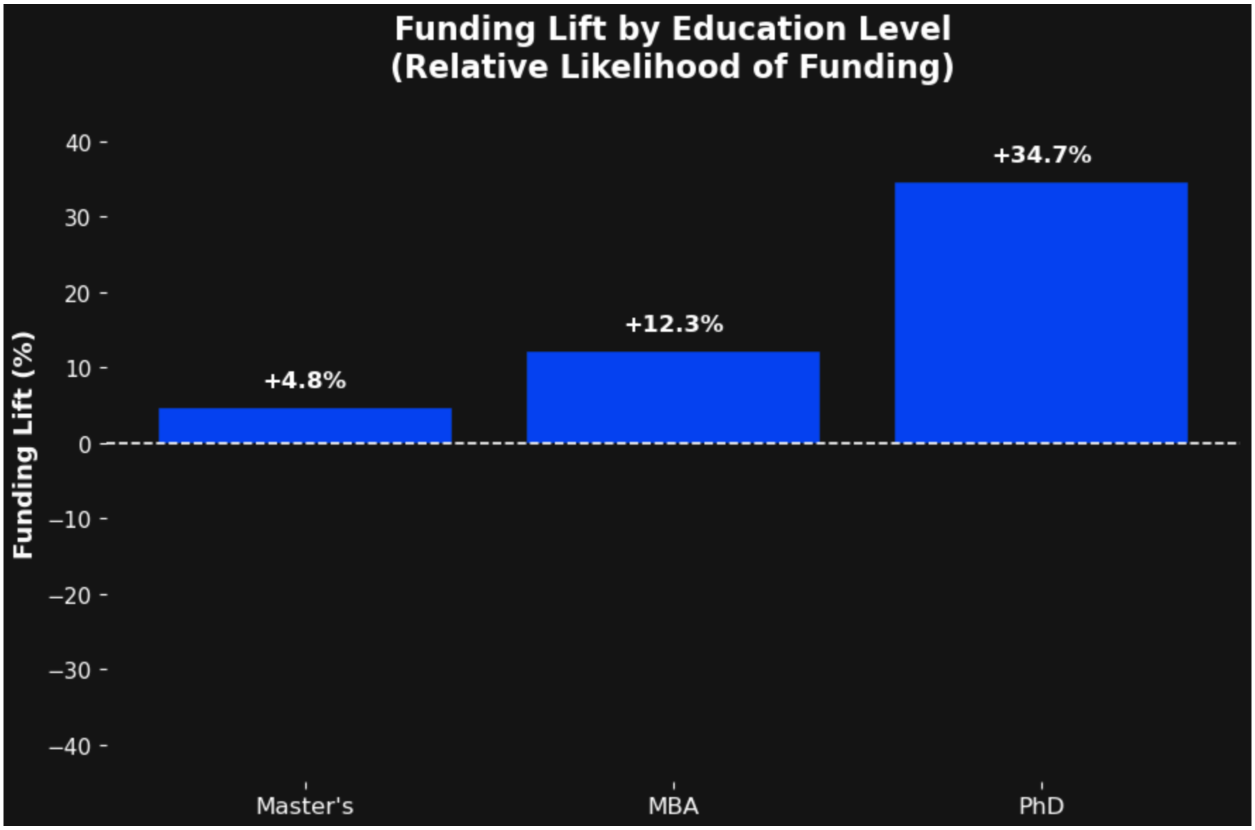

Figure 24: Funding Lift by Educational Level

4.3 AI in the Spotlight

4.3.1 Identifying AI/GenAI Mentions in Pitch Decks

To measure how often startups reference AI and generative AI over time, we created a standardized list of the most common AI-related buzzwords found in the dataset. This list included variations and synonyms such as “AI,” “artificial intelligence,” “machine learning,” “ML,” “generative AI,” “GenAI,” and “LLM,” among others. All text extracted from each deck was lowercased, punctuation was removed, and words were normalized to ensure consistent matching (e.g., “A.I.” → “ai”).

A Boolean check was then performed to flag whether any of these terms appeared at least once in a deck. For each quarter, we calculated the share of decks containing one or more of these terms and plotted the result. This approach ensured that variations in spelling or formatting did not cause undercounting, and that trends reflected consistent detection of AI-related mentions across time (Figure 25).

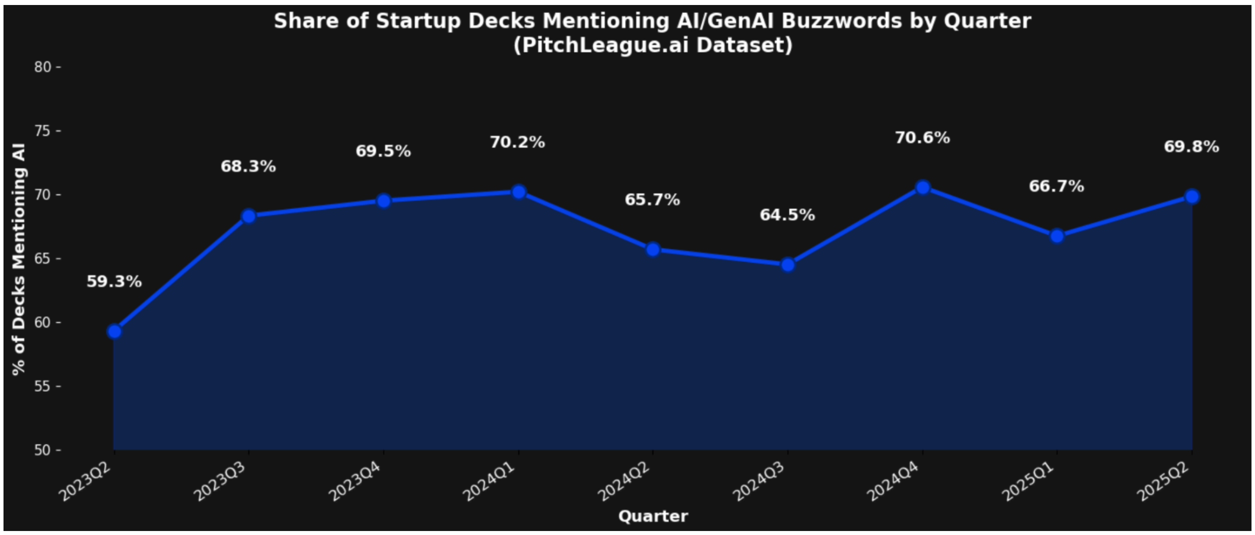

Figure 25: AI/Gen AI Buzzwords mentioned by Quarter

To track quarter-on-quarter AI adoption among product-focused startups, we identified all decks in which the extracted product description, market positioning, or business model explicitly referenced the use of AI technologies. For each quarter between Q2 2023 and Q2 2025, we calculated the percentage of product startups that incorporated AI into their offerings. The resulting trend shows a steady increase in AI adoption over time, with notable growth from under 41% in 2023 to nearly 59% by early 2025, before a slight decline in Q2 2025. Importantly, this growth rate outpaces the rise in generative AI buzzword mentions over the same period, not only because it has accelerated more sharply in recent quarters, but also because it started from a lower baseline percentage, highlighting a deeper, more sustained shift in actual product integration compared to surface-level hype (Figure 26).

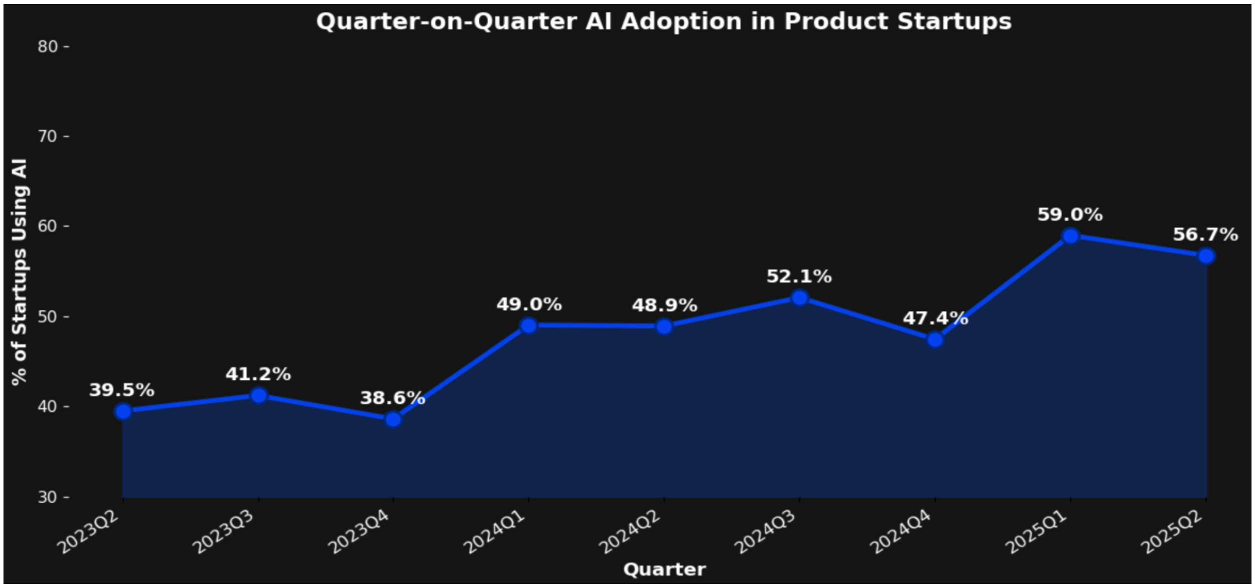

Figure 26: QoQ AI Adoption

4.3.2 Buzzwords in 2025 Pitch Decks

Beyond AI-specific mentions, it is useful to contextualize how often AI appears compared to other common buzzwords in 2025 pitch decks. Figure 19 highlights the twenty most frequently used terms, with “platform” (61.1%) dominating overall, followed by “AI” (32.3%), “impact” (30.8%), and “analytics” (25.2%). Other recurring themes include go-to-market strategies, MVPs, SaaS models, marketplaces, and ecosystems, all central to early-stage venture narratives (Figure 27).

Interestingly, while “AI” ranks second overall, its prevalence remains well below more generic framing devices such as “platform.” This contrast underscores a dual reality: on one hand, AI has rapidly become one of the most salient concepts in startup positioning; on the other, founders continue to rely heavily on broad, catch-all labels that frame their ventures in scalable, investor-friendly terms.

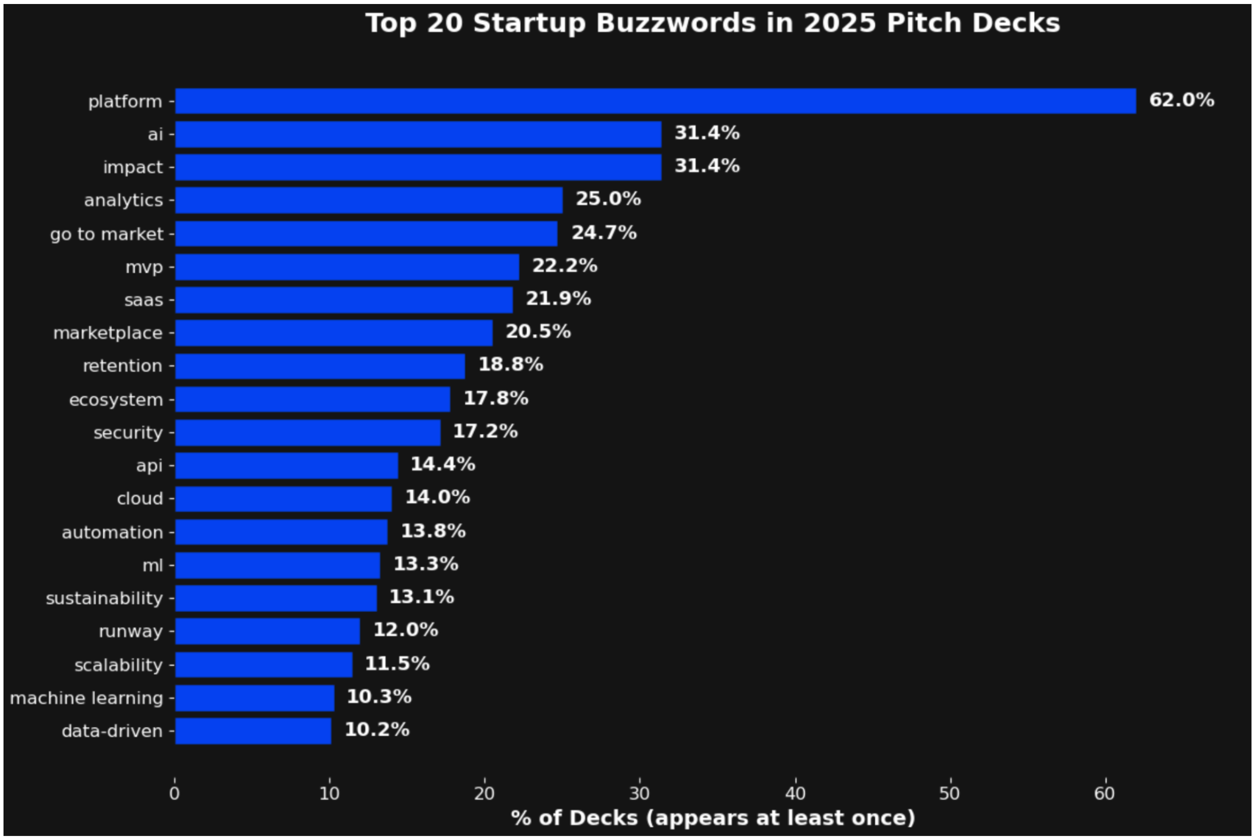

Figure 27: Top 20 Buzzwords in 2025 Pitch Decks

4.4 Inside the Pitch Deck: What Stands Out and What's Missing

4.4.1 Missing Deck components

A close analysis of the submitted pitch decks shows a strong adherence to core storytelling components (Figure 28). Nearly all founders clearly articulate the solution (97%) and the problem (91%) their startup addresses, while team information (88%), business model (83%), and market size (80%) are also featured in the majority of decks. This consistency aligns well with industry best practices, demonstrating that founders understand the importance of framing their story and providing context for investors. Slides covering progress or traction and market trends appear in more than two-thirds of decks, indicating a growing awareness among founders of the need to communicate both momentum and market positioning.

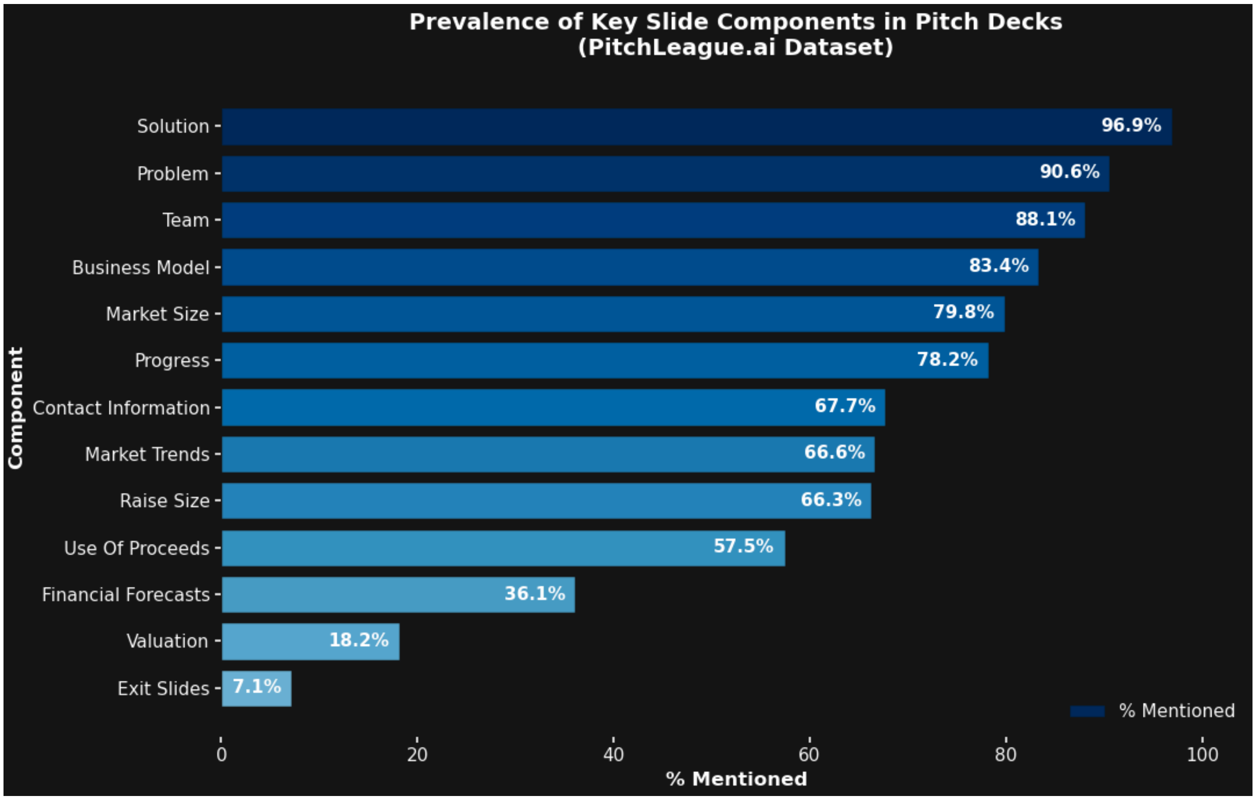

Figure 28: Prevalence of Key Slide Components

Yet, notable gaps remain in investor readiness. Indeed, the low presence of valuation (18.2%) and exit slides (7.1%) are a good sign. Indeed, at the Pre-Seed and Seed stages, it's not the founder's role to define valuation or project exit scenarios, that responsibility lies with investors aiming to return capital. Instead, more practical concerns are the 33% of decks missing raise size and the 32% lacking contact information, both of which directly hinder investor evaluation and follow-up.

4.4.2 Deck Components Funded vs Not Funded

This analysis compares the presence of key contact fields and core pitch deck components between funded and unfunded startups, expressed as percentage point differences (Figure 29). Contact details stand out as the most decisive gap, funded startups are 61.2 percentage points more likely to include a website and 21.2 percentage points more likely to list an email address. Together, these findings suggest that unfunded founders often neglect basic but critical credibility signals, making it harder for investors to engage. Beyond contact information, traction/progress slides (+20.4 pp) also show a notable advantage for funded startups, highlighting the importance of evidencing momentum. By contrast, sections like team, business model, market size, and problem/solution appear in most decks regardless of funding outcome, leading to smaller gaps. This pattern indicates that while most founders understand the need for these core sections, it is the combination of clear contact details and proof of traction that more sharply differentiates funded from unfunded startups.

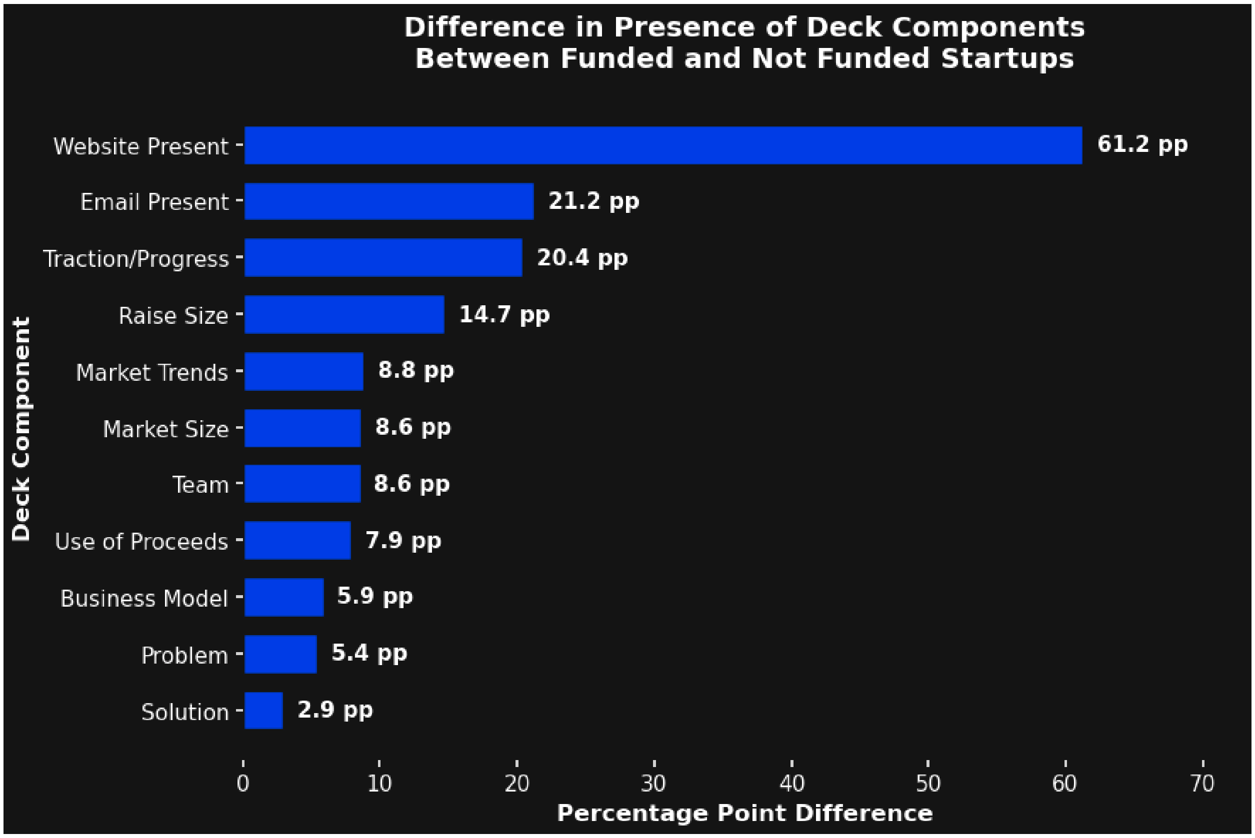

Figure 29: Deck Components Funded vs Not Funded

4.4.3 Most Overused Phrases

This analysis highlights the prevalence of recurring phrases in startup pitch decks (Figure 30). Frequently used expressions such as *"with a focus on"* and *"at the forefront of"* appear polished at first glance but add little substantive value. Indeed, nearly 200 decks in the dataset included the phrase *"with a focus on"*, underscoring how often investors are exposed to identical language.

Other common formulations, including *"a proven track record"* and *"our vision is to"*, are typically intended to establish credibility or signal ambition. However, their overuse has rendered them clichés. Without supporting evidence, claims of a *"proven track record"* carry limited weight, while visionary framings such as *"on a mission to"* or *"the next generation of"* lose their impact when repeated across many decks.

These findings suggest that many founders default to generic phrasing rather than articulating what truly differentiates their ventures. For investors, such language risks being perceived as filler, with attention shifting instead to sections of the deck that provide concrete traction, data, or unique insights

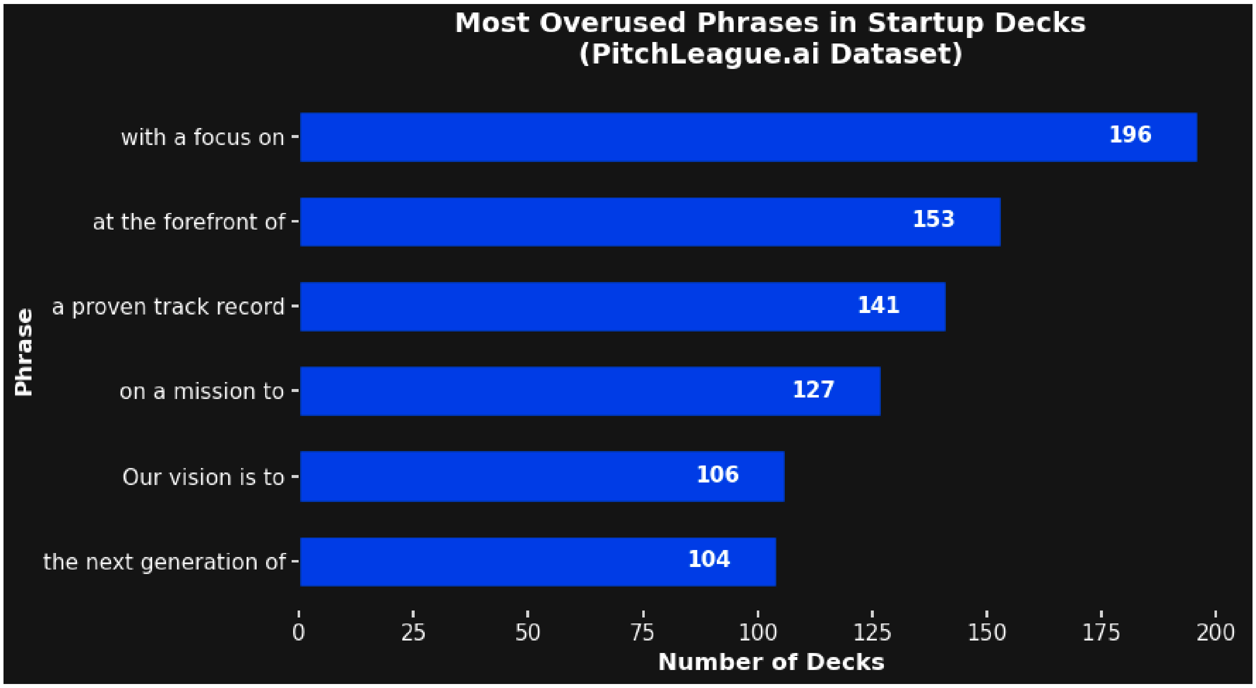

Figure 30: Most Overused Phrases

4.4.4 How Pitch Deck Colors Relate to Funding Success

The analysis of pitch deck design reveals both prevalence and funding impact of background color choices. As shown in the chart, the most common backgrounds in funded decks are blue (17.9%), black/very dark (16.3%), and light gray (14.7%), followed by white/off-white (8.0%) and dark gray (6.2%) (Figure 31).

However, median funding raised differs significantly by color choice. Decks with a black or very dark background achieved the highest median funding at $1.20M, suggesting investors may perceive darker tones as more polished or professional. White/off-white decks followed closely at $1.00M, reinforcing their position as a safe, clean design choice. Among bolder palettes, red decks secured a median of $0.90M, while orange decks attracted $0.65M. At the lower end, dark gray backgrounds corresponded with $0.64M in funding (Figure 32).

Overall, these results indicate that while blue and gray palettes dominate in frequency, it is the high-contrast (black/white) and bold accent (red) designs that perform best in terms of capital raised. Conversely, intermediate tones like orange and dark gray appear less effective, suggesting that investors respond more favorably to either classic minimalism or striking, high-impact color schemes.

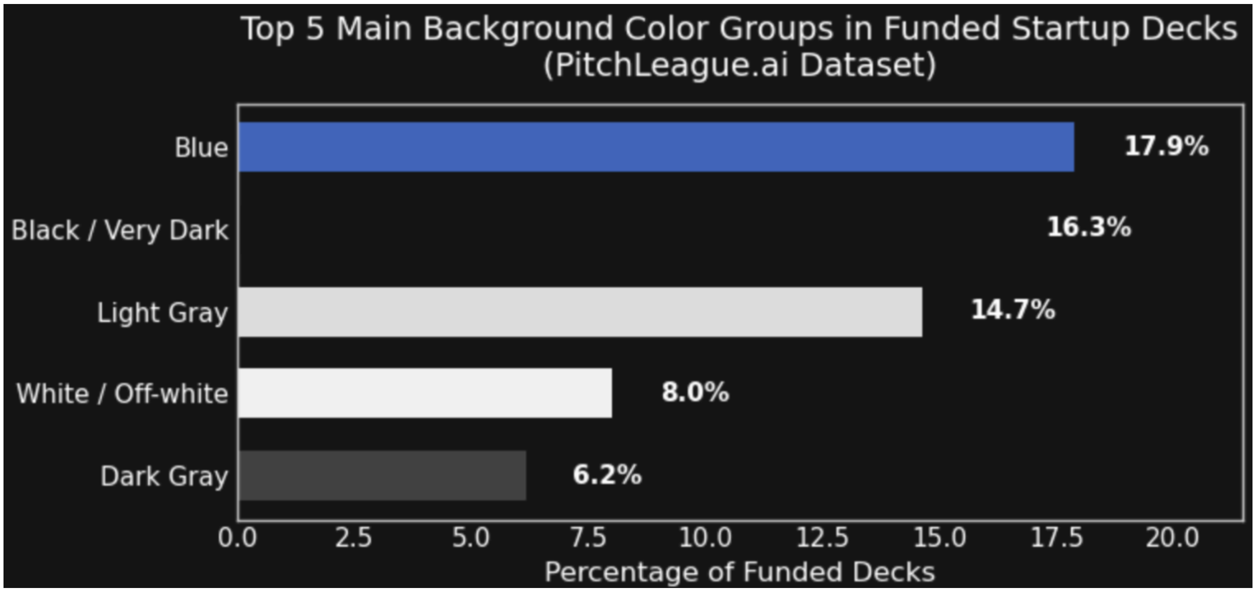

Figure 31: Top 5 Background Colors in Funded Startups

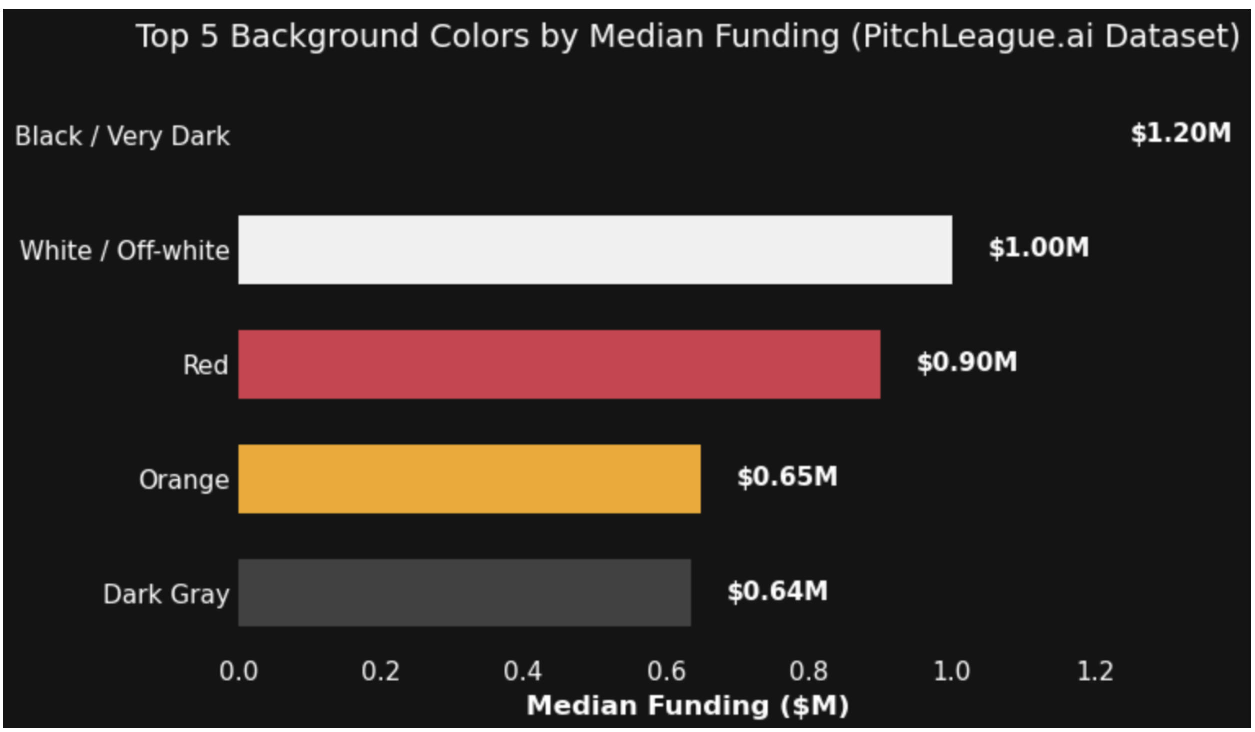

Figure 32: Top 5 Background Colors by Median Funding

4.4.5 ARR Availability

This analysis examines how the presence of Annual Recurring Revenue (ARR) in pitch decks relates to funding outcomes (Figure 33). Across all stages, startups that reported ARR were funded at notably higher rates. At the Pre-Seed stage, 28% of startups with ARR secured funding compared to only 12.1% without. The gap persists at Seed (33.6% vs. 17.9%) and Series A (40.7% vs. 24.1%). It is important to note that these percentages reflect funding success rates within each subgroup (those with ARR vs. those without), rather than proportions of the overall sample, hence the two figures at each stage do not add up to 100%. These results suggest that even modest early revenue signals can significantly strengthen investor confidence, serving as concrete evidence of traction beyond projections.

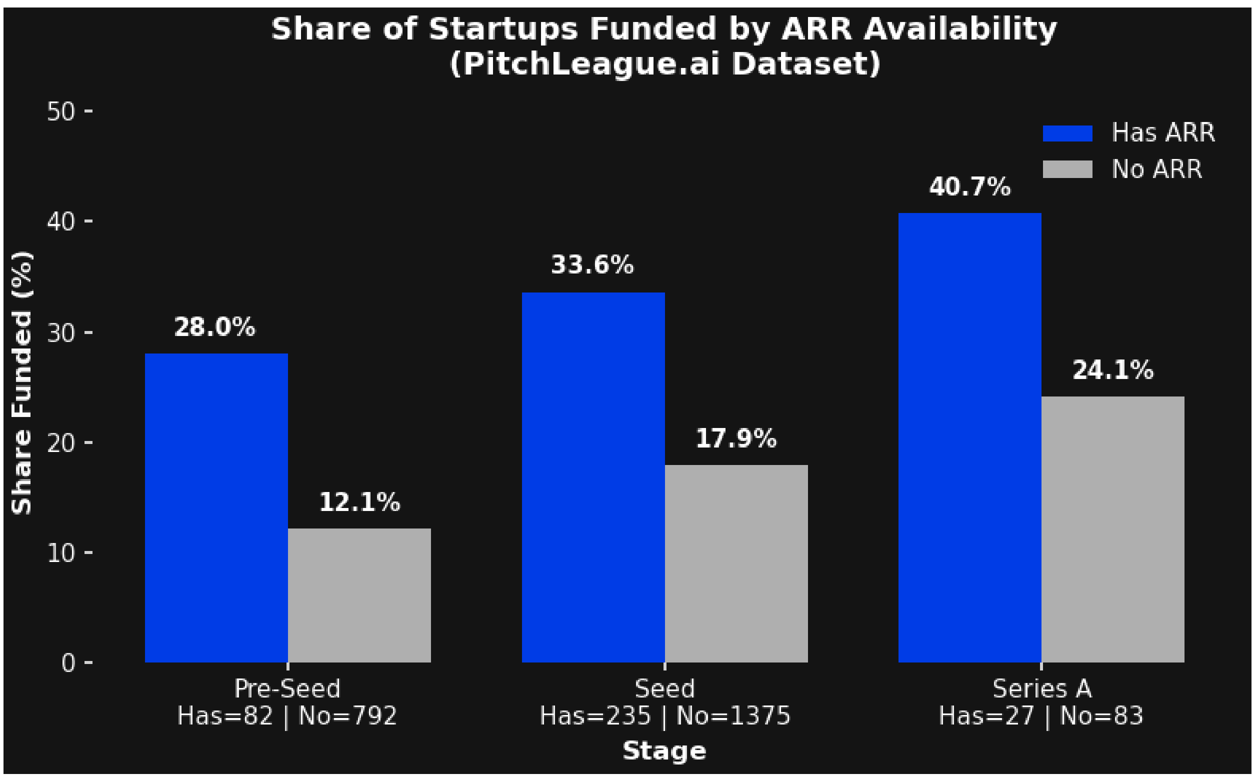

Figure 33: Share Funded by ARR Availability

4.4.6 Investor Mentions

This analysis looks at whether explicitly mentioning investors in pitch decks is linked to higher funding success. Even at the Pre-Seed stage, founders often reference investors, ranging from friends and family and early angels, to accelerators and incubators. We can see that startups that referenced investors had consistently higher funding rates than those that did not (Figure 34). At Pre-Seed, 33.1% of startups mentioning investors were funded compared to just 10.5% that did not. The same trend holds at Seed (36.5% vs. 16.2%) and Series A (38.6% vs. 21.2%). It is normal that the two numbers at each stage do not sum to 100%, since they represent funding success rates within each subgroup (mention vs. no mention) rather than proportions of the total sample. These findings highlight the signaling power of investor mentions: referencing prior or current investors appears to boost credibility and reassurance, helping founders stand out in competitive fundraising environments.

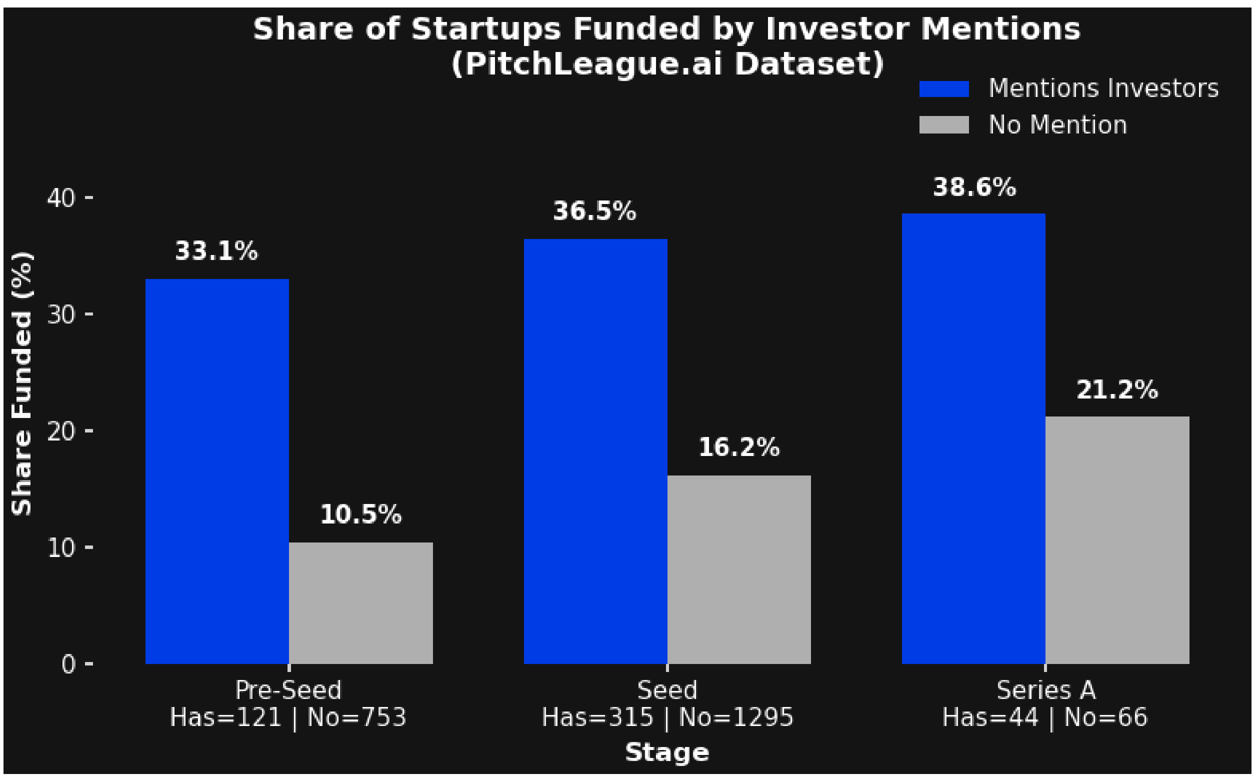

Figure 34: Share Funded by Investor Mentions

4.5 Pitchleague Scoring System

4.5.1 Median Scores by Funding Stage

In addition to funding outcomes, PitchLeague scores also display systematic variation across funding stages. As shown in Figure 35, overall median scores increase gradually as startups progress through the investment pipeline. Pre-Seed ventures record a median overall score of 41.2, rising modestly to 42.6 at Seed and 46.5 at Series A. A similar pattern is observed for Startup Quality Scores (32.0 to 34.0 to 39.5), suggesting that later-stage ventures demonstrate stronger fundamentals in line with investor expectations.

By contrast, Deck Quality Scores start higher at the Pre-Seed stage (78.0), dip slightly at Seed (76.0), and then decline further to 72.0 at Series A. This inverted pattern may reflect how early-stage teams, lacking traction, rely more heavily on polished presentation materials to compensate, whereas later-stage companies depend less on decks and more on demonstrated business performance.

Taken together, these results highlight that while Startup Quality Scores trend upwards with stage, Deck Quality is relatively strongest in the earliest phases, underscoring its role as a critical credibility signal when hard operating metrics are absent.

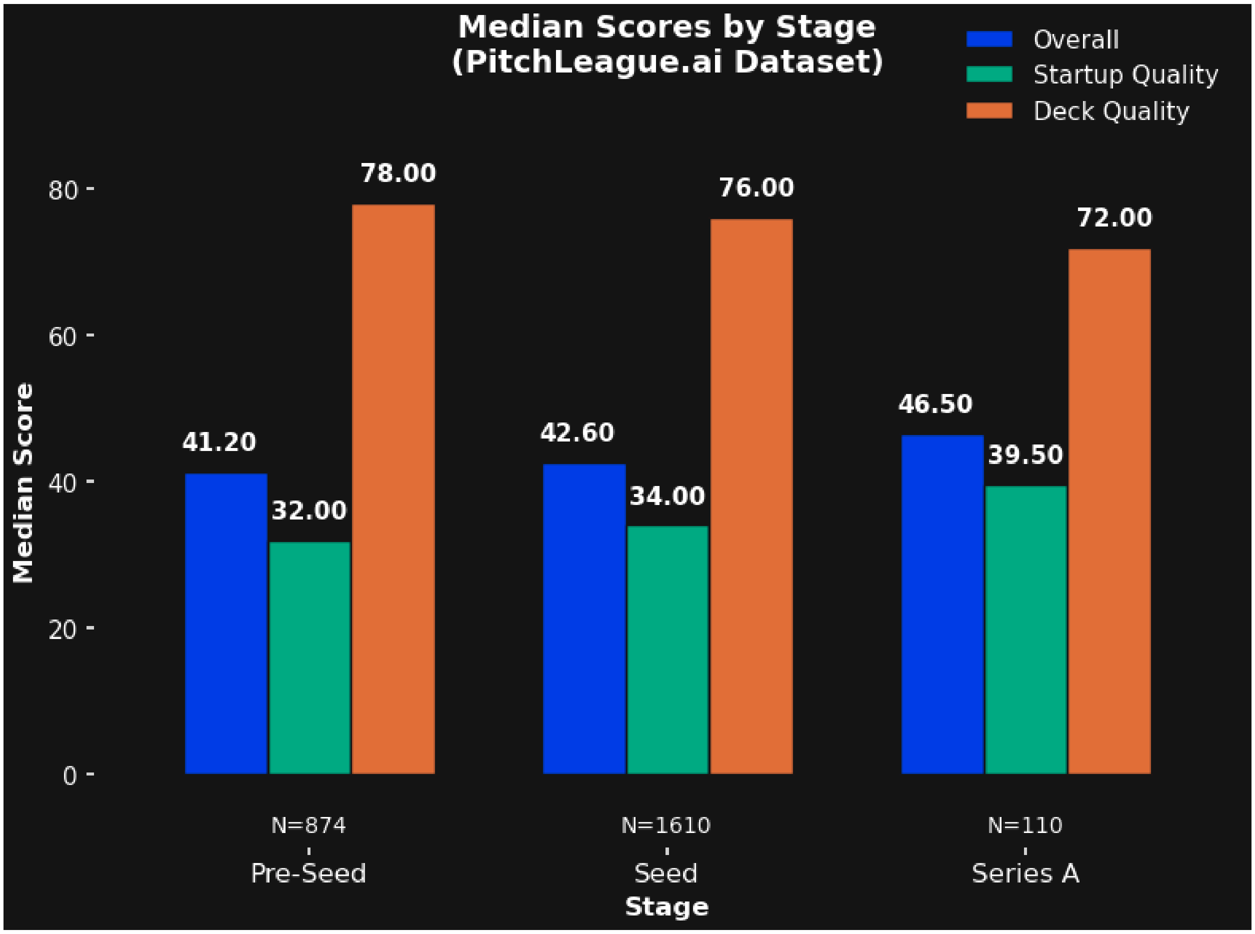

Figure 35: Median Scores by Stage

4.5.2 Investment Signals Across Regions: Deck Quality vs. Startup Potential

For this analysis, we considered only countries with at least 20 pitch decks to ensure fair comparisons. In terms of median "Deck Overall" scores, which assess spelling and grammar, structure, length, and clarity, we can see that Ukraine (79.5), Israel (79.0), and Poland (79.0) lead the way, followed by Saudi Arabia (77.5), Kenya (77.0), and Nigeria (77.0). This shows that several emerging ecosystems are producing decks on par with, or even stronger than, those from established hubs.

However, when looking instead at the median "Startup Overall" scores, which evaluate traction, market, and team, the picture changes significantly. Indeed, Finland (46.0) is far ahead, followed by Norway (35.0), Singapore (33.0), and Australia

(32.0). The United Kingdom (31.0) and the United States (28.0) also appear, alongside Spain and Switzerland (30.0 each). These results suggest that while some ecosystems excel in deck craftsmanship and communication quality, others stand out more for underlying venture fundamentals such as execution, market opportunity, and team strength.

The comparison highlights a key insight: regions strong in presentation are not always those with the strongest startups, and vice versa. For investors, this underscores the importance of looking beyond pitch quality to evaluate the actual potential of ventures across geographies.

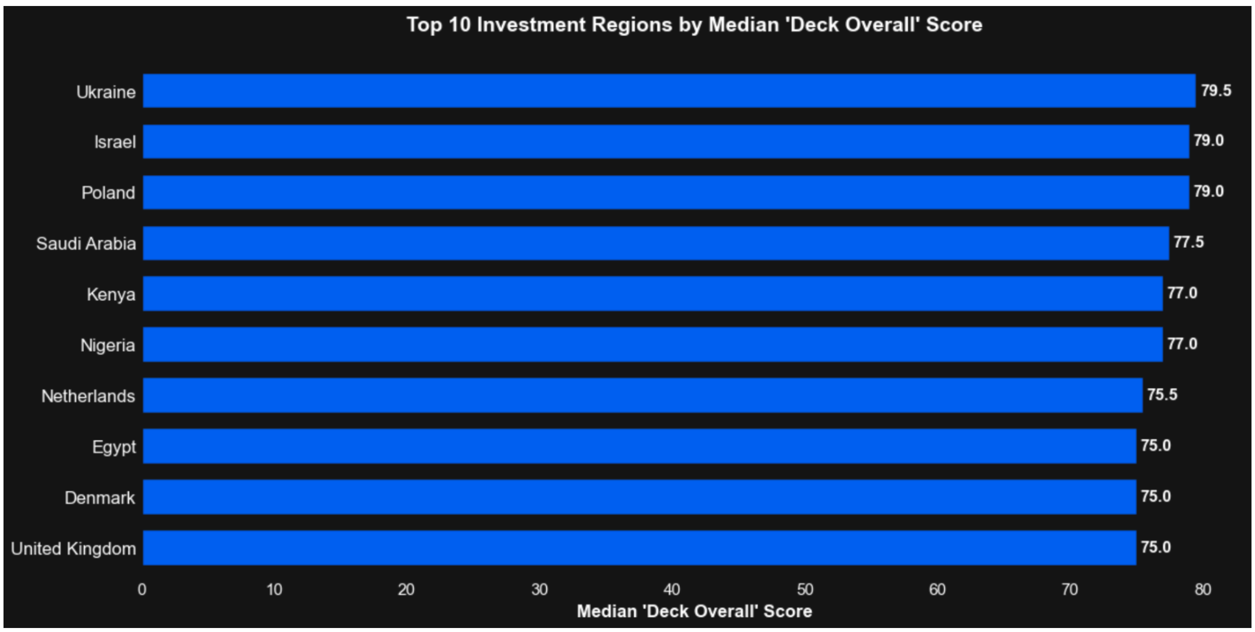

Figure 36: Top 10 Investment Regions by Median Deck Score

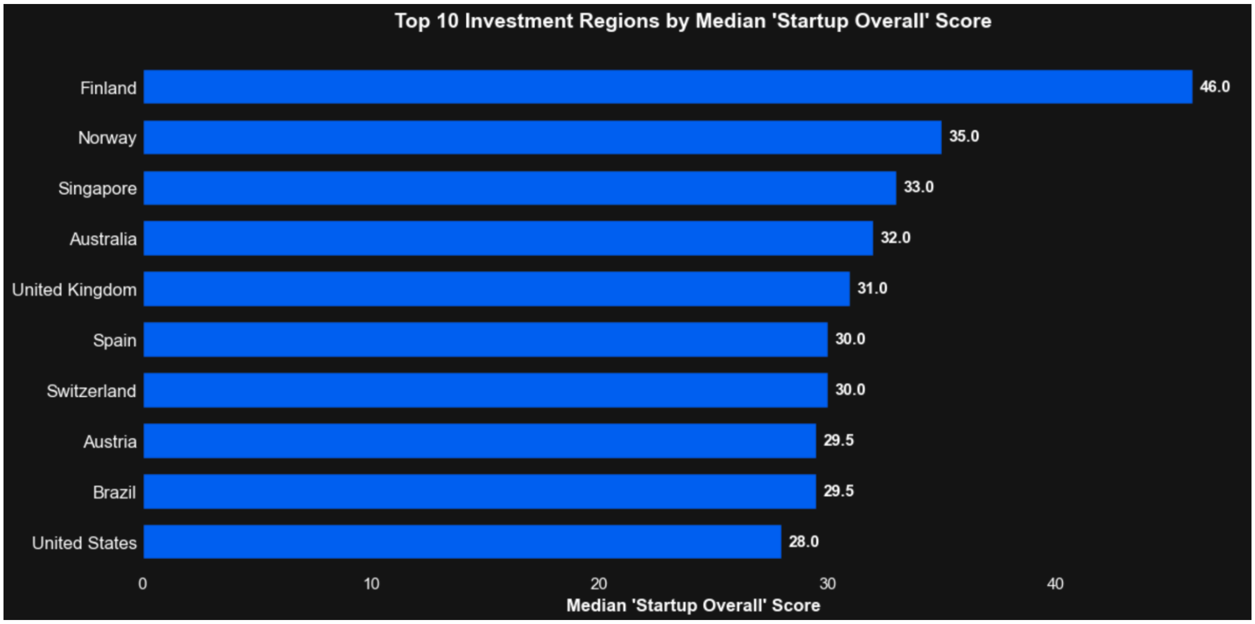

Figure 37: *Top 10 Investment Regions by Median Startup Score*

4.5.3 Median Scores: UK vs. US

This chart compares median scores by country for the United States and the United Kingdom, the two most represented geographies in the dataset (Figure 38). The results show that while both ecosystems produce strong decks, the United Kingdom edges out the United States across all score types. U.K. startups achieve a slightly higher overall score (39.6 vs. 38.0) and startup score (31.0 vs. 28.0), as well as a marginally stronger deck score (75.0 vs. 74.0). These consistent advantages suggest that, within this dataset, U.K. startups are better positioned both in how they present and in how their fundamentals are assessed, making the U.K. the stronger performer between the two.

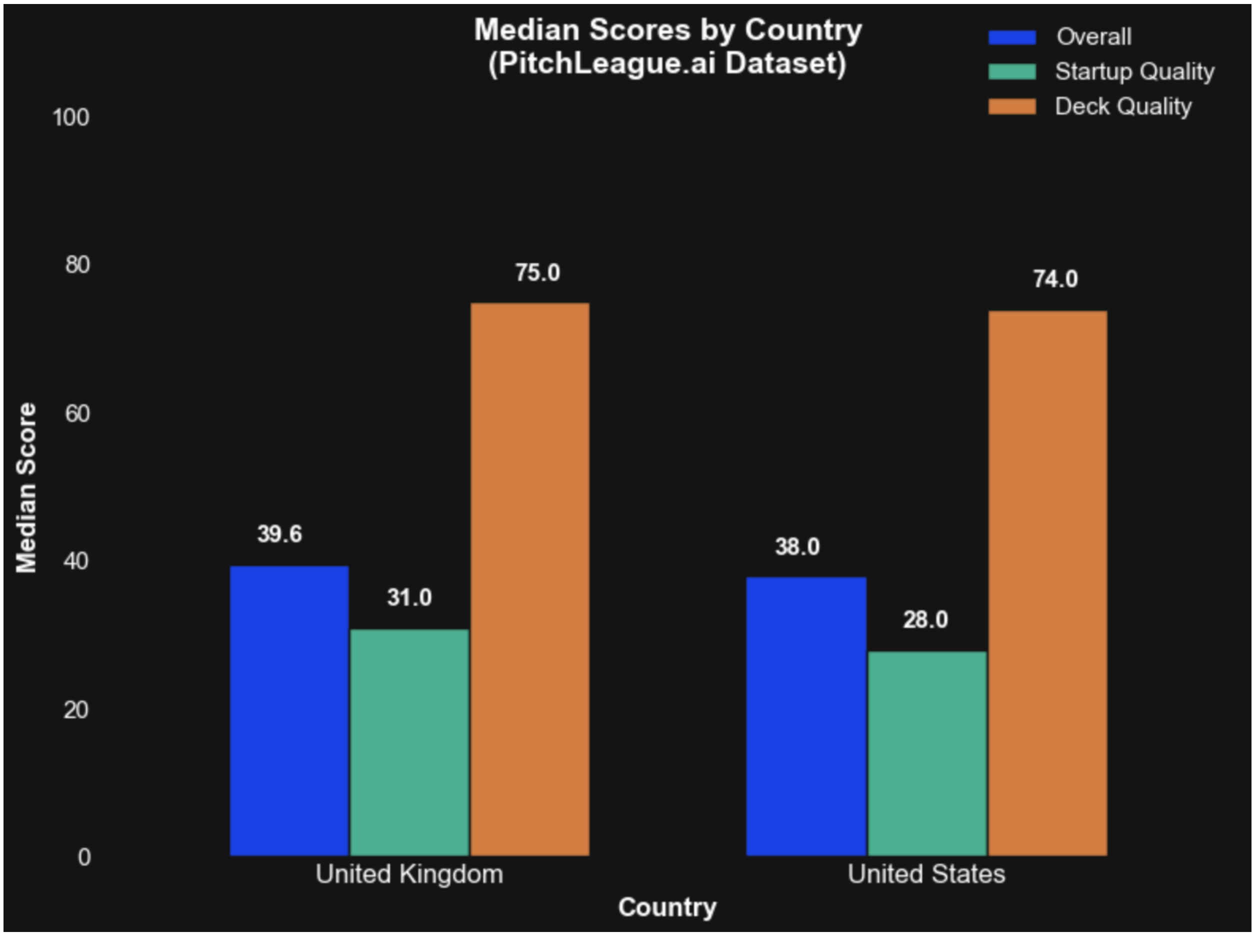

Figure 38: Median Scores UK vs US

4.5.4 Median Funding by Score Percentile

To assess whether PitchLeague's scoring system correlates with fundraising outcomes, we grouped startups into score percentiles (Top 25%, Top 10%, Top 5%, and Top 1%) and compared their median funding against the overall funded sample baseline of $0.88M (Figure 39). The results show a strong positive relationship: startups in the Top 25% scored a median of $1.04M (+18% vs. baseline), the Top 10% raised $1.30M (+47%), the Top 5% reached $1.50M (+70%), and the Top 1% peaked at $1.55M (+76%).

The corresponding score thresholds for each percentile across overall, deck, and startup dimensions are presented in Figure 40, illustrating the cut-off points startups need to reach in order to enter higher-performing cohorts.

Taken together, these patterns suggest that PitchLeague's scoring system captures meaningful quality signals that align with investor behavior: as scores climb into higher percentiles, startups consistently raise more capital.

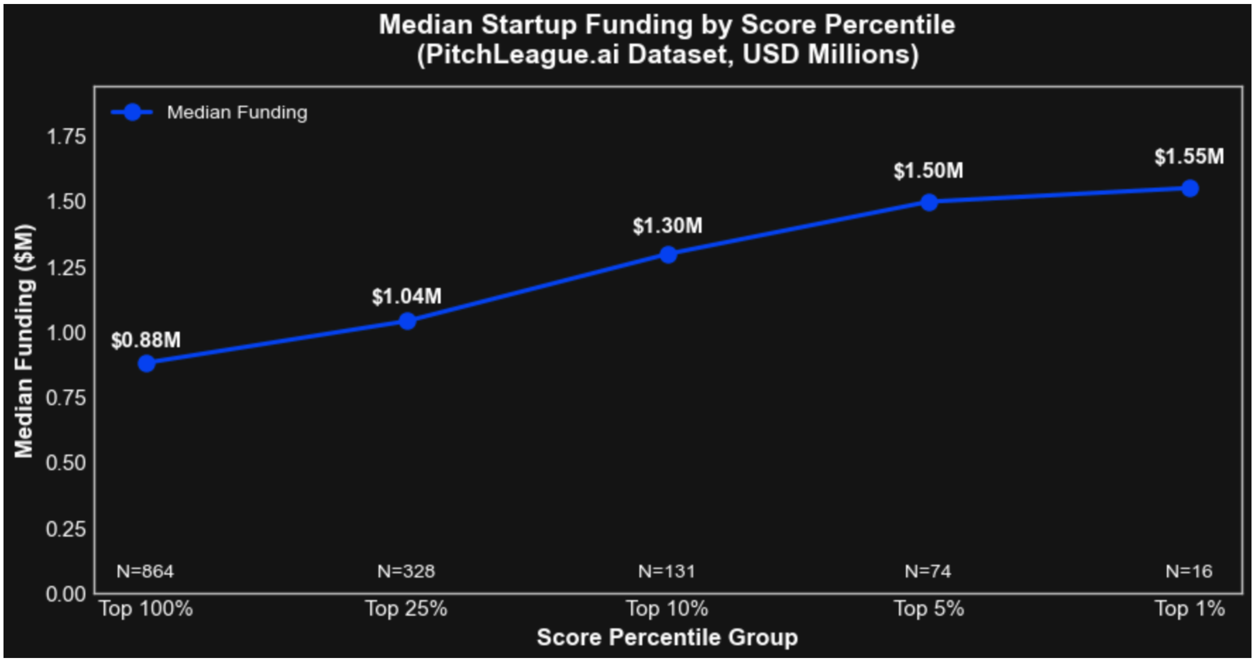

Figure 39: Median Funding by Score Percentile

Percentile	Overall Threshold	Deck Threshold	Startup Threshold
Top 100%	34.80	74.00	26.00
Top 25%	49.00	81.00	42.00
Top 10%	62.00	86.00	58.00
Top 5%	69.40	89.00	67.00
Top 1%	81.44	92.00	82.00

Figure 40: Scores by Percentile Groups

4.5.5 Startup Quality Score by Funding Status

Startups that successfully secured funding exhibited markedly higher Startup Quality Scores than those that did not (Figure 41). The funded group had a median score of 35.0 compared to 24.0 among non-funded startups, with corresponding means of 35.9 and 25.7. These results indicate that the Startup Quality Score captures meaningful dimensions of venture strength that align with investor preferences, reinforcing its value as a discriminative signal in the fundraising process.

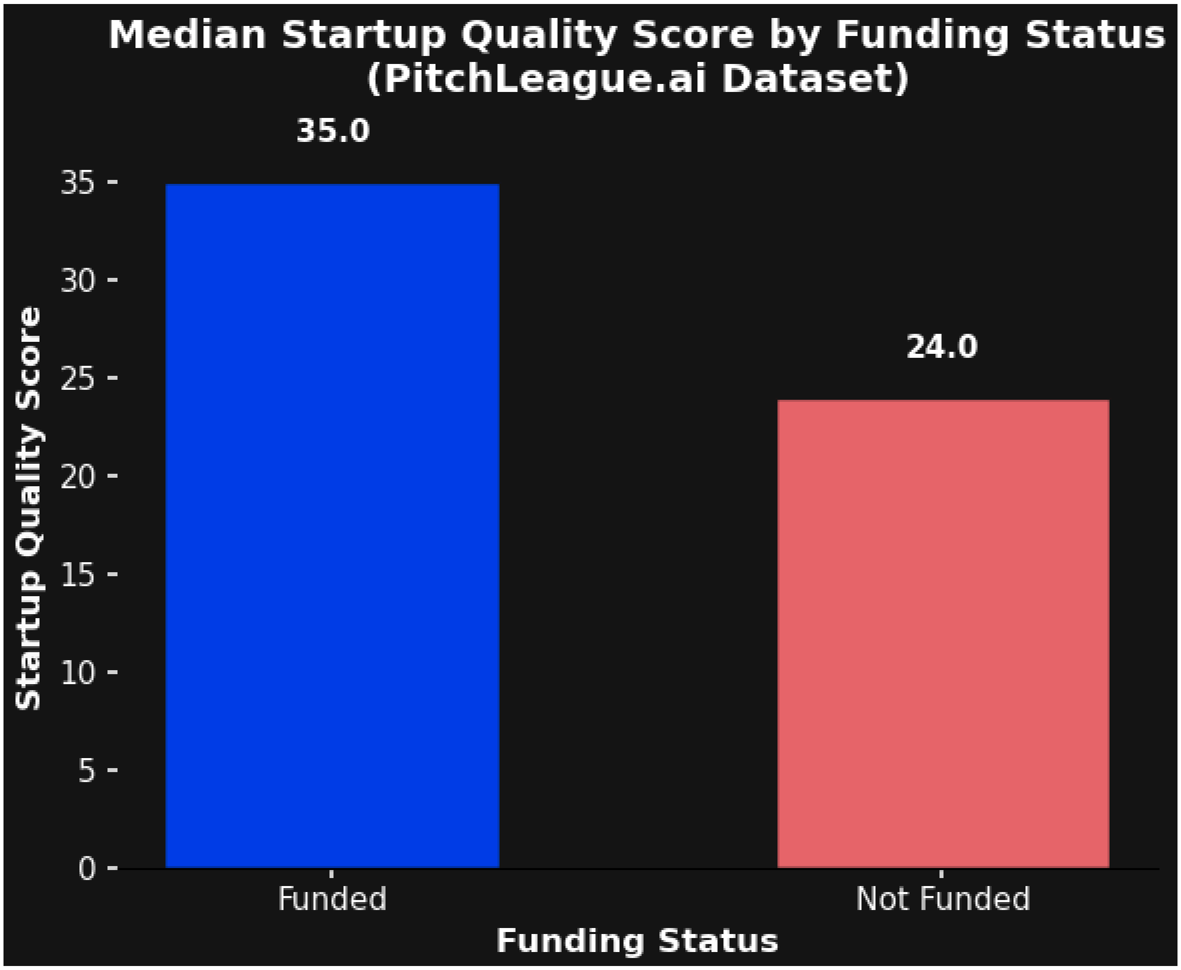

Figure 41: Startup Quality Score by Funding Status

4.5.6 Deck Quality Score by Funding Status

A similar but smaller pattern emerged when examining Deck Quality Scores (Figure 42). Funded startups achieved a median of 76.0, compared to 73.0 among their non-funded counterparts, with mean scores of 75.2 and 71.5 respectively. This suggests that pitch deck presentation quality does play a role in fundraising outcomes, though its explanatory power is weaker than that of the underlying Startup Quality Score.

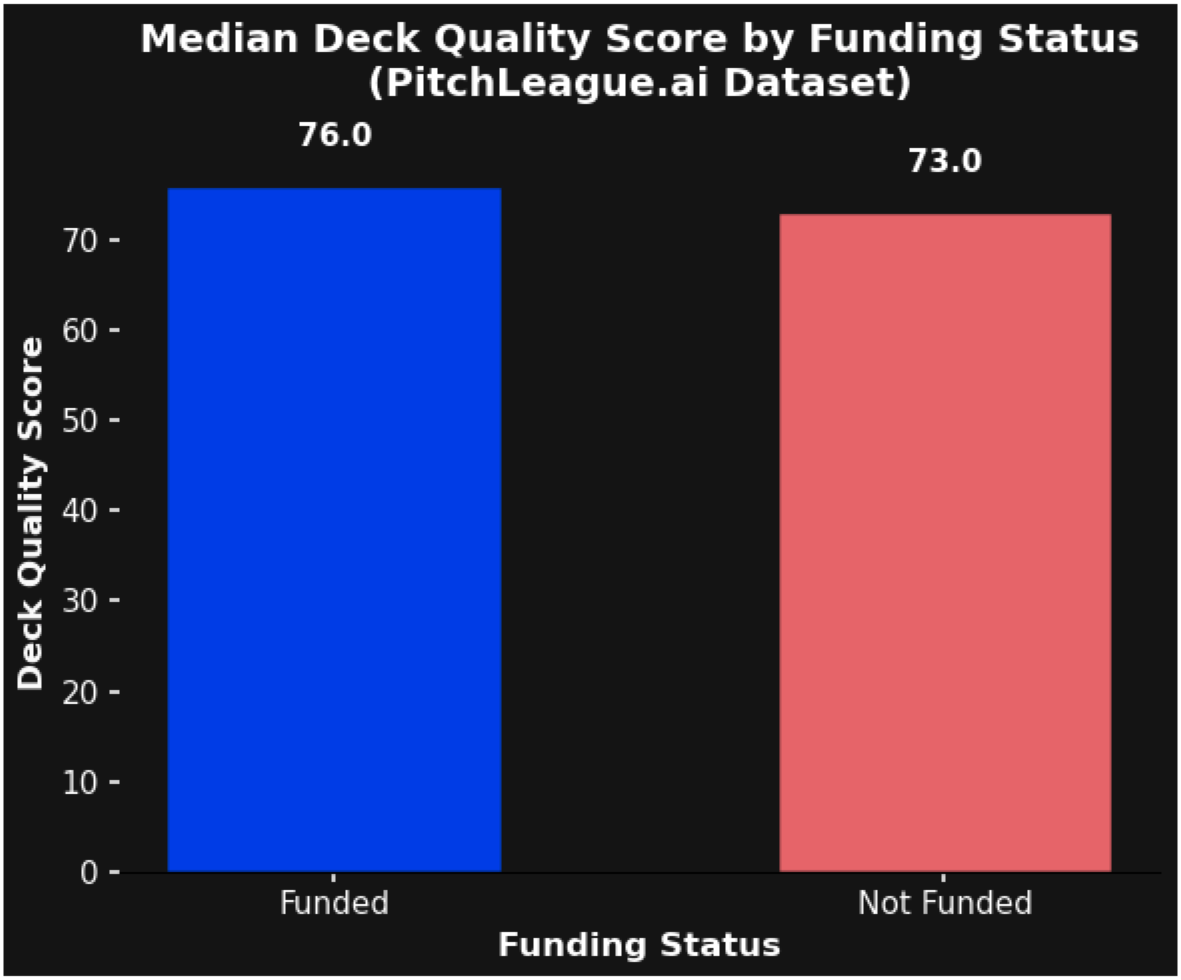

Figure 42: Deck Quality Score by Funding Status

5. Concluding Remarks

5.1 Key Findings

1. Funding Trends and Growth Dynamics

The early-stage funding environment expanded steadily from 2023 to 2025, with both median and total capital deployed rising. Startups progress along a predictable trajectory: capital raised, ARR, and market size claims increase sharply with each round, while gross margins dip during early scaling before stabilizing at Series A. These dynamics capture the shift from pre-revenue uncertainty toward evidence of traction and credible market opportunities.

2. Geographic and Sector Concentration

Funding is highly concentrated across a few hubs. The U.S. and U.K. dominate capital deployment, while India and Nigeria highlight growing activity in Asia and Africa. Sectorally, Fintech, Enterprise Software, and Healthcare together account for nearly half of all funding, underscoring investor preference for scalable, tech-driven markets.

3. Team as a Decisive Factor

Founder backgrounds strongly shape fundraising outcomes. Prior entrepreneurial experience, technical expertise, and startup exposure all correlate with higher funding rates. Multi-founder teams raise more often than solo founders, although the latter sometimes secure disproportionately larger rounds. Gender gaps remain striking: fewer than one in five funded startups are female-led, underscoring persistent structural inequities.

4. Concentration in Leading Employers and Universities

The founder pipeline is heavily shaped by elite institutions. Big Tech firms such as Google, Microsoft, and Amazon dominate prior employment histories, while universities like Oxford, Cambridge, Stanford, MIT, and Harvard supply a substantial share of founders. These networks concentrate opportunity, though successful founders also emerge from more diverse professional and educational backgrounds.

5. AI as Signal and Substance

AI has become both a narrative and a genuine trend. Mentions of AI/GenAI have risen sharply, but product-level adoption is expanding even faster. This indicates more than hype: AI integration reflects a deeper structural change in early-stage innovation.

6. Investor and Traction Signals Matter Most

Practical signals such as traction slides, ARR disclosure, investor mentions, and even basic contact details most clearly differentiate funded from unfunded ventures across all stages.

7. PitchLeague Scoring System

PitchLeague scores capture these dynamics effectively. Indeed, higher startup and deck quality scores align consistently with larger fundraising outcomes, but the startup quality score is the most representative measure of real investor behavior. Furthermore, funded startups score significantly higher than unfunded peers, and higher score percentiles correspond to higher median round sizes. This reinforces the potential of standardized, data-driven benchmarks to complement human judgment and reduce uncertainty in venture capital.

5.2 Limitations

While this study offers valuable insights into early-stage fundraising, there are important constraints to acknowledge. Pitch deck data captures only what founders choose to include, meaning that many decisive factors, such as market timing, founder perseverance, adaptability, resilience, and leadership qualities, are not directly observable. Broader ecosystem influences, including investor networks, "warm introductions," and sector sentiment, also play a major role in fundraising success but rarely appear in slide content.

The dataset is drawn solely from PitchLeague submissions, which may introduce selection bias, as these startups might not represent the wider founder population.

These constraints mean the findings should be interpreted as insights into what is visible in the deck, not as a complete measure of the full set of factors influencing investor decisions.

5.3 Future Work

Future research should extend beyond the current two-year snapshot by tracking startups over longer periods and multiple funding rounds. This would enable the analysis of fundraising persistence, post-deck outcomes, and even exit events which would provide a more complete picture of startup trajectories.

Additionally, integrating founder-level qualitative data, such as personality traits, resilience, and network strength, via surveys or follow-up studies would help capture the intangible drivers of startup success that are often missed by automated extraction.

As the field of data-driven VC evolves, there is growing value in collaborating with industry partners to share insights, test new features, and validate findings. Such efforts will ensure that future work not only advances academic understanding but also delivers actionable intelligence for investors and founders navigating the early-stage ecosystem.

References

- Armstrong, B. (2025) *How to create an investor pitch deck: essential skills for early-stage and Series A funding*, SVB Startup Insights, 2025. Available at: https://www.svb.com/startup-insights/startup-strategy/how-to-create-investor-pitch-deck-vc-angels/
- Buffington, M.B. (2025) 'MB on VC: The power law of venture capital—fact vs fiction', BIP Ventures. Available at: https://www.bipventures.vc/news/mb-on-vc-the-power-law-of-venture-capital-fact-vs-fiction
- Dealroom (2023) 'Venture Capital Unwrapped: Global Tech Q4 2023'. Available at: https://dealroom.co/guides/global
- Ewens, M., Nanda, R. and Rhodes-Kropf, M. (2022) 'Cost of experimentation and the evolution of venture capital', Journal of Financial Economics, 143(1), pp.138–161. Available at: https://www.nber.org/system/files/working_papers/w24523/w24523.pdf
- Gompers, P., Gornall, W., Kaplan, S.N. and Strebulaev, I.A. (2016) 'How do venture capitalists make decisions?', NBER Working Paper No. 22587. Available at: https://www.nber.org/papers/w22587
- Islam, M., Fremeth, A.R. and Marcus, A.A. (2018) 'Signaling by early stage startups: US government research grants and venture capital funding', Journal of Business Venturing, 33(1), pp.35–51. Available at: https://www.ivey.uwo.ca/media/3778210/fremeth-islam-marcus-forthcoming.pdf
- Pitchleague.ai (2025) Available at: https://www.pitchleague.ai/
- Sachs, M. & Unbescheiden, M. (2024) *Biases Influencing Venture Capitalists' Decision-Making: A Systematic Literature Review*. Available at: https://ssrn.com/abstract=4818073
- Sequel (2025). Available at: https://www.sequel.co/
- Tamaseb, A. (2021) 'Super Founders: What Data Reveals About Billion-DollarStartups'.
- Wang, X. & Ihlamur, Y. (2024) *An Automated Startup Evaluation Pipeline: Startup Success Forecasting Framework (SSFF)* [Preprint]. University of Oxford & Vela Partners. Available at:

https://www.researchgate.net/profile/Xisen-Wang-3/publication/381005574_An_Automated_Startup_Evaluation_Pipeline_Startup_Success_Forecasting_Framework_SSFF/links/6724da9aecbbde716b521c3d/An-Automated-Startup-Evaluation-Pipeline-Startup-Success-Forecasting-Framework-SSFF.pdf