The Selective Tailwind Effect of Artificial Intelligence

Annamaria Conti Danilo Messinese [*]

September 6, 2024

Abstract

What role does AI play in entrepreneurial decision making? We explore this question by analyzing the impact of AI predictive tools on the performance of a large sample of US startups. We exploit the random release of Google Analytics 4 (GA4) which introduced AI predictive tools especially useful for mobile app developers. Leveraging this shock in a difference-in-differences model, we find that post-GA4-release there is a boost in customer acquisition. However, the positive premium is driven by the upper tail of the treatment effect distribution, and not by marginal improvements. These effects are largest for innovative startups led by highly skilled founders. Shedding light on the mechanisms, we show that GA4 boosts the productivity of A/B testing tools. Overall, these findings suggest that AI predictive tools are useful for complementing skilled human capital in formulating new testable business hypotheses, especially relevant for the detection of breakthroughs.

Keywords: Entrepreneurship, Strategic Decision-making, Artificial Intelligence, Predictive Tools, Experimentation

[*]Conti: Department of Strategy, IE University, P.º de la Castellana, 259, 28046 Madrid, Spain, *annamaria.conti@ie.edu*. Messinese: Department of Strategy, IE University, P.º de la Castellana, 259, 28046 Madrid, Spain, *danilo.messinese@ie.edu*.

1 Introduction

While it is well established that entrepreneurs operate under uncertainty, how they should navigate this condition remains unclear (Zellweger and Zenger, 2023). This ambiguity is particularly concerning because, in the absence of guidance, entrepreneurs have been shown to take surprisingly few measures, such as researching potential competitors and their strategies, to hedge against uncertainty (Bennett and Chatterji, 2023). The consequences of this behavior are potentially profound, negatively affecting not only the performance of entrepreneurial ventures but also, more broadly, an economy's innovation, employment, and growth (Conti and Roche, 2021; Haltiwanger et al., 2013). To start filling this vacuum, recent literature advocated the role of a "scientist" entrepreneur, who applies a scientific-like method in formulating and testing hypotheses about the viability of new ideas and products (Camuffo et al., 2020, 2024; Gans et al., 2019). While this approach holds promise, it remains unclear whether entrepreneurs' time and resource constraints allow them to fully embrace this method. The relatively recent emergence of artificial intelligence (AI) predictive analytic tools might help entrepreneurs relax these constraints.

AI tools have all the characteristics of general purpose technologies, given their widespread use across industries and potential for innovation (Goldfarb et al., 2023). However, as of today, empirical evidence on their role is mixed. While studies have shown that the usage of AI predictive analytics tools is especially widespread among more innovative companies and skilled individuals (Brynjolfsson and McElheran, 2016; McElheran et al., 2024), some research indicates that these actors particularly benefit (Brynjolfsson et al., 2021), whereas other studies suggest that such tools may be especially beneficial for those at the lower end of the skill distribution (Dell'Acqua et al., 2023). There is also no consensus on whether AI predictive analytics tools help boost productivity within the innovation frontier (Wu et al., 2020), or if they also help users go beyond the frontier (Mullainathan and Rambachan, 2024), and whether they merely assist with hypothesis testing or also with the formulation of new hypotheses (Camuffo et al., 2023; Ludwig and Mullainathan, 2024). This study is the first to provide a large-scale empirical analysis of the role of AI predictive analytics tools for entrepreneurial ventures.

To guide our empirical analysis, we present a conceptual framework emphasizing AI's crucial role in anomaly detection, a key step in uncovering breakthroughs. However, pushing the innovation frontier requires more than just this function; it also needs human capital. The human ability to engage in theory-based causal reasoning enables entrepreneurs to generate new theories by mapping anomalies to relevant explanatory factors (Ehrig and Schmidt, 2022; Felin and Holweg, 2024). The combination of AI predictive tools and human capital then helps envision the "correct" model, paving the way for A/B testing tools to quantify the impact of the model's explanatory variables. We bring these conjectures to the data, investigating how the adoption of AI predictive tools affects startup customer acquisition, whether it helps boost productivity beyond the innovation frontier, and the interplay between these tools and human capital as well as their interaction with A/B testing tools.

To perform our analysis, we assemble a large dataset of startups, beginning with the full list of U.S. software-related startups listed on Crunchbase that were founded between 2016 and 2018. We merge this dataset with time-variant information about the technologies these startups use to build their websites, collected from BuiltWith. Additionally, we include data on total and mobile visits to startup websites, sourced from Semrush. The result is a comprehensive dataset of 36,835 U.S. software-related startups that integrates information on firms, technology, and performance.

We employ a difference-in-differences empirical design that takes advantage of the quasi-random timing of the release of Google Analytics 4 (GA4) in October 2020, which introduced AI predictive tools particularly useful for mobile app developers. Compared to its predecessor, Universal Analytics, GA4 comes with substantially enhanced machine learning capabilities. These include improved attribution models, more accurate anomaly detection, and the introduction of new key predictive metrics.[1] GA4 also makes it easier to track users across multiple devices, allowing for more precise measurement of the contribution of mobile visits to a company's website.[2] This feature is especially useful for mobile app developers, enabling them to enhance user experience, engagement, and retention while lowering acquisition costs.[3] After distinguishing our startups based on whether they develop apps, we assess how their website traffic—used as a proxy for consumer acquisition—changes following the launch of GA4 for mobile app developers, compared to other startups. In doing so, we saturate our model with relevant fixed effects that absorb fixed characteristics of startups and macroeconomic shocks that could bias our estimates.

The results are striking. Absent significant pre-trends, we find that in the months after the release of GA4, monthly mobile visits for app developers grow by 93% on average, relative to the control group. This increment boosts the total visits a startup's website receives each month by 84%. Not only are the effects precisely estimated and economically relevant, but they also persist over time, suggesting that the release of GA4 has had a long-lasting positive impact on consumer acquisition through mobile apps. Delving deeper into these results, we find that the positive customer acquisition premium is driven by the upper tail of the treatment effect distribution, rather than by marginal improvements. Indeed, we show that the likelihood of a startup's website receiving a number of mobile visits above the median increases by 34%, while the likelihood of falling into the 90th percentile for mobile visits increases by a considerably larger percentage of 66%.

[1] Refer to: `https://easyinsights.ai/blog/ten-ways-ga4-is-better-than-universal-analytics` and `https://www.winwithmcclatchy.com/blog/ga4-vs-ua-google-analytics`.

[2] Refer to: `https://www.narrative.bi/analytics/how-to-use-google-analytics-4-for-mobile-apps`.

[3] Refer to: `https://www.owox.com/blog/articles/guide-to-mobile-app-analytics`.

Having uncovered our main effects, we then delve into exploring their sources. By applying machine learning methods to assess the level of innovativeness of a startup's venture, we find that our main effects are primarily driven by startups at the innovation frontier. In fact, the likelihood of being in the top percentile for mobile and total visits increases by 110% and 60%, respectively, for innovative startups, compared to only 16% and 22% for those less innovative. This result challenges the common belief that AI predictive tools are more helpful for incremental innovations, as they typically extrapolate from past trends (Wu et al., 2020). Instead, our findings show that these tools help startups identify breakthroughs by efficiently detecting anomalies.

Next, we categorize founders' human capital based on their education and work experience. Our classification provides strong evidence that the availability of new AI predictive tools leads to uneven performance effects among founders, with highly skilled founders capturing the greatest benefits. Indeed, the performance premium for app startups led by highly skilled founders is at least 73%, while for those managed by less skilled founders, it is at most 38%. Connecting together innovativeness and human capital, we highlight that the strongest effects of GA4 on mobile and total visits derive from frontier startups led by highly skilled founders. Finally, to bring our exploration full circle, we uncover strong complementarities between AI predictive tools and A/B testing tools. Specifically, we show that following the release of GA4, app developers have increased their use of A/B testing tools, which complement GA4 in enhancing startup customer acquisition.

Our findings contribute to two main literature strands. First, we inform the strand on theory-based decision making in entrepreneurship (Camuffo et al., 2020, 2024; Felin et al., 2020; Felin and Zenger, 2017; Gans et al., 2019). This literature has compared entrepreneurs to scientists who form beliefs, test them, and advance by integrating the test results into their theoretical constructs. Extending this literature, we add the role of AI predictive tools and demonstrate empirically that these tools complement human cognition in generating breakthroughs, particularly in novel fields where human cognitive capabilities are more constrained. We conjecture that the underlying mechanism is AI's ability to detect anomalies, which supplements its traditional role of extrapolating trends from existing data.

Second, we contribute to the literature on the conditions explaining the adoption of AI in organizations (Agrawal et al., 2024; Bresnahan, 2019; Brynjolfsson and McElheran, 2016; McElheran et al., 2024) and the role of AI in strengthening (Brynjolfsson and McElheran, 2016; Gaessler and Piezunka, 2023; Goldfarb et al., 2023; McElheran et al., 2024) or substituting human capabilities (Acemoglu et al., 2022; Acemoglu and Restrepo, 2019). In contrast to these studies, our focus is on the role of AI predictive tools in informing and guiding scientist-like entrepreneurs to produce innovations. Challenging the view that AI predictive tools produce low-variance ideas unless properly trained (Meincke et al., 2024; Wu et al., 2020), we demonstrate that these tools help overcome human cognitive limitations, enabling founders to achieve top performance gains in novel fields. By doing so, we complement the findings of Otis et al. (2023) and Koning et al. (2022). Our large-scale analysis shows that AI not only complements founders' human capital but also that the positive performance premium from AI adoption is driven by the upper tail of the treatment effect distribution, rather than by marginal improvements. By assisting founders in formulating their business theories, AI predictive tools boost the adoption and productivity of A/B testing.

2 Theoretical Framework

Entrepreneurial decision-making is inherently fraught with uncertainty, forcing entrepreneurs to navigate conditions where outcomes are typically unknown (McDonald and Eisenhardt, 2020; McMullen and Shepherd, 2006; Ott and Eisenhardt, 2020). Coping with uncertainty requires founders to continuously engage with stakeholders and update their understanding of key elements related to products and customers as new information becomes available.

Accordingly, scholars have emphasized experimentation as a critical approach in entrepreneurial strategy (Gans et al., 2019; Koning et al., 2022; Levinthal, 2017). A well-known example of such an experimental approach is the "lean startup" methodology, which is particularly popular among software startups.[4] This methodology underscores the value of validating hypotheses through customer interactions and data collection (Osterwalder, 2010; Ries, 2011). However, this approach has been criticized for overemphasizing "learning by testing" at the expense of "learning by thinking" (Felin et al., 2020).

Building on and moving beyond the lean startup model, recent research has introduced the concept of entrepreneurs as "scientific" experimenters (Camuffo et al., 2020, 2024; Zellweger and Zenger, 2023). Unlike the lean startup, which relies on iterative testing, this newer approach emphasizes generating hypotheses based on value-creating causal theories. (Camuffo et al., 2023; Chavda et al., 2024; Ehrig and Schmidt, 2022; Felin and Zenger, 2017; Wuebker et al., 2023).

These theories provide a framework for navigating uncertainty, enabling entrepreneurs to formulate more precise hypotheses and conduct more informative experiments (Gambardella and Messinese, 2024). Theories are not just tools for understanding known scenarios, but also for envisioning unexplored opportunities through causal reasoning (Camuffo et al., 2023). This forward-looking approach allows entrepreneurs to identify and capitalize on opportunities that might be hidden within anomalies or less apparent data (Felin and Holweg, 2024).

In recent years, the increased availability of data and advances in artificial intelligence (AI) have made data-driven analysis an essential component of strategic decision-making. AI,

[4] https://hbr.org/2013/05/why-the-lean-start-up-changes-everything

defined by Agrawal et al. (2022) as a "prediction technology" is increasingly being leveraged by established firms to analyze vast amounts of data and make predictions about future events (Agrawal et al., 2018; Brynjolfsson et al., 2021; Cockburn et al., 2019). For example, AI has been shown to help firms identify and prioritize the most-valuable opportunities, allowing them to minimize false positives (Tranchero, 2023). This role is especially valuable when opportunities are clearly identifiable, as in the case of drug target identification.

Although AI's role in more established contexts is well recognized (Camuffo et al., 2022), its application to entrepreneurship and innovation is more nuanced due to the difficulties in making sense of available data. Returning to the concept of a "scientific" entrepreneur who builds and tests new theories to envision unexplored opportunities, AI could play a supportive role in this process.

As we elaborate formally in the Appendix, a theory can be likened to a statistical model, whose specification and relevant features the entrepreneur must uncover(Montiel Olea et al., 2022). Here, AI can help with 1) the selection of relevant features predicting a given outcome and 2) the choice of the statistical model that links the features to the outcome. As such, entrepreneurs using AI tools may be more likely to develop more valuable projects. Moreover, AI can help with 3) the refinement of the selected model by uncovering new and surprising information, that is, anomalies. By providing non-obvious insights, AI tools can challenge and enrich an existing model.

Elaborating on this third point, many novel AI-based software products, such as GA4, provide services that go beyond standard predictive metrics useful for model specification and estimation. For instance, GA4 introduced the *Anomaly Detection* feature that leverages machine learning algorithms to analyze historical data and establish a baseline of expected behavior. GA4's Anomaly Detection is integrated with *Explorations*, an advanced analysis tool that helps users uncover deeper insights about consumer behavior. The two tools combined allow entrepreneurs to identify deviations from the baseline and possible sources of such deviations, for example, specific user segments. As such, the usage of AI tools might lead not only to more valuable projects, on average, but to the discovery of breakthroughs.

Envisioning opportunities that extend beyond simple predictive logic requires a certain level of experience (Dew et al., 2009). Utilizing AI to develop more robust theories, harness the value in detected anomalies, and avoid false positives through selective focus on key elements is a sophisticated task. This process demands expertise and the ability to effectively *connect the dots*. In this context, AI can broaden the search beyond familiar domains, expanding the scope of opportunity identification (Tranchero, 2023). Strategizing with AI requires the capability to transcend data and engage in forward-looking theorizing (Camuffo et al., 2023; Felin and Holweg, 2024). This process necessitates experience in strategic decision-making and an aptitude for navigating complexity. As a result, the human expertise of the founders is likely to complement rather than substitute the value of AI tools.

Once AI has assisted an entrepreneur in crafting and refining a theory, the next step in the "scientific" process is empirical testing. This is where tools like A/B testing become critical. A/B testing serves as a rigorous method for validating a crafted theory, offering empirical evidence to either discard or confirm the key features of the theory. By combining A/B testing with AI, entrepreneurs can achieve high precision in their decision-making, thereby enhancing the marginal productivity of their experiments.

To summarize, AI can empower entrepreneurs to navigate uncertainty not merely by reacting to data but by proactively shaping it. The integration of AI-driven insights, founder expertise, and rigorous experimentation might generate a powerful toolkit for producing breakthrough innovations and seizing new opportunities in an ever-evolving landscape.

In the next sections, we bring our intuitions to the data.

3 Data

To build our dataset, we combine data on U.S. startups, their founders, and investors from Crunchbase with technology descriptions from LinkedIn, website technologies from BuiltWith, and website visit statistics from Semrush.

Crunchbase is an online directory that provides detailed information on a wide range of technology startups, their founders, and investors. Much of this data is entered by Crunchbase staff, with the rest being crowdsourced (Conti and Roche, 2021). Registered members can contribute information to the database, which is then reviewed by the Crunchbase team. Compared to databases like VentureXpert and VentureSource, Crunchbase offers broader coverage of technology startups, including those that have not secured venture capital.

From Crunchbase, we extract information on all recorded U.S. software-related startups[5] founded between 2017 and 2019. This amounts to 57,138 startups, for which we have data on founding dates, industry group keywords, locations, financing rounds, and participating investors.

To precisely identify startups developing apps, we use publicly available information from LinkedIn on the description of startups' businesses and technologies. This information is accessible for 94% of the startups. We do so because, by inspection, founders often list the keyword "app" on their Crunchbase profile to describe their technologies even when developing apps is not their core business. According to these descriptions, 6% of our software startups develop apps as their core business. In robustness checks, we either include in the app developers category the additional 7% of startups described by Crunchbase as app developers, or use Crunchbase's "app" group keyword to identify app developers.

[5]Software-related startups are startups defined by the following Crunchbase industry group keywords: advertising, analytics, apps, artificial intelligence, B2B, B2C, blockchain, collaboration, commerce and shopping, community and lifestyle, content and publishing, database, design, digital entertainment, financial services, gaming, ICT, information technology, internet, messaging and telecommunications, media and entertainment, mobile, payments, platforms, privacy and security, ride sharing, sales and marketing, search engine, social network, software, and video.

We match our dataset with information on the technology stacks utilized by startups' websites, obtained from BuiltWith, using the startups' website domains as the matching key. BuiltWith is a technology profiler whose data has been increasingly utilized by the academic community. It offers fine-grained, time-stamped data covering over 25,000 web technologies (for example, analytics, A/B testing, advertising, hosting, and content management system) (Koning et al., 2022; Roche et al., 2024; Stroube and Dushnitsky, 2023). By implementing our matching procedure, we were able to find information for approximately 93% of the startups in our dataset.

Finally, we merged data from Crunchbase and BuiltWith with information on the visits startup websites receive. This data is provided by Semrush, a SaaS platform offering detailed time-stamped data on total, desktop, and mobile visits companies' websites receive. Relying on third-party providers, Semrush collects clickstream data, which records an individual's clicks through their internet journey, from millions of internet users who agree to share their data.[6] Recently, Semrush entered into a partnership with Crunchbase to enhance the accuracy of website tracking for Crunchbase startups. This highlights the growing importance of monitoring website visits as a metric for gauging a startup's technology.

By merging all these datasets together, we arrive at a final sample of 36,835 U.S. startups developing software technologies. These startups were founded during the years 2017-2019 and were observed monthly from the year they were founded until November 2023. Descriptive statistics are provided in Table 1. As shown, Approximately 28% of the startups had raised a financing round as of the end of our sample period, November 2023. On average, the number of website technologies used by a startup by November 2023 is at most 93, 50% had used GA4, while 37% had used A/B testing tools.

To measure how innovative a startup is, we built an innovation score on a scale from 0

[6]Please refer to: https://www.semrush.com/blog/what-is-clickstream-data/.

(least innovative) to 100 (most innovative) using machine learning, including input from Gemini. In essence, we used a training dataset and asked Gemini to assign an innovation score to each of the startups in the dataset. We then used a combination of Term Frequency-Inverse Document Frequency (TF-IDF) and a stochastic regression model (SGDRegressor) to predict the level of venture novelty. As reported, the mean of the score is 55.6. Based on this measure, we consider startups above the median of the score distribution to be innovative.

Employing data on founder biographies and their education degrees available from Crunchbase, we built a measure of ”top” founder human capital. In practice, we identified highly-skilled founders as individuals with backgrounds such as serial entrepreneurship, experience in top consulting firms like McKinsey and Bain & Company, academic positions, scientific expertise, and education degrees from renowned schools in the US and internationally. According to this classification, 13% percent of the startups in our dataset are led by highly-skilled founders. Finally, and perhaps not surprisingly, Table 1 shows that over 43% of the startups in our dataset are concentrated in California, Massachusetts, and New York.

Table 2 presents descriptive statistics on website visits for startups, distinguishing between those categorized as app developers and others. In 2020, just before the release of GA4, mobile developer websites received fewer visits on average compared to other startups. However, the differences are not statistically significant.

⟨ Insert Table 1 and Table 2 about here ⟩

4 Empirical Methodology

To assess the impact of AI predictive tools on startup performance, we employ a difference-in-differences regression design, evaluating how a startup's performance changes after the release of GA4, distinguishing between startups more or less at risk of adopting GA4. This tool was first introduced in mid-October 2020. As previously discussed, GA4 significantly advances the machine learning capabilities of its predecessor, Universal Analytics. For example, it refined the attribution model by utilizing AI to distribute credit among various points in a user's journey instead of giving all the credit to the last-clicked advertisement. It also enhanced the algorithms used for anomaly detection, identifying atypical patterns in the data that could signal key insights. Furthermore, GA4 added new metrics, such as *Purchase Probability*, *Churn Probability*, and *Predicted Revenue*, which enable companies to more accurately predict customer behavior.

Remarkably, by moving to an *event*-based model where every interaction of a user is captured as a distinct event[7], GA4 made it easier to track users across multiple devices. As a result, it allowed for a more precise measurement of the contribution of mobile visits to a company's website, increasing the expected value of GA4 for app developers. Building on this discussion, we identify app developers as startups relatively more at risk of adopting GA4 relative to other software startups. This conjecture is confirmed by our data. Indeed, Figure 1 illustrates that following the introduction of GA4, the likelihood of adopting GA4 increases both among mobile developers and other software startups. However, the surge is more pronounced among the former category of startups. App developers have become more likely to adopt G4 compared to other startups starting from March 2021, the date at which the new features of GA4 became available for mobile tracking.

⟨ Insert Figure 1 about here ⟩

The main startup outcomes we analyze are total mobile visits to the startups' websites, which we consider proxies for customer acquisition. In alternative analyses presented in the Appendix, we also examine users and user retention. As our outcomes are count variables that only take positive and integer, we estimate the following difference-in-differences Poisson Quasi-Maximum Likelihood (PQML) model at the level of startup i observed in

[7]https://support.google.com/analytics/answer/9846734?hl=en.

year-month t:

$$E(Y_{it}|X_{it}) = \gamma_i \exp(\beta \cdot DevelopApps_i \cdot I[t > October\ 2020] + \delta \cdot TechStack_{it} + \tau_t). \quad (1)$$

$DevelopApps_i$ is an indicator equal to 1 if startup i develops apps. The coefficient of interest in this model is β. It represents the change in customer acquisition for app developers following the launch of GA4 in October 2020, replacing Universal Analytics, in comparison to other software startups. Note that while app developers use GA4 more intensively than other software startups, the latter group still adopts GA4. As a result, the findings we discuss in the next section represent lower bound estimates of the true effects.

$TechStack_{it}$ is the cumulative count of the technologies a startup uses to build its website, other than analytics tools. We use this measure as a proxy for the technology requirements of a startup's website. The γ_i are startup fixed effects that absorb fixed differences across startups, while τ_t are year-month fixed effects, which absorb the effects of possible macroeconomic shocks.

In alternative specifications, we will evaluate whether the release of GA4 allowed mobile startups to produce breakthroughs or whether it helped introduce marginal technology improvements. For this purpose, we will modify Eq. 1 estimating instead linear probability models for whether a startup's number of website visits received falls in the 90th percentile, serving as a proxy for breakthroughs, or if the startup attained a number of website visits above the median, serving as a proxy for marginal technology improvements.

The descriptive statistics reported in Figure 2 offer a glance into the phenomenon we plan to investigate. As displayed in Panel A, during the 12 months preceding the release of GA, the average number of website visits followed the same trend regardless of whether startups developed apps. However, after the launch of GA4, app developers began to outperform other software startups after month 7. These patterns are more accentuated in

Panel B, where we examine the likelihood of being in the top percentile for the number of total website visits received.

⟨ Insert Figure 2 about here ⟩

5 Results

5.1 Baseline Results

We report the results from the PQML model in Eq. 1 for the monthly number of website visits a startup received in Table 3. We cluster standard errors by startup. As shown in column 1, after the launch of GA4, the hazard that an app developer's website receives an additional visit each month from a mobile device increases by 0.658 compared to the control group. This coefficient corresponds to a 93% increase in the monthly number of visits from mobile devices. In column 2, we show that such a boost in the number of mobile visits leads to an increase in the total number of monthly visits by 84%.[8]

To inspect the presence of pre-trends, Figure 3 reports the results from a dynamic version of Eq. 1 that replaces the $I[t > October\ 2020]$ term with time indicators for each year-month to/from the release of GA4. The reported coefficients represent the additional proportion of new visits that the websites of app developers receive each month relative to the other software startups. Panel A reports the effects of the GA4 launch on mobile visits, while Panel B reports the effects on total visits. We highlight three main patterns. First, we observe no pre-trends regardless of the outcome examined, suggesting that the number of mobile (total) website visits by app developers and other software startups would have followed a similar trend in the absence of the GA4 launch. Second, the increase in website visits accelerates starting from the fifth month after the release of GA4, which is in line with

[8] The number of observations reported at the end of Table 3 is lower than the total number of observations because the Stata command *ppmlhdfe* discards singleton observations by default. However, when we force *ppmlhdfe* to retain these observations, the results (reported in Table A1) do not change. Moreover, as we report in Table A2, these results are not driven by a few outliers, as they remain robust after winsorizing the top 1% of observations of the number of visits.

the fact that the full release of GA4 features for mobile app tracking occurred in March 2021. Third, the effects of the GA4 release remain positive and significant thereafter.

In columns 3 and 4 of Table 3, we repeat the same exercise, having restricted the sample to startups that raised at least one financing round from founding until the end of 2023. With this exercise, we aim to assess whether the potential adoption of AI predictive tools could have helped app developers not only survive but also grow. As shown, the effects on app developers' mobile (total) visits of the GA4 release are no longer statistically significant. Overall, these results suggest that, on average, the availability of AI predictive tools might have helped the treated startups achieve enough website visits to stay afloat, but not enough to grow with the support of investor capital.

⟨ Insert Table 3 and Figure 3 about here ⟩

5.2 Impact of AI predictive tools on the tails of website visits distribution

Having assessed the average effect of the availability of new AI predictive tools, we next examine their impact on the tails of the website visits distribution. To do so, we estimate linear probability models for the likelihood that a startup's number of visits is above the median (Panel A of Table 4) and above the 90th percentile cutoff (Panel B of Table 4). The results highlight a remarkable pattern. While the effects—especially on website visits from mobile devices—are positive and robust regardless of the cutoff chosen, they are stronger when we examine the 90th percentile cutoff.

As reported in columns 1 and 2, post-GA4 release, app developers become 0.0189 (0.0136) more likely to be above the median of the number of mobile (total) visits received, equivalent to a 34% (8%) increase relative to the outcome mean. However, they become 0.007 (0.014) more likely to be in the 90th percentile, equivalent to a 62% (43%) increase in the outcome mean. We observe similar patterns in columns 3 and 4, where we restrict the sample to startups that raised at least one funding round during the period we observe.

These findings are confirmed by the event studies reported in Figure 4. We continue to observe no significant pre-trends. Moreover, the effects of the GA4 release are considerably stronger when we examine the 90th percentile cutoff. Overall, these results suggest that the availability of AI predictive tools helped startups boost customer acquisition. However, the positive premium is driven by the upper tail of the treatment effect distribution, and not by marginal improvements. This provides an indication that AI predictive tools may help startups achieve breakthroughs rather than incremental innovations.

⟨ Insert Table 4 and Figure 4 about here ⟩

To verify the robustness of these results, we first show in Table A3 that our results are not driven by a few outliers, as they remain robust after winsorizing the top 1% observations of the number of visits. Additionally, our results are not an artifact of the econometric methodology we employed. To verify this, in Table A4, we estimated IV models to assess the impact of the *actual* adoption of GA4 on the likelihood of a startup being in the top percentile for the number of mobile (total) visits its website received. We apply this analysis to the full set of U.S. software startups that were founded after the launch of GA4, extending the sample to startups founded up to December 2023. Specifically, we instrumented the adoption of GA4 by startups -measured with a (0/1) indicator that takes value 1 the months GA4 is being used- with a Bartik instrument (Bartik, 1991), which isolates treatment variation due to the differential impact of common shocks on startups with distinct predetermined exposures. The idea of this instrument is to purge the treatment variation of possibly confounding factors varying across units over time (Breuer, 2022). In practice, we exploited the fact that nation-wide shocks in the adoption of GA4 impact cities' adoption differently, depending on their pre-existing industrial structure. This subpart of variation in cities' usage of GA4 is less likely to reflect changes due to local technology shocks, growth opportunities, or financing conditions. To build the instrument, we computed the predetermined usage of GA4 in 2020 (prior to the release of GA4) by sector and U.S. city. We then calculated the inner product of the predetermined city-industry share of GA4 usage in 2020 and the time-varying U.S.-wide industry share. As a result of this estimation strategy, we observe a strong impact of the actual usage of GA4 on both mobile and total website visits received.

To further assess the robustness of our results, we show in tables A5 and A6 that the findings are not sensitive to the definition of app developers we employ. In fact, whether we use a broader definition of app developers that includes startups identified by Crunchbase's 'app' group keyword, or we only use Crunchbase's "app" group keyword to identify app developers, the results remain very similar in both the significance and magnitude of the coefficients. These last results are noteworthy because by analyzing an extraction of the Crunchbase database from 2018, we could confirm that the industry group keywords used by startups to describe their technologies remained largely unchanged. This indicates that startups did not alter their descriptions even after new AI predictive tools became available.

Finally, to verify that our findings are not specific to the kids of outcomes observed, we assess the effect of GA4 tools on the number of visits a startup's website receives, *weighted* by one minus the bounce rate. As the bounce rate is the percentage of visitors who enter a website and then leave rather than continue viewing the other pages, this weighted measure is a proxy for website retention. The results reported in Table A7 show that the availability of AI predictive tools helped startups boost customer retention. However, the positive premium continues to be driven by the upper tail of the treatment effect distribution, and not by marginal improvements. Finally, we observe a very similar pattern in Table A8, where we examine the likelihood that a startup is above the median (Panel A) and in the top percentile (Panel B) for the number of users visiting its website.

5.3 Distinguishing Based on the Innovativeness of a Startup's Venture

After identifying our main effects, we differentiate startups based on the innovativeness of their ventures to evaluate how the effects may vary between more and less innovative startups. As we mentioned in Section 3, we employed machine learning methods and input from Gemini to identify the innovativeness of a startup, building an innovation score on a scale from 0 (least innovative) to 100 (most innovative). Based on this score, we identify startups undertaking innovative projects as those above the median of the score distribution. The results using this cutoff are reported in Table 5. In columns 1 and 2, we focus on innovative startups, while in columns 3 and 4, we examine startups undertaking more traditional projects.

As shown, the effects we uncovered so far are considerably stronger for the subsample of innovative startups. In this subsample, app developers become 110% (60%) more likely to be in the 90th percentile of the number of mobile (total) visits after the release of GA4. Conversely, in the subsample of less innovative startups, app developers become 16% (22%) more likely to be in the 90th percentile of the number of mobile (total) visits after the release of GA4.

These differences in effects are confirmed in Table A9, where instead of performing a split-the-sample analysis, we augment our estimated equation with a triple interaction between $DevelopApps_i$, $I[t > October\ 2020]$, and the indicator identifying innovative startups. Additionally, Table A9 shows that these differences persist when we build the same innovation score using ChatGPT rather than Gemini. Finally, these differences emerge clearly in the event studies reported in Figure 5. Here, we show that the effects associated with the months succeeding the release of GA4 are stronger in the sample of innovative startups (Panels A and C) than in the sample of startups pursuing more traditional projects (Panels B and D). Referring back to our conceptual framework, these results suggest that AI predictive tools are particularly beneficial for startups undertaking innovative projects, as they could assist founders in identifying anomalies that may pave the way to breakthroughs.

⟨ Insert Table 5 and Figure 5 about here ⟩

5.4 Distinguishing By the Human Capital of Startup Founders

In this section, we differentiate startups based on how skilled their founders are. We identify highly-skilled founders (”founders with top expertise”) as individuals with backgrounds such as serial entrepreneurship, experience in top consulting firms like McKinsey and Bain & Company, academic positions, scientific expertise, and education degrees from renowned schools in the US and internationally. The split-the-sample analysis using this distinction is reported in Table 6. In columns 1 and 2, we focus on startups led by founders with top expertise, while in columns 3 and 4, we examine the remaining startups.

The effects of the availability of new AI predictive tools are stronger in the subsample of startups managed by founders with top expertise. Here, app developers become 147% (73%) more likely to be in the 90th percentile of the number of mobile (total) visits after the release of GA4. In contrast, in the subsample of startups led by less skilled founders, app developers become 38% (26%) more likely to be in the 90th percentile of the number of mobile (total) visits after the release of GA4.

These results are confirmed in Table A10, where we add to our estimated equation a triple interaction between $DevelopApps_i$, $I[t > October\ 2020]$, and an indicator identifying startups led by founders with top expertise. Here, we show that startups developing apps become 1.96 (3.5) percentage points more likely to be in the 90th percentile of mobile (total) visits after the release of GA4 relative to other software startups. The event studies reported in Figure 6 support and strengthen these results. Taken together, our findings suggest that extracting value from AI predictive tools necessitates significant cognitive effort and superior human capital. Consequently, the benefits of adopting new AI predictive tools are uneven, disproportionately favoring founders with top expertise.

⟨ Insert Table 6 and Figure 6 about here ⟩

5.5 Distinguishing By the Innovativeness of Startup Ventures and the Human Capital of Startup Founders

To complete the analysis, Table 7 presents a 2-by-2 matrix displaying the outcomes obtained from estimating a linear probability model for the likelihood of a startup being in the 90th percentile of the total number of visits across four subsamples. These subsamples are defined by the innovativeness of a startup's venture and the expertise of its founders.

Regardless of the outcome examined, we demonstrate that the most pronounced effects of the availability of new AI predictive tools are observed in the subsample of innovative startups led by founders with top expertise (top left). In this group, we observe a 88% increase in the likelihood of being in the 90th percentile in terms of the number of visits, which stands in stark contrast to the 9% premium observed in the subsample of startups developing more traditional technologies and led by founders with less expertise (bottom right). In the other quadrants, the premium observed is at most 48%. Similar premiums are observed in Table A11, where we repeat the same analysis as described earlier, but this time focusing on the number of mobile visits.

⟨ Insert Table 7 about here ⟩

5.6 AI Predictive Tools and A/B Testing

To bring our exploration full circle, we examine the relationship between AI predictive tools and A/B testing. As mentioned in Section 2, by providing predictions about covariates that could influence startup performance, AI predictive tools can enhance the marginal productivity of A/B testing tools, which are primarily used for hypothesis testing.

We begin our investigation by showing in column 1 of Table 8 that after the release of GA4, mobile startups significantly increased their utilization of A/B testing tools. This result is reassuring, as it suggests that the potential effects of A/B testing tools may not be confounded by the adoption of GA4.

In columns 2 to 5, we assess whether there exists any relationship between GA4 and the utilization of A/B testing tools in driving visits to the startups' websites. Columns 2 and 3 report the results of a PQML model for the number of mobile and total visits a startup's website receives. Here, we detect no significant relationship between GA4 and A/B testing tools in the number of mobile and total visits received by a startup's website in any given month. However, the linear probability models presented in columns 4 and 5 demonstrate that following the release of GA4, app developers that have utilized A/B testing are 1.5 (2.5) percentage points more likely to be in the 90th percentile of mobile (total) visits.

Clearly, the decision to utilize A/B testing tools is an endogenous one and, therefore, the results presented in columns 4 and 5 should be interpreted as correlations. Despite this, the event studies shown in Figure 7 offer some reassurance. In fact, they reveal flat pre-trends prior to the release of GA4 (Panel A and B), and a significant rise in the probability of a startup being in the last percentile of the number of mobile visits for startups that have employed A/B testing only after GA4 was launched (Panel A). In Panel B, the positive impact of the GA4 release on the likelihood that a startup is in the last percentile of the total number of visits is not as prominent.

⟨ Insert Table 8 and Figure 7 about here ⟩

To go beyond the correlations just presented, we estimate an instrumental variables (IV) model applied to the full set of U.S. software startups that were founded after the launch of GA4, extending the sample to startups founded up to December 2023. Within this sample, we assess whether app developers—which we found to benefit particularly from GA4—might have experienced differential improvements in customer acquisition from adopting A/B testing tools compared to other software startups. We instrument for startup usage of A/B testing tools with a Bartik instrument isolating treatment variation due to the differential impact of common shocks on startups with distinct predetermined exposures. Again, the rationale of this instrument is to purge the treatment variation of possibly confounding factors varying across units over time. In particular, we exploit the fact that nation-wide shocks in the adoption of A/B testing tools impact cities' adoption differently, depending on their pre-existing industrial structure. This subpart of variation in cities' usage of A/B testing tools is less likely to reflect changes due to local technology shocks, growth opportunities, or financing conditions. To build the instrument, we computed the predetermined usage of A/B testing tools in 2020 (prior to the release of GA4) by sector and U.S. city. We then calculated the inner product of the predetermined city-industry share of A/B testing usage in 2020 and the time-varying U.S.-wide industry share. The results are displayed in Table A12. In columns 1 and 2, we report the results from a naive linear probability model, which suggests a complementarity between the adoption of A/B testing tools and GA4 in increasing the likelihood of a startup being in the 90th percentile for the number of mobile (total) visits received. In columns 3 and 4, we report the IV estimates showing that app developers benefit more from using A/B testing tools than other software startups. The first stages are quite strong. In fact, the F tests of excluded instruments take values 15 and 313 in the first stages for A/B, and the interaction between A/B and the indicator identifying app developers, respectively. Overall, our results underscore the complementary relationship between AI predictive tools and A/B testing within the framework of entrepreneurial decision making.

⟨ Insert Table 9 about here ⟩

6 Concluding Remarks

What role does AI play in entrepreneurial decision making? Implementing a difference-in-differences empirical approach exploiting the quasi-random release of GA4 in a large dataset of U.S. software startups, this paper uncovers that AI predictive tools are especially effective in achieving breakthroughs and less so in attaining average performance gains. This is particularly the case when these tools are used by skilled founders and deployed to develop novel projects. AI does not just help startups survive, but conditional on survival, it helps achieve customer acquisition gains in the upper tail of the distribution.

Taken together, our findings suggest that AI predictive tools aid in the development of more valuable ideas by detecting causal patterns that existing statistical models overlook. This improves the marginal productivity of A/B testing tools, which focus on enhancing the precision of pre-formulated theories.

This paper advances and offers implications for existing entrepreneurial literature. While the common perception so far has been that AI predictive tools help entrepreneurs develop low-variance, incremental innovations by extrapolating trends from historical data, our study offers a different perspective of AI predictive tools. Although the role of these tools is clearly non-democratic, as they particularly complement skilled human capital, they assist founders in achieving breakthroughs, especially in risky settings where data is either not abundantly available or is less obvious to interpret relying solely on accumulated experience. These breakthroughs have the potential to considerably boost an economy's innovation, employment, and growth.

References

Acemoglu, D., Autor, D., Hazell, J., and Restrepo, P. (2022). "Artificial intelligence and jobs: Evidence from online vacancies." *Journal of Labor Economics*, *40*(S1), S293–S340.

Acemoglu, D., and Restrepo, P. (2019). "Automation and new tasks: How technology displaces and reinstates labor." *Journal of Economic Perspectives*, *33*(2), 3–30.

Agrawal, A., Gans, J., and Goldfarb, A. (2018). "Prediction, judgment, and complexity: a theory of decision-making and artificial intelligence." In *The economics of artificial intelligence: An agenda*, 89–110, University of Chicago Press.

Agrawal, A., Gans, J., and Goldfarb, A. (2022). *Prediction Machines, Updated and Expanded: The Simple Economics of Artificial Intelligence*. Harvard Business Press.

Agrawal, A., Gans, J. S., and Goldfarb, A. (2024). "Artificial intelligence adoption and system-wide change." *Journal of Economics & Management Strategy*, *33*(2), 327–337.

Bartik, T. J. (1991). "Who benefits from state and local economic development policies?"

Bennett, V. M., and Chatterji, A. K. (2023). "The entrepreneurial process: Evidence from a nationally representative survey." *Strategic Management Journal*, *44*(1), 86–116.

Bresnahan, T. (2019). "Artificial intelligence technologies and aggregate growth prospects." *Prospects for economic growth in the United States*, 132–172.

Breuer, M. (2022). "Bartik instruments: An applied introduction." *Journal of Financial Reporting*, *7*(1), 49–67.

Brynjolfsson, E., Jin, W., and McElheran, K. (2021). "The power of prediction: predictive analytics, workplace complements, and business performance." *Business Economics*, *56*, 217–239.

Brynjolfsson, E., and McElheran, K. (2016). "The rapid adoption of data-driven decision-making." *American Economic Review*, *106*(5), 133–139.

Camuffo, A., Cordova, A., Gambardella, A., and Spina, C. (2020). "A scientific approach to entrepreneurial decision making: Evidence from a randomized control trial." *Management Science*, *66*(2), 564–586.

Camuffo, A., Gambardella, A., Maccheroni, F., Marinacci, M., and Pignataro, A. (2022). "Microfoundations of low-frequency high-impact decisions." Tech. rep., CEPR Discussion Paper No. DP17392.

Camuffo, A., Gambardella, A., Messinese, D., Novelli, E., Paolucci, E., and Spina, C. (2024). "A scientific approach to entrepreneurial decision-making: Large-scale replication and extension." *Strategic Management Journal*.

Camuffo, A., Gambardella, A., and Pignataro, A. (2023). "Theory-driven strategic management decisions." *CEPR DP 17664v2*.

Chavda, A., Gans, J. S., and Stern, S. (2024). "Theory-driven entrepreneurial search." *Available at SSRN 4706860*.

Cockburn, I. M., Henderson, R., and Stern, S. (2019). "4. the impact of artificial intelligence on innovation: An exploratory analysis." In *The economics of artificial intelligence*, 115–148, University of Chicago Press.

Conti, A., and Roche, M. P. (2021). "Lowering the bar? external conditions, opportunity costs, and high-tech start-up outcomes." *Organization Science*, *32*(4), 965–986.

Dell'Acqua, F., McFowland, E., Mollick, E. R., Lifshitz-Assaf, H., Kellogg, K., Rajendran, S., Krayer, L., Candelon, F., and Lakhani, K. R. (2023). "Navigating the jagged technological frontier: Field experimental evidence of the effects of ai on knowledge worker productivity and quality." *Harvard Business School Technology & Operations Mgt. Unit Working Paper*, (24-013).

Dew, N., Read, S., Sarasvathy, S. D., and Wiltbank, R. (2009). "Effectual versus predictive logics in entrepreneurial decision-making: Differences between experts and novices." *Journal of Business Venturing*, *24*(4), 287–309.

Ehrig, T., and Schmidt, J. (2022). "Theory-based learning and experimentation: How strategists can systematically generate knowledge at the edge between the known and the unknown." *Strategic Management Journal*, *43*(7), 1287–1318.

Felin, T., Gambardella, A., Stern, S., and Zenger, T. (2020). "Lean startup and the business model: Experimentation revisited." *Long Range Planning*, *53*(4), 101889.

Felin, T., and Holweg, M. (2024). "Theory is all you need: Ai, human cognition, and decision making." *Human Cognition, and Decision Making (February 23, 2024)*.

Felin, T., and Zenger, T. R. (2017). "The theory-based view: Economic actors as theorists." *Strategy Science*, *2*(4), 258–271.

Gaessler, F., and Piezunka, H. (2023). "Training with ai: Evidence from chess computers." *Strategic Management Journal*, *44*(11), 2724–2750.

Gambardella, A., and Messinese, D. (2024). "Design-based and theory-based approaches to strategic decisions.", working paper.

Gans, J. S., Stern, S., and Wu, J. (2019). "Foundations of entrepreneurial strategy." *Strategic Management Journal*, *40*(5), 736–756.

Goldfarb, A., Taska, B., and Teodoridis, F. (2023). "Could machine learning be a general purpose technology? a comparison of emerging technologies using data from online job postings." *Research Policy*, *52*(1), 104653.

Haltiwanger, J., Jarmin, R. S., and Miranda, J. (2013). "Who creates jobs? small versus large versus young." *Review of Economics and Statistics*, *95*(2), 347–361.

Koning, R., Hasan, S., and Chatterji, A. (2022). "Experimentation and start-up performance: Evidence from a/b testing." *Management Science*, *68*(9), 6434–6453.

Levinthal, D. A. (2017). "Mendel in the c-suite: Design and the evolution of strategies." *Strategy Science*, *2*(4), 282–287.

Ludwig, J., and Mullainathan, S. (2024). "Machine learning as a tool for hypothesis generation." *The Quarterly Journal of Economics*, *139*(2), 751–827.

McDonald, R. M., and Eisenhardt, K. M. (2020). "Parallel play: Startups, nascent markets, and effective business-model design." *Administrative Science Quarterly*, *65*(2), 483–523.

McElheran, K., Li, J. F., Brynjolfsson, E., Kroff, Z., Dinlersoz, E., Foster, L., and Zolas, N. (2024). "Ai adoption in america: Who, what, and where." *Journal of Economics & Management Strategy*.

McMullen, J. S., and Shepherd, D. A. (2006). "Entrepreneurial action and the role of uncertainty in the theory of the entrepreneur." *Academy of Management review*, *31*(1), 132–152.

Meincke, L., Mollick, E. R., and Terwiesch, C. (2024). "Prompting diverse ideas: Increasing ai idea variance." *arXiv preprint arXiv:2402.01727*.

Montiel Olea, J. L., Ortoleva, P., Pai, M. M., and Prat, A. (2022). "Competing models." *The Quarterly Journal of Economics*, *137*(4), 2419–2457.

Mullainathan, S., and Rambachan, A. (2024). "From predictive algorithms to automatic generation of anomalies." Tech. rep., National Bureau of Economic Research.

Osterwalder, A. (2010). "Business model generation: A handbook for visionaries, game changers, and challengers."

Otis, N., Clarke, R. P., Delecourt, S., Holtz, D., and Koning, R. (2023). "The uneven impact of generative ai on entrepreneurial performance." *Available at SSRN 4671369*.

Ott, T. E., and Eisenhardt, K. M. (2020). "Decision weaving: Forming novel, complex strategy in entrepreneurial settings." *Strategic Management Journal*, *41*(12), 2275–2314.

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

Roche, M. P., Oettl, A., and Catalini, C. (2024). "Proximate (co-) working: Knowledge spillovers and social interactions." *Management Science*.

Stroube, B., and Dushnitsky, G. (2023). "Built with builtwith: A technical note." *Available at SSRN*.

Tranchero, M. (2023). "Finding diamonds in the rough: Data-driven opportunities and pharmaceutical innovation.", working Paper.

Wu, L., Hitt, L., and Lou, B. (2020). "Data analytics, innovation, and firm productivity." *Management Science*, *66*(5), 2017–2039.

Wuebker, R., Zenger, T., and Felin, T. (2023). "The theory-based view: Entrepreneurial microfoundations, resources, and choices." *Strategic Management Journal*, *44*(12), 2922–2949.

Zellweger, T., and Zenger, T. (2023). "Entrepreneurs as scientists: A pragmatist approach to producing value out of uncertainty." *Academy of Management Review*, *48*(3), 379–408.

Figures

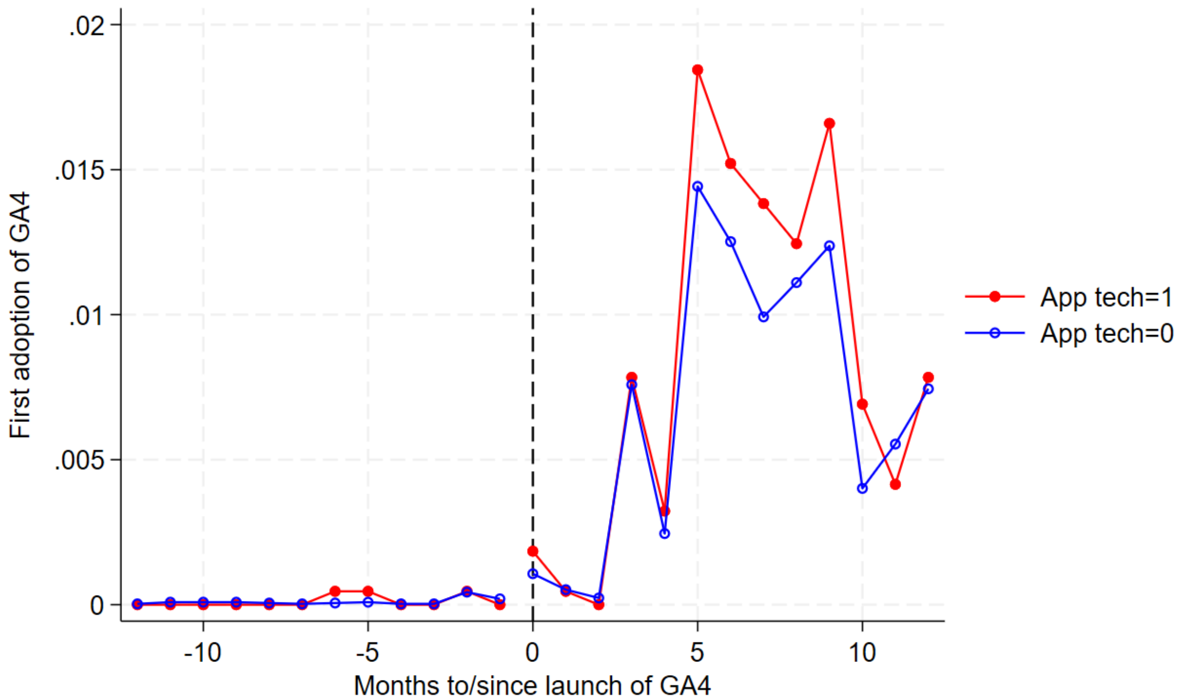

Figure 1: Adoption of Google Analytics 4 Over Time

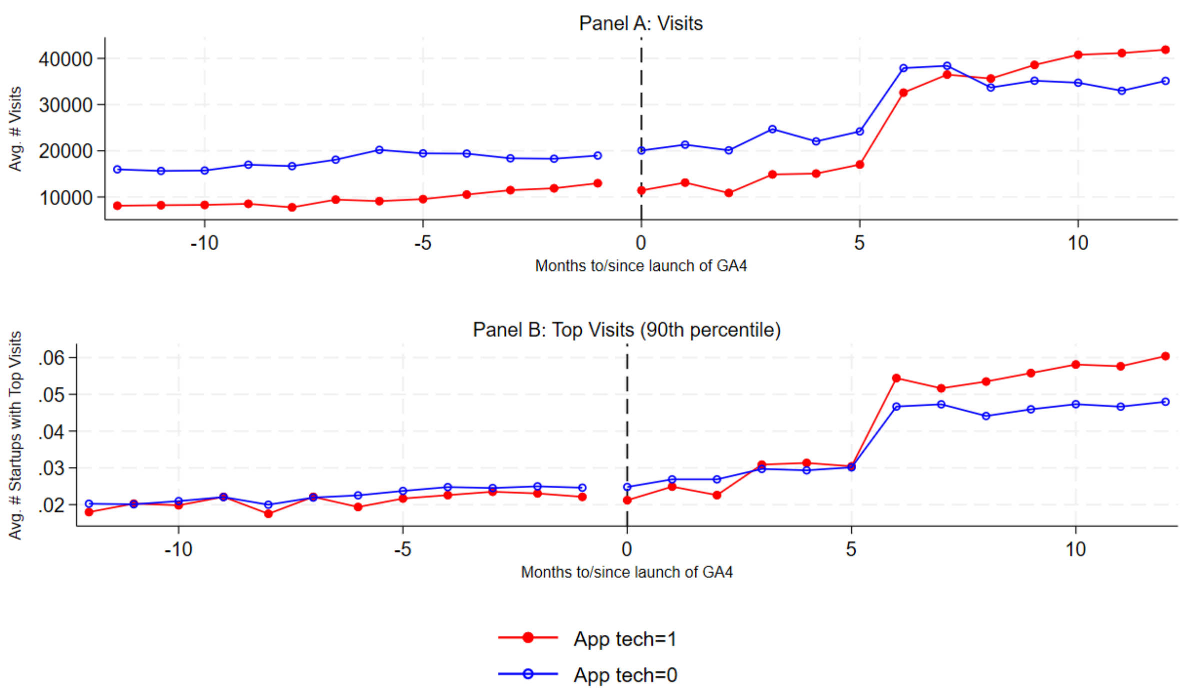

Figure 2: Website Visits Over Time

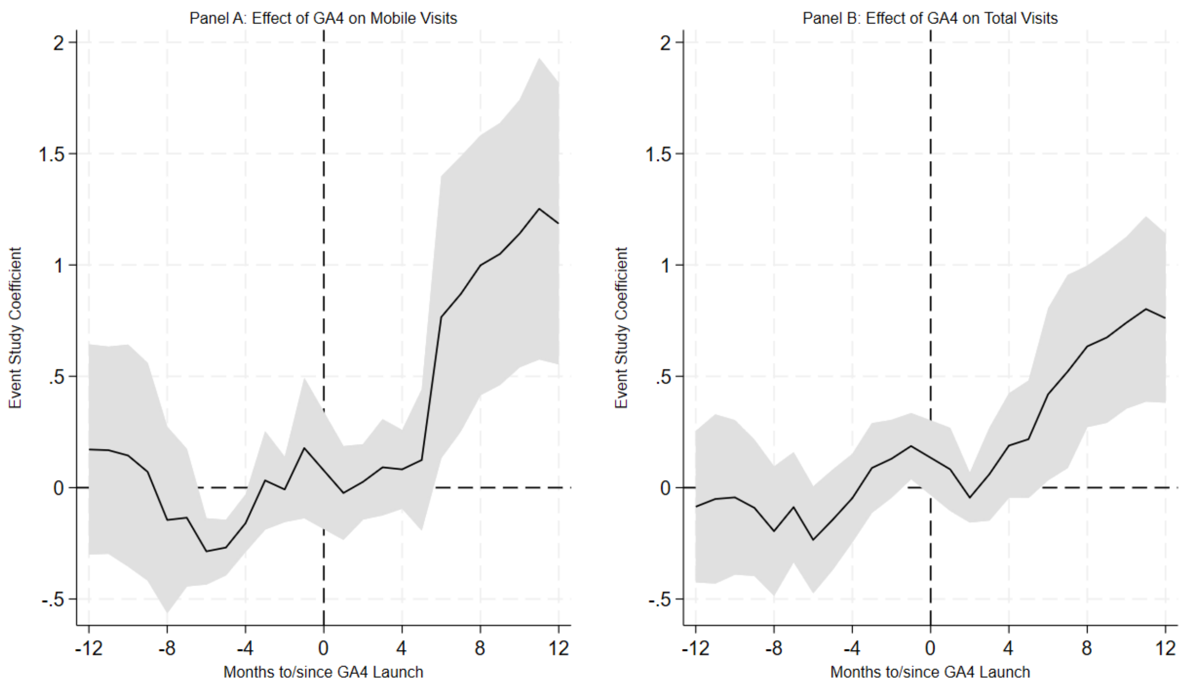

Figure 3: Event Studies for the Number of Total (Mobile) Visits

In this figure, we estimate a dynamic version of Eq. (1) that replaces $PostGA4$ with time indicators for each year-month to/from the release of Google Analytics 4. In Panel A, we examine the number of mobile visits a startup's website receives in month t as an outcome. In Panel B, the outcome is the total number of visits a startup's website receives in month t. Startup and year-month fixed effects are included, and standard errors are clustered at the startup level. The shaded area represents the 95 percent confidence interval.

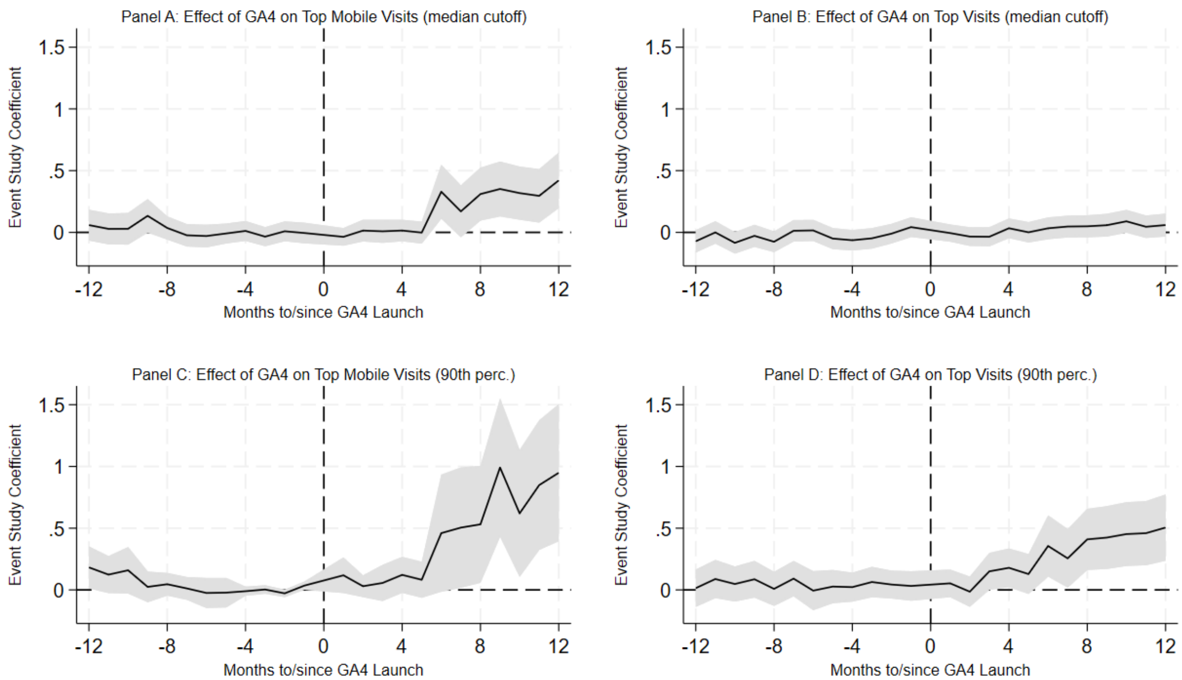

Figure 4: Event Studies for Top Total (Mobile) Visits

In this figure, we estimate a dynamic version of Eq. (1) that replaces $PostGA4$ with time indicators for each year-month to/from the release of Google Analytics 4. In Panel A (Panel B), startups with top (top mobile) visits are those with a number of total (mobile) monthly visits above the median. In Panel C (Panel D), startups with top (top mobile) visits are those with a number of total (mobile) monthly visits above the 90th percentile cutoff. Startup and year-month fixed effects are included, and standard errors are clustered at the startup level. The shaded area represents the 95 percent confidence interval.

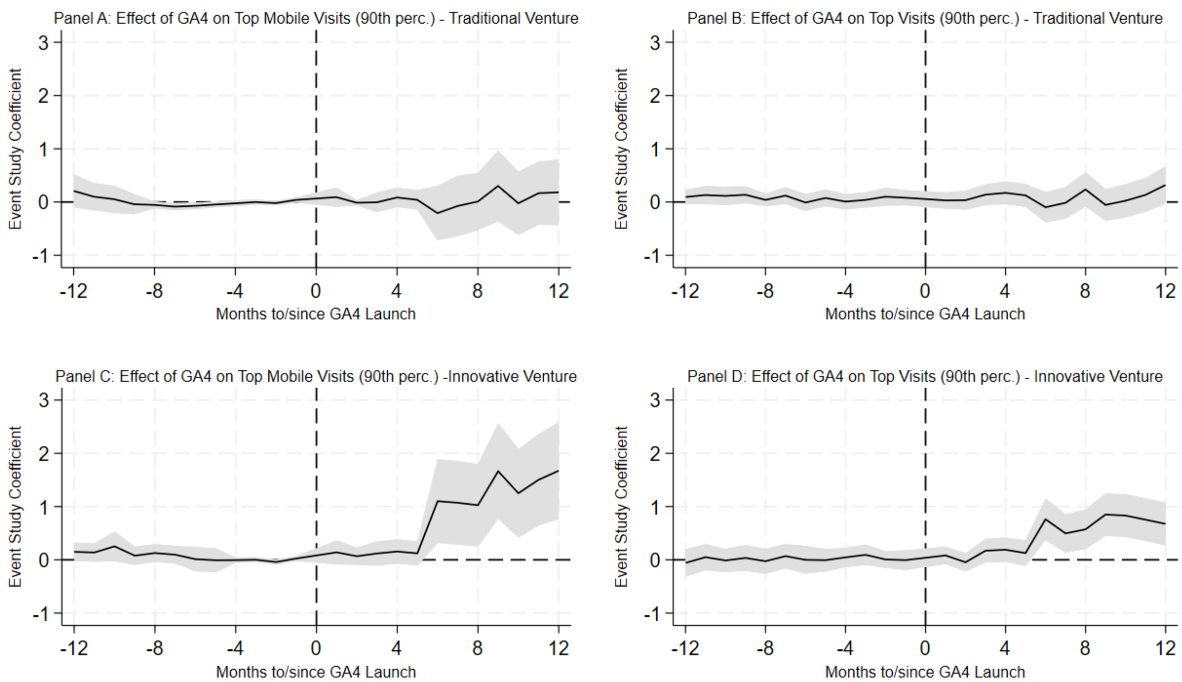

Figure 5: Event Studies for Top Total (Mobile) Visits: By the Innovativeness of a Startup's Venture
In this figure, we estimate a dynamic version of Eq. (1) that replaces $PostGA4$ with time indicators for each year-month to/from the release of Google Analytics 4. In Panels A and B, we examine traditional ventures. In Panels C and D, we examine innovative startups. The outcome is the likelihood that a startup is in the top percentile for the number of mobile (total) visits its website receives. Startup and year-month fixed effects are included, and standard errors are clustered at the startup level. The shaded area represents the 95 percent confidence interval.

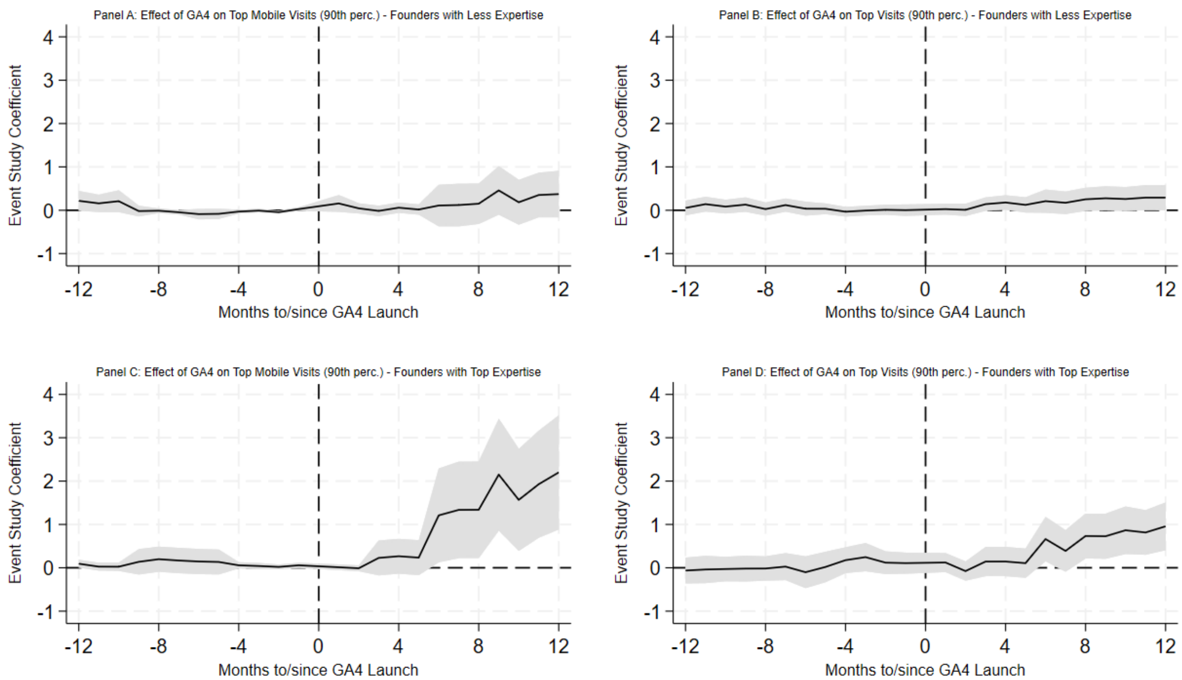

Figure 6: Event Studies for Top Total (Mobile) Visits: By the Expertise of a Founding Team

In this figure, we estimate a dynamic version of Eq. (1) that replaces $PostGA4$ with time indicators for each year-month to/from the release of Google Analytics 4. In Panels A and B, we examine startups led by founders with less expertise. In Panels C and D, we examine startups led by founders with top expertise. The outcome is the likelihood that a startup is in the top percentile for the number of mobile (total) visits its website receives. Startup and year-month fixed effects are included, and standard errors are clustered at the startup level. The shaded area represents the 95 percent confidence interval.

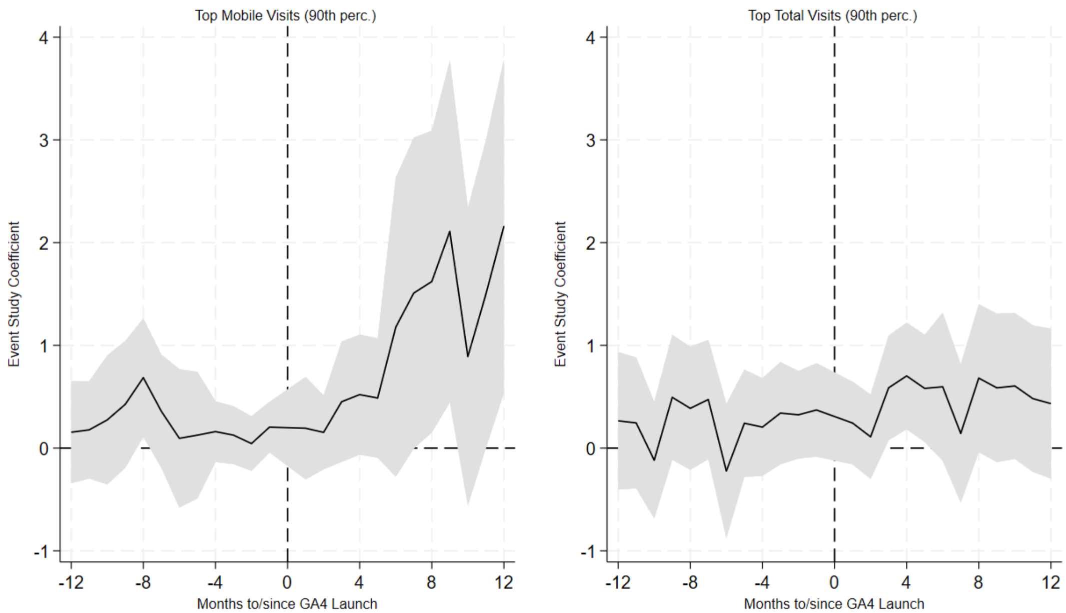

Figure 7: The Joint Effect of A/B Testing and Machine Learning

In this figure, we report the coefficients associated with the triple interaction between A/B, *Develops Apps*, and each year-month to/from the release of Google Analytics 4. The outcome is the likelihood that a startup is in the top percentile for the number of mobile (total) visits its website receives. Startup and year-month fixed effects are included, and standard errors are clustered at the startup level. The shaded area represents the 95 percent confidence interval.

Tables

Table 1: Descriptive Statistics

Variable	**Mean**	**Std. Dev.**	**Min.**	**Max.**	**N**
Number of Visits in 2020	221961	5516438	0	493751540	36835
Number of Mobile Visits in 2020	111776	3814692	0	372844679	36835
Number of Desktop Visits in 2020	110185	2278550	0	206259309	36835
Raised Financing Round	0.277	0.448	0	1	36835
Used Analytics Tools	0.888	0.315	0	1	36835
Used Google Analytics 4	0.498	0.5	0	1	36835
Used A/B Tools	0.371	0.483	0	1	36835
Max. Number Technologies Used	93.431	47.156	0	536	36835
Develops Apps	0.059	0.235	0	1	36835
Innovation Score	78.831	5.781	27.517	101.629	36835
Founders Have Top Expertise	0.128	0.335	0	1	36835
Located in California	0.267	0.443	0	1	36835
Located in Massachusetts	0.034	0.181	0	1	36835
Located in New York	0.131	0.337	0	1	36835

Table 2: Summary Statistics by Whether a Startup Develops App Technologies

	Develops Apps		
	0	1	Test
N	34,666 (94.1%)	2,169 (5.9%)	
Number of Visits in 2020	227926 (5676286)	126632 (1352219)	0.364
Number of Mobile Visits in 2020	116151 (3929413)	41864 (590226)	1.000
Number of Desktop Visits in 2020	111775 (2338090)	84769 (893548)	0.438

Notes: Mean (Standard deviation). In the last column, we report the p-value from a Kruskal-Wallis test of difference in means.

Table 3: The Effect of the Introduction of Google Analytics 4 on Startup Website Visits

	(1) Mobile Visits	(2) Total Visits	(3) Mobile Visits	(4) Total Visits
	Full Sample		*Startups That Raised a Round*	
Post GA4 × Develops Apps	0.658**	0.611**	-0.526	-0.283
	(0.325)	(0.241)	(0.396)	(0.302)
Technology Stack	0.000278	0.000999	0.00690	0.00818**
	(0.00211)	(0.00154)	(0.00437)	(0.00319)
Startup FEs	Yes	Yes	Yes	Yes
Year-month FEs	Yes	Yes	Yes	Yes
Observations	1624912	2486326	503885	684018

Notes: We report the results of PQML count models for the number of mobile visits (columns 1 and 3), and total visits (columns 2 and 4) a startup's website receives in month t. $PostGA4$ is a (0/1) indicator that takes the value one after October 2020, when Google Analytics 4 was released. *Develops Apps* is a (0/1) indicator that identifies startups developing apps. $Technology\ Stack$ is the count of technologies (except those related to analytics) a startup used in month t. Standard errors (in parentheses) are clustered by startup. Significance noted as: *p<0.10; **p<0.05; ***p<0.01.

Table 4: The Effect of the Introduction of Google Analytics 4 on Startup Website Visits: By Cutoffs of the Visits Distribution

	Panel A: Visits Above Median			
	(1) Mobile Visits	(2) Total Visits	(3) Mobile Visits	(4) Total Visits
	Full Sample		*Startups That Raised a Round*	
Post GA4 × Develops Apps	0.0189***	0.0136***	0.0292***	0.00464
	(0.00323)	(0.00489)	(0.00664)	(0.00946)
Technology Stack	0.00122***	0.00299***	0.00171***	0.00410***
	(0.0000304)	(0.0000442)	(0.0000574)	(0.0000791)
Startup FEs	Yes	Yes	Yes	Yes
Year-month FEs	Yes	Yes	Yes	Yes
Observations	2492033	2492033	684656	684656
R2	0.460	0.519	0.439	0.515
Mean D.V.	0.0557	0.166	0.0718	0.213
	Panel B: Top Visits (90th percentile)			
	(1) Mobile Visits	(2) Total Visits	(3) Mobile Visits	(4) Total Visits
	Full Sample		*Startups That Raised a Round*	
Post GA4 × Develops Apps	0.00735***	0.0143***	0.0135***	0.0236***
	(0.00207)	(0.00319)	(0.00453)	(0.00671)
Technology Stack	0.000381***	0.000913***	0.000454***	0.00113***
	(0.0000209)	(0.0000296)	(0.0000405)	(0.0000562)
Startup FEs	Yes	Yes	Yes	Yes
Year-month FEs	Yes	Yes	Yes	Yes
Observations	2492033	2492033	684656	684656
R2	0.461	0.520	0.396	0.493
Mean D.V.	0.0111	0.0333	0.0144	0.0426

Notes: In Panel A, we report the results of linear probability models for the likelihood that a startup is above the median for the number of mobile visits (columns 1 and 3) and total visits (columns 2 and 4) a startup's website receives in month t. In Panel B, We report similar results but consider the top percentile of (mobile) visits as a cutoff. $PostGA4$ is a (0/1) indicator that takes the value one after October 2020, when Google Analytics 4 was released. *Develops Apps* is a (0/1) indicator that identifies startups developing apps. $Technology\ Stack$ is the count of technologies (except those related to analytics) a startup used in month t. Standard errors (in parentheses) are clustered by startup. Significance noted as: *p<0.10; **p<0.05; ***p<0.01.

Table 5: The Effect of the Introduction of Google Analytics 4 on Startup Website Visits: By the Innovativeness of a Startup's Venture

	(1)	(2)	(3)	(4)
	Top Mobile Visits	Top Total Visits	Top Mobile Visits	Top Total Visits
	Novel Projects		*Traditional Projects*	
Post GA4 × Develops Apps	0.0118***	0.0198***	0.00177	0.00725*
	(0.00302)	(0.00460)	(0.00263)	(0.00417)
Technology Stack	0.000310***	0.000809***	0.000434***	0.000992***
	(0.0000255)	(0.0000401)	(0.0000313)	(0.0000420)
Startup FEs	Yes	Yes	Yes	Yes
Year-month FEs	Yes	Yes	Yes	Yes
Observations	1245962	1245962	1246071	1246071
R2	0.444	0.517	0.476	0.523
Mean D.V.	0.0107	0.0329	0.0113	0.0334

Notes: We report the results of linear probability models for the likelihood that a startup is in the top percentile for the number of mobile visits (columns 1 and 3), and total visits (columns 2 and 4) a startup's website receives in month t. The sample in columns 1 and 2 encompasses innovative startups, while the sample in columns 3 and 4 encompasses less-innovative startups. $PostGA4$ is a (0/1) indicator that takes the value one after October 2020, when Google Analytics 4 was released. *Develops Apps* is a (0/1) indicator that identifies startups developing apps. *Technology Stack* is the count of technologies (except those related to analytics) a startup used in month t. Standard errors (in parentheses) are clustered by startup. Significance noted as: *p<0.10; **p<0.05; ***p<0.01.

Table 6: The Effect of the Introduction of Google Analytics 4 on Startup Website Visits: By the Expertise of the Founding Team

	(1)	(2)	(3)	(4)
	Top Mobile Visits	Top Total Visits	Top Mobile Visits	Top Total Visits
	Founders with Top Expertise		*Other Founders*	
Post GA4 × Develops Apps	0.0234***	0.0431***	0.00371**	0.00753**
	(0.00850)	(0.0115)	(0.00183)	(0.00303)
Technology Stack	0.000583***	0.00138***	0.000348***	0.000838***
	(0.0000706)	(0.0000864)	(0.0000216)	(0.0000312)
Startup FEs	Yes	Yes	Yes	Yes
Year-month FEs	Yes	Yes	Yes	Yes
Observations	320372	320372	2171661	2171661
R2	0.460	0.537	0.462	0.515
Mean D.V.	0.0204	0.0594	0.00978	0.0294

Notes: We report the results of linear probability models for the likelihood that a startup is in the top percentile for the number of mobile visits (columns 1 and 3) and total visits (columns 2 and 4) a startup's website receives in month t. The sample in columns 1 and 2 encompasses startups led by founders with top expertise, while the sample in columns 3 and 4 encompasses startups led by founders with less expertise. $PostGA4$ is a (0/1) indicator that takes the value one after October 2020, when Google Analytics 4 was released. *Develops Apps* is a (0/1) indicator that identifies startups developing apps. *Technology Stack* is the count of technologies (except those related to analytics) a startup used in month t. Standard errors (in parentheses) are clustered by startup. Significance noted as: *p<0.10; **p<0.05; ***p<0.01.

Table 7: The Effect of the Introduction of Google Analytics 4 on Startup Website Visits: By the Expertise of the Founding Team and the Innovativeness of a Startup's Venture

	$Top\ Expertise = 1$	$Top\ Expertise = 0$
$Innovativeness = 1$	0.052*** (0.015) [0.059]	0.012** (0.004) [0.025]
$Innovativeness = 0$	0.028(0.017) [0.060]	0.003 (0.004) [0.033]

Notes: We report the results of linear probability models for the likelihood that a startup is in the top percentile for the number of total visits. We examine four subsamples corresponding to the four cells in the table. The coefficients reported are those associated with the interaction between $PostGA4$ and *Develops Apps*. Standard errors (in parentheses) are clustered by startup. The figures reported in square brackets refer to the mean of the dependent variable for the subsample examined. Significance noted as: *p<0.10; **p<0.05; ***p<0.01.

Table 8: The Effect of the Introduction of Google Analytics 4 on A/B Testing's Contribution to Startup Website Visits

	(1) A/B Testing	(2) Mobile Visits	(3) Total Visits	(4) Top Mobile Visits	(5) Top Total Visits
Post GA4 × Develops Apps	0.00566	0.978***	0.648**	0.00173	0.00499**
	(0.00625)	(0.373)	(0.275)	(0.00142)	(0.00249)
A/B		0.626**	0.661**	-0.00289**	0.00525**
		(0.313)	(0.260)	(0.00131)	(0.00217)
Post GA4 × A/B		-0.498**	-0.388**	0.0125***	0.0260***
		(0.251)	(0.181)	(0.00111)	(0.00176)
A/B × Develops Apps		0.723	0.636*	0.00316	0.00467
		(0.448)	(0.345)	(0.00580)	(0.00932)
Post GA4 × A/B × Develops Apps		-0.455	-0.165	0.0154***	0.0253***
		(0.479)	(0.346)	(0.00591)	(0.00856)
Technology Stack	0.00410***	0.000296	0.00109	0.000347***	0.000796***
	(0.0000615)	(0.00206)	(0.00150)	(0.0000220)	(0.0000305)
Startup FEs	Yes	Yes	Yes	Yes	Yes
Year-month FEs	Yes	Yes	Yes	Yes	Yes
Observations	2492033	1624912	2486326	2492033	2492033
R2	0.757			0.462	0.522

Notes: Column 1 reports the results of a linear probability model for the cumulative likelihood that a startup performs A/B testing in month t. Columns 2 and 3 report the results of PQML count models for the number of mobile and total visits a startup's website received in t. Columns 4 and 5 report the results of linear probability models for the likelihood that a startup is in the top percentile for the number of mobile (total) visits its website receives in t. A/B is a (0/1) indicator that takes the value one from the month that a startup begins to perform A/B testing. $PostGA4$ is a (0/1) indicator that takes the value one after October 2020, when Google Analytics 4 was released. *Develops Apps* is a (0/1) indicator that identifies startups developing apps. *Technology Stack* is the count of technologies (except those related to analytics) a startup used in month t. Standard errors (in parentheses) are clustered by startup. Significance noted as: *p<0.10; **p<0.05; ***p<0.01.

Appendix

A1 Theoretical framework: main intuitions

Let $x \in X$ be some vector of features, $y \in Y$ the set of outcomes, and $D = \{(x_1, y_1), ..., (x_n, y_n)\}$ a finite dataset in a certain business domain. Let us define a theory T as the causal mapping between D and the collection C of all bijective mappings $c(\cdot) : X \rightrightarrows Y$, i.e. $T(\cdot) : D \rightarrow C$. Given a certain dataset in a certain business domain, the entrepreneur focuses on the features x that they believe affect the likelihood of occurrence of a chosen interest outcome (Camuffo et al., 2023). An entrepreneur has a prior belief π in her initial theory T and considers certain variables x as relevant for prediction. She additionally holds a vector of prior beliefs p over the selected variables. We assume that there is a *true* theory T^* that provides the highest value for an entrepreneurial venture, conditional on the business domain. The goal of our entrepreneur is, therefore, to explore the search space and select the theory that is closest to T^*, given D.[1] Thus, our entrepreneur must frame 1) general theories, represented by statistical models $T(\cdot, D) \in C$, and 2) within the selected model, alternative (sub)theories, rising from the selection of a specific vector of explanatory variables x, such that $T(x, D) \subset Y$. They select both the model that causally links x to y and the xs to include in the model. As an illustrative example, y can represent standard performance metrics in the entrepreneurial context, such as revenue, customer acquisition, and customer retention. Instead, the covariates in x, represent elements that an entrepreneur believes could influence the outcome of interest y. These could include, for example, user experience, an entrepreneur's marketing capabilities, and the level of market competition.

Therefore, assuming all entrepreneurs are endowed with the same dataset D, they encounter two distinct layers of uncertainty: i) uncertainty regarding the relevant statistical model and represented by the prior π and ii) uncertainty regarding the explanatory variables embedded in the model, represented by priors p. These two types of uncertainty are linked to the concepts of i) model misspecification and ii) model estimation uncertainty, respectively (Montiel Olea et al., 2022).

To improve their search, let our entrepreneur utilize a collection of predictive tools $\mathfrak{P}$. We posit that some of these tools can be used to *explore* the search space, thereby addressing model misspecification, while others are used to *test* the theory, namely, to test the coefficients of the selected explanatory variables. Indeed, some predictive tools can help identify more potential valuable features within the selected dataset (Tranchero, 2023), while others are useful for rejecting (or supporting) the null hypotheses regarding the relevance of selected features (Koning et al., 2022).

A1.0.1 Theory selection: Model misspecification

We denote $\mathbb{P}_1 \in \mathfrak{P}$ the family of tools used to explore the search space and predict the right model. These tools, built on machine learning and AI, aim to analyze large amounts of data to make predictions about customer behavior and the business context, more generally. For simplicity, we label $\mathbb{P}_1$ as *predictive AI*. Predictive AI generates signals about the likelihood p that a certain feature x is salient for the chosen theory and sheds light on potential causal links between the features in x and the outcome of interest y, providing information to update the prior belief π about the statistical model.

In sum, an entrepreneur starts with an initial theory $T(\cdot)$ and explores the search space with AI to select potential features x that are deemed salient for generating value. They start from a certain model or theory, $y = \beta x' + \epsilon$, where x' is a vector of potentially relevant features and ϵ a Gaussian error, and leverage predictive AI to update both π and p. As such, predictive AI helps to enhance existing theories and create more valuable projects.

[1] In more formal terms, the entrepreneur seeks to minimize the expected value of a standard quadradic loss function, equal to the square of the difference between the true theory T^* and her theory T, under the priors π and p, for a given dataset D. For the purposes of our discussion, we will not delve into these technicalities.

Intuition 1. *Entrepreneurs using predictive AI are more likely to develop more valuable projects.*

Predictive AI can also help entrepreneurs identify hidden patterns in the data[2], which we refer to as logical *anomalies*. Building upon Mullainathan and Rambachan (2024), a logical anomaly is a subset of the given dataset $D' \subset D$ that is not compatible with a pre-selected theory $T(\cdot)$, i.e. $T(x, D') = \varnothing$ for each $x \in X$. Therefore, predictive AI can help entrepreneurs identify patterns that cannot be explained by existing theories. It is important to clarify that anomalies should not be considered merely as initially unidentified explanatory variables, but as causal patterns that are not captured by the pre-selected statistical model[3]. Thus, the detection of anomalies can aid in the identification of the correct model and reduce errors due to model misspecification.

However, it is reasonable to assume that simply detecting logical anomalies does not necessarily lead to the selection of a better theory. While detecting logical anomalies can pave the way for breakthroughs by revealing new directions, integrating these anomalies into a new theory can be challenging and is likely to result in a skewed distribution of innovative outcomes. In other words, the *true* theory T^* is the most valuable theory among all possible theories, including those incorporating such logical anomalies. It is straightforward that T^* not only outperforms the others but envisions a novel idea, reflecting the incorporation of new unexplored patterns in the data. A direct consequence of such a decision-making process is that envisioning novel theories is a challenging task that requires substantial cognitive effort and superior human capital. Overall, this discussion translates into the following intuitions:

Intuition 2. *Entrepreneurs utilizing predictive AI are more likely to generate breakthrough projects than to achieve average performance improvements.*

Intuition 3. *Highly skilled entrepreneurs using predictive AI perform relatively better.*

A1.0.2 Theory testing: Model estimation

We denote $\mathbb{P}_2 \in \mathfrak{P}$ the family of tools aimed at testing an pre-chosen theory. A clear example of such tools is *A/B testing* (Azevedo et al., 2020; Koning et al., 2022). After selecting a model, A/B testing tools allow an entrepreneur to test the β coefficients. Note the difference here. Predictive AI provides predictions about p over covariates x and the potential causal link with y, while A/B testing tools test each coefficient β and provide a clear cutoff to reject or support the null hypothesis. This last step is akin to a standard experimentation phase in entrepreneurship, which improves performance by enhancing precision.

What emerges from our framework is the sequential structure of the entrepreneurial decision-making process. Our entrepreneur frames a theory, improves it by incorporating features that generate more value and anomalies, then tests their theory. A direct consequence of adopting predictive AI tools is the improvement of the marginal productivity of A/B testing tools, since they are now used to test more valuable theories. By reducing model estimation errors through AI, an entrepreneur can generate significantly more value:

Intuition 4. *Predictive AI increases the marginal productivity of A/B testing tools.*

[2] `https://medium.com/techmagic/ai-anomaly-detection-what-you-need-to-know-ba2bae7cb510`.

[3] Anomalies could lead to the inclusion of other covariates $x \in X$ as a result of the adoption of a new theory.

A2 Method to construct our innovation measure

To measure how innovative a startup is, we use machine learning techniques to build an innovation score, on a scale from 0 to 100. We followed the following steps. First, we used a training dataset of 903 startups that were founded in 2019. We focus on these startups because they were founded prior to the release of GA4 in October 2020. We then asked Gemini to evaluate the level of innovation of these startups by analyzing the startups' descriptions written by their founders on LinkedIn and Crunchbase as well as the texts of their patents, whenever startups were granted U.S. patents. As a robustness check, we built the same measure using input from ChatGPT rather than Gemini.

Secondly, we developed a predictive model to forecast the innovation level of our sample startups. This model has two main milestones. First, we adopted TF-IDF Vectorization, which converts text data into a matrix of TF-IDF features, capturing the importance of words in each document relative to the entire dataset. TF measures the frequency of a word in a document. It assigns higher weights to words that appear more frequently. Conversely, IDF measures the importance of a word across the entire corpus of documents (in our case, startup descriptions), assigning lower weights to common words and higher weights to rarer words. The TF-IDF score for each word in a document is the product of its TF and IDF scores, capturing both the word's importance in the document and its uniqueness across the corpus of documents. As a last milestone of our predictive model, we employed a Stochastic Gradient Descent Regressor (SGDRegressor) with an epsilon-insensitive loss function. SGDRegressor finds the best-fitting line for a set of data points. It does so by making small adjustments (gradients) to the fitting line, based on random sub-samples of the data. We chose this regressor due to its efficiency and suitability for handling large-scale linear regression problems. The model was configured with no penalty (regularization) and a high maximum iteration count (10,000) to ensure thorough convergence. The final output from this regression is normalized on a scale from 0 to 100.

A3 Tables

Table A1: The Effect of the Introduction of Google Analytics 4 on Startup Website Visits: Having Kept Singleton Observations

	(1)	(2)	(3)	(4)
	Mobile Visits	Total Visits	Mobile Visits	Total Visits
	Full Sample		*Startups That Raised a Round*	
Post GA4 × Develops Apps	0.658**	0.611**	-0.526	-0.283
	(0.325)	(0.241)	(0.396)	(0.302)
Technology Stack	0.000278	0.000999	0.00690	0.00818**
	(0.00211)	(0.00154)	(0.00437)	(0.00319)
Startup FEs	Yes	Yes	Yes	Yes
Year-month FEs	Yes	Yes	Yes	Yes
Observations	1624912	2486326	503885	684018

Notes: We report the results of PQML count models for the number of mobile visits (columns 1 and 3) and total visits (columns 2 and 4) a startup's website receives in month t. This time, we keep singleton observations. $PostGA4$ is a (0/1) indicator that takes the value one after October 2020, when Google Analytics 4 was released. *Develops Apps* is a (0/1) indicator that identifies startups developing apps. *Technology Stack* is the count of technologies (except those related to analytics) a startup used in month t. Standard errors (in parentheses) are clustered by startup. Significance noted as: *p<0.10; **p<0.05; ***p<0.01.

Table A2: The Effect of the Introduction of Google Analytics 4 on Startup Website Visits: Having Winsorized the Top 1% of Observations of the Number of Visits

	(1)	(2)	(3)	(4)
	Mobile Visits	Total Visits	Mobile Visits	Total Visits
	Full Sample		*Startups That Raised a Round*	
Post GA4 × Develops Apps	0.454***	0.322***	0.277*	0.212*
	(0.108)	(0.0815)	(0.158)	(0.113)
Technology Stack Technology Stack	0.00515***	0.00581***	0.00999***	0.0108***
	(0.000709)	(0.000709)	(0.00122)	(0.00121)
Startup FEs	Yes	Yes	Yes	Yes
Year-month FEs	Yes	Yes	Yes	Yes
Observations	1624912	2486326	503885	684018

Notes: We report the results of PQML count models for the number of mobile visits (columns 1 and 3), and total visits (columns 2 and 4) a startup's website receives in month t. We winsorize the top 1% of observations of the number of (mobile) visits. $PostGA4$ is a (0/1) indicator that takes the value one after October 2020, when Google Analytics 4 was released. *Develops Apps* is a (0/1) indicator that identifies startups developing apps. $Technology\ Stack$ is the count of technologies (except those related to analytics) a startup used in month t. Standard errors (in parentheses) are clustered by startup. Significance noted as: *p<0.10; **p<0.05; ***p<0.01.

Table A3: The Effect of the Introduction of Google Analytics 4 on Startup Top Website Visits: Having Winsorized the Top 1% of Observations of the Number of Visits

	(1)	(2)	(3)	(4)
	Top Mobile Visits	Top Total Visits	Top Mobile Visits	Top Total Visits
	Full Sample		*Startups That Raised a Round*	
Post GA4 × Develops Apps	0.00735***	0.0143***	0.0135***	0.0236***
	(0.00207)	(0.00319)	(0.00453)	(0.00671)
Technology Stack	0.000381***	0.000913***	0.000454***	0.00113***
	(0.0000209)	(0.0000296)	(0.0000405)	(0.0000562)
Startup FEs	Yes	Yes	Yes	Yes
Year-month FEs	Yes	Yes	Yes	Yes
Observations	2492033	2492033	684656	684656
R2	0.461	0.520	0.396	0.493
Mean D.V.	0.0111	0.0333	0.0144	0.0426

Notes: We report the results of linear probability models for the likelihood that a startup is in the top percentile for the number of mobile visits (columns 1 and 3), and total visits (columns 2 and 4) a startup's website receives in month t. We winsorize the top 1% of observations of the number of (mobile) visits. $PostGA4$ is a (0/1) indicator that takes the value one after October 2020, when Google Analytics 4 was released. *Develops Apps* is a (0/1) indicator that identifies startups developing apps. $Technology\ Stack$ is the count of technologies (except those related to analytics) a startup used in month t. Standard errors (in parentheses) are clustered by startup. Significance noted as: *p<0.10; **p<0.05; ***p<0.01.

Table A4: The Effect of the Introduction of Google Analytics 4 on Startup Website Visits: IV

	(1) Top Mobile Visits	(2) Top Total Visits
GA4	0.140***	0.105***
	(0.0373)	(0.0338)
Startup FEs	Yes	Yes
Year-month FEs	Yes	Yes
Observations	2072307	2072307

Notes: We instrument the actual usage of GA4 by a startup (i.e. an indicator equal to 1 during the period a startup uses GA4) with a Bartik instrument capturing the exogenous usage of GA4 in the cities in which the startups -operating in given industries- are located. In all models, we examine all the year months starting from January 2021 (thus encompassing the release of Google Analytics 4) and include startups founded after 2020. The F test of excluded instruments is 51. Standard errors (in parentheses) are clustered by startup. Significance noted as: *p<0.10; **p<0.05; ***p<0.01.

Table A5: The Effect of the Introduction of Google Analytics 4 on Startup Website Visits: By Cutoffs of the Visits Distribution and Using Crunchbase Industry Group Keywords to Extend the Sample of App Developers

	Panel A: Visits Above Median			
	(1)	(2)	(3)	(4)
	Mobile Visits	Total Visits	Mobile Visits	Total Visits
	Full Sample		*Startups That Raised a Round*	
Post GA4 × Develops Apps	0.0172***	0.0110***	0.0262***	-0.00351
	(0.00247)	(0.00373)	(0.00526)	(0.00746)
Technology Stack	0.00122***	0.00299***	0.00172***	0.00410***
	(0.0000304)	(0.0000443)	(0.0000575)	(0.0000791)
Startup FEs	Yes	Yes	Yes	Yes
Year-month FEs	Yes	Yes	Yes	Yes
Observations	2492033	2492033	684656	684656
R2	0.460	0.519	0.439	0.515
Mean D.V.	0.0557	0.166	0.0718	0.213
	Panel B: Top Visits (90th percentile)			
	(1)	(2)	(3)	(4)
	Mobile Visits	Total Visits	Mobile Visits	Total Visits
	Full Sample		*Startups That Raised a Round*	
Post GA4 × Develops Apps	0.00729***	0.0124***	0.0134***	0.0203***
	(0.00149)	(0.00237)	(0.00341)	(0.00505)
Technology Stack	0.000382***	0.000915***	0.000456***	0.00113***
	(0.0000209)	(0.0000296)	(0.0000405)	(0.0000562)
Startup FEs	Yes	Yes	Yes	Yes
Year-month FEs	Yes	Yes	Yes	Yes
Observations	2492033	2492033	684656	684656
R2	0.461	0.520	0.397	0.493
Mean D.V.	0.0111	0.0333	0.0144	0.0426

Notes: We employ Crunchbase industry group keywords to extend our sample of app developers. In Panel A, we report the results of linear probability models for the likelihood that a startup is above the median for the number of mobile visits (columns 1 and 3) and total visits (columns 2 and 4) a startup's website receives in month t. In Panel B, We report similar results but consider the top percentile of (mobile) visits as a cutoff. $PostGA4$ is a (0/1) indicator that takes the value one after October 2020, when Google Analytics 4 was released. *Develops Apps* is a (0/1) indicator that identifies startups developing apps. *Technology Stack* is the count of technologies (except those related to analytics) a startup used in month t. Standard errors (in parentheses) are clustered by startup. Significance noted as: *p<0.10; **p<0.05; ***p<0.01.

Table A6: The Effect of the Introduction of Google Analytics 4 on Startup Website Visits: By Cutoffs of the Visits Distribution and Using Crunchbase Industry Group Keywords to Identify App Developers

	Panel A: Visits Above Median			
	(1)	(2)	(3)	(4)
	Mobile Visits	Total Visits	Mobile Visits	Total Visits
	Full Sample		*Startups That Raised a Round*	
Post GA4 × Develops Apps	0.0158***	0.00912**	0.0194***	-0.00989
	(0.00265)	(0.00403)	(0.00571)	(0.00823)
Technology Stack	0.00118***	0.00296***	0.00169***	0.00403***
	(0.0000269)	(0.0000392)	(0.0000530)	(0.0000732)
Startup FEs	Yes	Yes	Yes	Yes
Year-month FEs	Yes	Yes	Yes	Yes
Observations	3188775	3188775	815668	815668
R2	0.458	0.510	0.440	0.514
Mean D.V.	0.0511	0.157	0.0700	0.210
	Panel B: Top Visits (90th percentile)			
	(1)	(2)	(3)	(4)
	Mobile Visits	Total Visits	Mobile Visits	Total Visits
	Full Sample		*Startups That Raised a Round*	
Post GA4 × Develops Apps	0.00518***	0.00862***	0.00851**	0.0135**
	(0.00158)	(0.00272)	(0.00372)	(0.00587)
Technology Stack	0.000381***	0.000913***	0.000455***	0.00113***
	(0.0000209)	(0.0000296)	(0.0000405)	(0.0000562)
Startup FEs	Yes	Yes	Yes	Yes
Year-month FEs	Yes	Yes	Yes	Yes
Observations	2492033	2492033	684656	684656
R2	0.461	0.520	0.396	0.493
Mean D.V.	0.0111	0.0333	0.0144	0.0426

Notes: We employ Crunchbase industry group keywords to identify app developers. In Panel A, we report the results of linear probability models for the likelihood that a startup is above the median for the number of mobile visits (columns 1 and 3) and total visits (columns 2 and 4) a startup's website receives in month t. In Panel B, We report similar results but consider the top percentile of (mobile) visits as a cutoff. $PostGA4$ is a (0/1) indicator that takes the value one after October 2020, when Google Analytics 4 was released. *Develops Apps* is a (0/1) indicator that identifies startups developing apps. $Technology\ Stack$ is the count of technologies (except those related to analytics) a startup used in month t. Standard errors (in parentheses) are clustered by startup. Significance noted as: *p<0.10; **p<0.05; ***p<0.01.

Table A7: The Effect of the Introduction of Google Analytics 4 on Startup Website Visits: Weighted by (1-Bounce Rate)

	Panel A: Visits Above Median			
	(1)	(2)	(3)	(4)
	Mobile Visits	Total Visits	Mobile Visits	Total Visits
	Full Sample		*Startups That Raised a Round*	
Post GA4 × Develops Apps	0.0154***	0.00635	0.0195***	-0.0132
	(0.00290)	(0.00460)	(0.00579)	(0.00905)
Technology Stack	0.000818***	0.00292***	0.00132***	0.00430***
	(0.0000286)	(0.0000441)	(0.0000530)	(0.0000776)
Startup FEs	Yes	Yes	Yes	Yes
Year-month FEs	Yes	Yes	Yes	Yes
Observations	2492033	2492033	684656	684656
R2	0.432	0.476	0.423	0.473
Mean D.V.	0.0557	0.166	0.0718	0.229
	Panel B: Top Visits (90th percentile)			
	(1)	(2)	(3)	(4)
	Mobile Visits	Total Visits	Mobile Visits	Total Visits
	Full Sample		*Startups That Raised a Round*	
Post GA4 × Develops Apps	0.00580***	0.0102***	0.0101***	0.0178***
	(0.00182)	(0.00291)	(0.00387)	(0.00627)
Technology Stack	0.000340***	0.000816***	0.000425***	0.00119***
	(0.0000186)	(0.0000273)	(0.0000333)	(0.0000530)
Startup FEs	Yes	Yes	Yes	Yes
Year-month FEs	Yes	Yes	Yes	Yes
Observations	2492033	2492033	684656	684656
R2	0.434	0.494	0.378	0.478
Mean D.V.	0.0111	0.0333	0.0144	0.0495

Notes: We employ Crunchbase industry group keywords to identify app developers. In Panel A, we report the results of linear probability models for the likelihood that a startup is above the median for the number of mobile visits (columns 1 and 3) and total visits (columns 2 and 4) a startup's website receives in month t. These numbers are weighted by one minus the bounce rate. The bounce rate is defined as the percentage of visitors who enter the site and then leave rather than continuing to view other pages within the same site. In Panel B, We report similar results but consider the top percentile of (mobile) visits as a cutoff. $PostGA4$ is a (0/1) indicator that takes the value one after October 2020, when Google Analytics 4 was released. *Develops Apps* is a (0/1) indicator that identifies startups developing apps. $Technology\ Stack$ is the count of technologies (except those related to analytics) a startup used in month t. Standard errors (in parentheses) are clustered by startup. Significance noted as: *p<0.10; **p<0.05; ***p<0.01.

Table A8: The Effect of the Introduction of Google Analytics 4 on Startup Website Users

	Panel A: Visits Above Median			
	(1)	(2)	(3)	(4)
	Mobile Visits	Total Visits	Mobile Visits	Total Visits
	Full Sample		*Startups That Raised a Round*	
Post GA4 × Develops Apps	0.0202***	0.0153***	0.0310***	0.00510
	(0.00330)	(0.00487)	(0.00681)	(0.00956)
Technology Stack	0.00123***	0.00301***	0.00173***	0.00440***
	(0.0000305)	(0.0000441)	(0.0000576)	(0.0000799)
Startup FEs	Yes	Yes	Yes	Yes
Year-month FEs	Yes	Yes	Yes	Yes
Observations	2492033	2492033	684656	684656
R2	0.464	0.523	0.437	0.514
Mean D.V.	0.0557	0.166	0.0718	0.232
	Panel B: Top Visits (90th percentile)			
	(1)	(2)	(3)	(4)
	Mobile Visits	Total Visits	Mobile Visits	Total Visits
	Full Sample		*Startups That Raised a Round*	
Post GA4 × Develops Apps	0.00813***	0.0127***	0.0146***	0.0256***
	(0.00216)	(0.00312)	(0.00471)	(0.00690)
Technology Stack	0.000401***	0.000975***	0.000477***	0.00138***
	(0.0000221)	(0.0000299)	(0.0000432)	(0.0000584)
Startup FEs	Yes	Yes	Yes	Yes
Year-month FEs	Yes	Yes	Yes	Yes
Observations	2492033	2492033	684656	684656
R2	0.481	0.516	0.411	0.489
Mean D.V.	0.0111	0.0333	0.0144	0.0499

Notes: We employ Crunchbase industry group keywords to identify app developers. In Panel A, we report the results of linear probability models for the likelihood that a startup is above the median for the number of mobile users (columns 1 and 3) and total users (columns 2 and 4) visiting a startup's website in month t. In Panel B, We report similar results but consider the top percentile of (mobile) users as a cutoff. *PostGA4* is a (0/1) indicator that takes the value one after October 2020, when Google Analytics 4 was released. *Develops Apps* is a (0/1) indicator that identifies startups developing apps. *Technology Stack* is the count of technologies (except those related to analytics) a startup used in month t. Standard errors (in parentheses) are clustered by startup. Significance noted as: *p<0.10; **p<0.05; ***p<0.01.

Table A9: The Effect of the Introduction of Google Analytics 4 on Startup Visits: By the Innovativeness of a Startup's Venture - Triple Interactions & Robustness with the ChatGPT Innovation Score

	(1) Top Mob. Visits	(2) Top Tot. Visits	(3) Top Mob. Visits	(4) Top Tot. Visits
Post GA4 × Dev. Apps × Novel P. (Gemini)	0.0104***	0.0130**		
	(0.00400)	(0.00620)		
Post GA4 × Dev. Apps × Novel P. (ChatGPT)			0.00848**	0.0120*
			(0.00405)	(0.00622)
Technology Stack	0.000380***	0.000912***	0.000380***	0.000912***
	(0.0000209)	(0.0000295)	(0.0000209)	(0.0000295)
Startup FEs	Yes	Yes	Yes	Yes
Year-month FEs	Yes	Yes	Yes	Yes
Observations	2492033	2492033	2492033	2492033
R2	0.461	0.520	0.461	0.520
Mean D.V.	0.0111	0.0333	0.0111	0.0333

Notes: We augment Eq. 1 with a triple interaction between $DevelopApps_i$, $PostGA4$, and an indicator identifying startups developing innovative projects. In columns 1 and 2, the innovation score was built using Gemini. In columns 3 and 4, the innovation score was built using input from ChatGPT. $PostGA4$ is a (0/1) indicator that takes the value one after October 2020, when Google Analytics 4 was released. *Develops Apps* is a (0/1) indicator that identifies startups developing apps. *Technology Stack* is the count of technologies (except those related to analytics) a startup used in month t. Standard errors (in parentheses) are clustered by startup. Significance noted as: *p<0.10; **p<0.05; ***p<0.01.

Table A10: The Effect of the Introduction of Google Analytics 4 on Startup Website Visits: By the Expertise of the Founding Team - Triple Interaction

	(1) Top Mobile Visits	(2) Top Total Visits
Post GA4 × Develops Apps × Top Expertise	0.0196**	0.0352***
	(0.00874)	(0.0120)
Technology Stack	0.000376***	0.000904***
	(0.0000209)	(0.0000295)
Startup FEs	Yes	Yes
Year-month FEs	Yes	Yes
Observations	2492033	2492033
R2	0.462	0.521
Mean D.V.	0.0204	0.0594

Notes: We augment Eq. 1 with a triple interaction between $DevelopApps_i$, $PostGA4$, and an indicator identifying startups managed by founders with top expertise. $PostGA4$ is a (0/1) indicator that takes the value one after October 2020, when Google Analytics 4 was released. *Develops Apps* is a (0/1) indicator that identifies startups developing apps. $Technology\ Stack$ is the count of technologies (except those related to analytics) a startup used in month t. Standard errors (in parentheses) are clustered by startup. Significance noted as: *p<0.10; **p<0.05; ***p<0.01.

Table A11: The Effect of the Introduction of Google Analytics 4 on Startup Website (Mobile) Visits: By the Expertise of the Founding Team and the Innovativeness of a Startup's Venture

	$Top\ Expertise = 1$	$Top\ Expertise = 0$
$Innovativeness = 1$	0.026** (0.011) [0.019]	0.008** (0.003) [0.008]
$Innovativeness = 0$	0.020 (0.014) [0.023]	-0.0016 (0.002) [0.011]

Notes: We report the results of linear probability models for the likelihood that a startup is in the top percentile for the number of website visits from mobile devices. We examine four subsamples corresponding to the four cells in the table. The coefficients reported are those associated with the interaction between $PostGA4$ and *Develops Apps*. Standard errors (in parentheses) are clustered by startup. The figures reported in squared brackets refer to the mean of the dependent variable for the subsample examined. Significance noted as: *p<0.10; **p<0.05; ***p<0.01.

Table A12: The Joint Effect of A/B testing and Google Analytics 4 on Startup Visits - IV

	(1) Top Mobile Visits	(2) Top Total Visits	(3) IV - Mobile	(4) IV - Total
A/B	0.00801***	0.0118***	0.300***	0.299***
	(0.00158)	(0.00168)	(0.0916)	(0.0922)
GA4	0.00416***	0.00366***		
	(0.000843)	(0.000846)		
A/B × GA4	0.00953***	0.00620***		
	(0.00147)	(0.00154)		
A/B × Develops Apps			0.0852***	0.0940***
			(0.0250)	(0.0266)
Startup FEs	Yes	Yes	Yes	Yes
Year-month FEs	Yes	Yes	Yes	Yes
Observations	2072307	2072307	2072307	2072307

Notes: Columns 1 and 2 report the results of a linear probability model for the probability that a startup is in the top percentile for the number of mobile (total) visits its website received in month t. The regressors of interest are A/B, a (0/1) indicator that becomes 1 from the moment a startup performs an A/B test, and $GA4$, a (0/1) indicator that becomes 1 from the moment a startup utilizes Google Analytics 4. Columns 3 and 4, we report the results of an IV model where we instrument A/B and the interaction between A/B and whether a startup develops apps using a Bartik instrument capturing the exogenous usage of A/B tools in the cities in which the startups -operating in given industries- are located and an interaction between this instrument and whether the startup develops apps. In all models, we examine all the year months starting from January 2021 (thus encompassing the release of Google Analytics 4) and include startups founded after 2020. The F tests of excluded instruments take values 15 and 313 in the first stages for A/B, and the interaction between A/B and $DevelopsApps$, respectively. Standard errors (in parentheses) are clustered by startup. Significance noted as: *p<0.10; **p<0.05; ***p<0.01.

References

Azevedo, E. M., Deng, A., Montiel Olea, J. L., Rao, J., and Weyl, E. G. (2020). "A/b testing with fat tails." *Journal of Political Economy*, *128*(12), 4614–000.

Camuffo, A., Gambardella, A., and Pignataro, A. (2023). "Theory-driven strategic management decisions." *CEPR DP 17664v2*.

Koning, R., Hasan, S., and Chatterji, A. (2022). "Experimentation and start-up performance: Evidence from a/b testing." *Management Science*, *68*(9), 6434–6453.

Montiel Olea, J. L., Ortoleva, P., Pai, M. M., and Prat, A. (2022). "Competing models." *The Quarterly Journal of Economics*, *137*(4), 2419–2457.

Mullainathan, S., and Rambachan, A. (2024). "From predictive algorithms to automatic generation of anomalies." Tech. rep., National Bureau of Economic Research.