The Selective Tailwind Effect of Artificial Intelligence in Entrepreneurship

Annamaria Conti Danilo Messinese*†

March 21, 2025

Abstract

What role does AI play in entrepreneurial decision-making? We explore this question by exploiting large-scale data on US startups and the random release of Google Analytics 4 (GA4), which introduced AI tools especially beneficial for mobile app developers. Leveraging this shock in a difference-in-differences model, we find that the GA4 release significantly boosted customer acquisition, with the performance gains from AI adoption driven by the upper tail of the treatment effect distribution. These effects are most pronounced for innovative startups led by highly skilled founders. A survey of tech founders and an ad hoc experiment elucidate the mechanism: Entrepreneurs rely on causal theories to extract values from AI-detected data anomalies to generate valuable innovations. Our findings underscore the selective tailwind effect of AI in complementing skilled human capital to generate breakthroughs.

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

---

*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*.

†We are especially grateful to Ajay Agrawal and all the participants at the 2nd Bayesian Entrepreneurship Conference at the ION Management Science Lab, Economics of AI Conference (IESE) and Crete Innovation Entrepreneurship workshop by Esade for their insightful comments and feedback. We are also indebted to several startups for their time and effort in completing the survey. We sincerely thank Kevin Bryan, Emanuele Cagnolo, Marco Ceccagnoli, George Chondrakis, Gaétan de Rassenfosse, Riccardo Fini, Mike Impink, Aldona Kapacinskaite, Petteri Leppänen, Mike Nolet, Leonardo Ortega, Joe Ploog, and Benedikt Seigner, for their invaluable assistance in facilitating access to survey respondents. We also extend our gratitude to James Miner and Enrico Bergamini from FONGIT - The Innovation Platform, Raphael Hellmann from the EPFL Entrepreneur Club, Victor Vildé and Louis Ramard from Station F/HEC Paris, as well as Hervé Lebret from Inria Startup Studio for their support in this effort. Annamaria Conti acknowledges funding from the Swiss National Science Foundation (Project ID: 100013_188998). The experiment presented in this manuscript is included in the AEA RCT Registry (AEARCTR-0015201).

1 Introduction

> *"Novelty ordinarily emerges only for the man who, knowing with precision what he should expect, is able to recognize that something has gone wrong [...]. Anomalies that lead to paradigm change will penetrate existing knowledge to the core."*
>
> — Thomas Kuhn, *The Structure of Scientific Revolutions* (1962), p. 65

While it is well established that entrepreneurs operate under uncertainty, how they should navigate it remains unclear (Zellweger and Zenger, 2023). This ambiguity is particularly concerning because, in the absence of guidance, entrepreneurs tend to take surprisingly few measures—such as building a prototype and testing demand—to hedge against uncertainty (Bennett and Chatterji, 2023), while simultaneously collecting and using information with different scopes (Gambardella and Messinese, 2024). The consequences can be profound, affecting not only venture performance but also broader economic outcomes, including innovation and growth (Conti and Roche, 2021; Haltiwanger et al., 2013). Recent advances in AI might offer entrepreneurs powerful tools to navigate uncertainty, helping them generate and assess new strategies more effectively (Csaszar et al., 2024). By broadening the scope of viable innovations and reducing the risk of costly mistakes, AI has the potential to fundamentally influence how entrepreneurs approach decision-making under uncertainty.

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 tools is especially widespread among more innovative companies and skilled individuals (Brynjolfsson and McElheran, 2016; McElheran et al., 2024), the consensus over which actors benefit the most from these tools and through what mechanisms is mixed. Some research indicates that these tools might be especially beneficial for actors at the highest end of the skill distribution (Brynjolfsson et al., 2021), other studies suggest that such tools may elevate the skills and capabilities of actors at the lower end (Dell'Acqua et al., 2023). There is also no consensus on whether AI helps boost productivity within the innovation frontier (Dell'Acqua et al., 2023; Wu et al., 2020), or if they also help users go beyond the frontier (Mullainathan and Rambachan, 2024), and whether they merely assist with improving the efficiency of already established operations or whether they help formulate new strategies (Agrawal et al., 2023; Ludwig and Mullainathan, 2024). This study is the first to offer a large-scale empirical analysis of the impact of AI tools on entrepreneurial outcomes. To deepen our insights, we complement our core findings with a targeted survey of entrepreneurs and an experiment, both elucidating the mechanisms driving our large-scale results.

To guide our empirical analysis, we propose a conceptual framework highlighting AI's role in fostering innovation through anomaly detection and predictive analytics. While these tools uncover hidden patterns that signal opportunities, their value depends on human theory-based causal reasoning (Camuffo et al., 2023), enabling skilled entrepreneurs to generate new theories by mapping anomalies to relevant explanatory factors (Mullainathan and Rambachan, 2024; Ortoleva, 2012). We bring these conjectures to the data, exploring how the adoption of AI tools affects startup customer acquisition.

In the first part of our analysis, we constructed a dataset of 36,835 U.S. software startups founded between 2016 and 2018, as listed on Crunchbase. We enriched this with time-variant web technology data from BuiltWith and website traffic metrics from Semrush, creating a comprehensive dataset that links firm characteristics, technology adoption, and performance. We then estimated a difference-in-differences model leveraging the quasi-random timing of Google Analytics 4 (GA4)'s release in October 2020, which introduced AI-driven predictive tools particularly beneficial for mobile app developers.

Compared to its predecessor, Universal Analytics, GA4 offers enhanced machine learning capabilities, including improved attribution models, more accurate anomaly detection, and new predictive metrics.[1] It also enables more precise measurement of mobile visit contributions to company websites,[2] helping mobile app developers optimize user experience, engagement, and retention.[3] We classify startups based on whether they develop apps and examine how GA4's introduction affected their website traffic—used as a proxy for consumer acquisition—compared to other startups.

The results are striking. Absent significant pre-trends, we find that 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. Indeed, a startup's likelihood of receiving above-median mobile visits increases by 34%, while the probability of reaching the 90th percentile rises by a larger margin of 66%.

After uncovering our main effects, we examine their underlying sources. Using machine learning to assess a startup's innovativeness, we find that these effects are primarily driven by startups at the innovation frontier. In fact, the likelihood of a startup ranking in the top decile for mobile and total visits increases by 110% and 60%, respectively, for innovative ventures, compared to only 16% and 22% for those less innovative. This result highlights the pivotal role of AI tools in enabling frontier startups to produce more valuable innovations,

[1] Refer to: `https://easyinsights.ai/blog/ten-ways-ga4-is-better-than-universal-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`.

driving breakthrough performance. Next, we categorize founders' human capital based on their education and work experience. Our classification provides strong evidence that the availability of new AI 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%. Finally, we highlight that the strongest effects of GA4 derive from frontier startups led by highly skilled founders.

Our results thus far suggest that AI tools do more than help founders extrapolate past trends for operational efficiency—they may be pivotal in shaping novel business strategies. To explore this, we surveyed 135 founders through our entrepreneurial network. We uncover that startups developing apps are 27 percentage points more likely to have used AI tools, particularly GA4, than other surveyed startups. These tools not only optimize operations but are also instrumental for developing new projects and for strategic decision-making. Of the entrepreneurs surveyed, 54% reported using AI for anomaly detection at least occasionally, leveraging detected anomalies to refine their existing conceptual models or develop new ones, thereby driving both incremental and radical innovation. These findings complement our large-scale analysis, suggesting that the skewed impact of GA4 on startup performance may stem from founders using AI tools to refine strategies and uncover "hidden" opportunities through anomaly detection.

As anomalies appear to play a key role, we conducted an experiment to examine how they help founders identify valuable innovations and interact with entrepreneurs' causal reasoning. We developed a data analysis platform and recruited 275 entrepreneurs or aspiring entrepreneurs through Prolific. Participants, acting as CEOs of a food delivery company, analyzed an ad hoc dataset containing information on the firm's delivery plans and customer characteristics. They were randomly assigned to one or two interventions:

an anomaly detection tool similar to GA4 or a tutorial on causal-based reasoning in innovation. After analyzing the data, they proposed at least one strategy to improve customer acquisitions. The results are striking: entrepreneurs leveraging anomalies relied less on prior knowledge to identify new opportunities, though this effect was partially offset by the causal reasoning intervention. Moreover, exposure to both the anomaly detection tool and the causal reasoning tutorial led to the identification of more impactful innovations.

We contribute to two main literatures. First, we advance the literature on AI adoption in organizations (Agrawal et al., 2024; Bresnahan, 2019; Brynjolfsson and McElheran, 2016; McElheran et al., 2024) and the role of AI in enhancing (Brynjolfsson and McElheran, 2016; Gaessler and Piezunka, 2023; Goldfarb et al., 2023; McElheran et al., 2024; Toner-Rodgers, 2024) or displacing human tasks (Acemoglu et al., 2022; Acemoglu and Restrepo, 2019). Our focus is on the role of AI tools in informing and guiding entrepreneurs to generate valuable innovations. Challenging the view that AI tools produce low-variance ideas (Meincke et al., 2024; Wu et al., 2020), we demonstrate that these tools help overcome human cognitive limitations, enabling highly skilled founders to achieve top performance gains in novel fields. While this large-scale data analysis complements the findings of Otis et al. (2023) and Koning et al. (2022), our survey data and lab experiment further demonstrate that AI tools help founders uncover hidden patterns that deviate from the norm, which they integrate into their causal reasoning to refine strategies and uncover novel opportunities.

Second, we inform the strand on theory-based decision-making in entrepreneurship (Agrawal et al., 2025; Camuffo et al., 2020, 2024; Csaszar et al., 2024; Felin et al., 2024; Felin and Holweg, 2024; Gans et al., 2019; Zellweger and Zenger, 2023). This literature has compared entrepreneurs to scientists who form beliefs, test them, and integrate the test results into their theoretical constructs. We demonstrate that AI complements human cognition in producing valuable innovations, particularly in novel fields where human cognitive capabilities are more constrained. We show that at least one relevant mechanism is AI's ability to detect anomalies, which supplements its traditional role of extrapolating trends from existing data. The main message of this study is that AI provides a selective tailwind, enhancing skilled humans' causal reasoning through anomaly detection and driving the generation of valuable innovations.

2 Theoretical Framework

To navigate the uncertainty they face, entrepreneurs typically engage in iterative *experimentation*, gathering and interpreting information from customers and stakeholders—an approach often central to entrepreneurial strategy (Gans et al., 2019; Levinthal, 2017; Ries, 2011). While promising, critics argue that this approach overrelies on "learning by testing" while neglecting deeper causal reasoning (Felin et al., 2020).

To address this concern, recent research suggests that entrepreneurs should act as "scientific" experimenters, developing hypotheses grounded in *value-creating causal theories* (Camuffo et al., 2020, 2024; Ehrig and Schmidt, 2022; Zellweger and Zenger, 2023). These theories provide a coherent structure for hypothesis testing, leading to more insightful experimentation. In parallel, scholars emphasize the importance of integrating *anomalies* into entrepreneurial reasoning, as unexpected information can reveal valuable opportunities (Felin and Zenger, 2017; Mullainathan and Rambachan, 2024). This forward-looking approaches extend beyond evaluating known scenarios, uncovering previously unrecognized paths to value creation.

Advances in AI and the growing availability of data have made data-driven analysis integral to entrepreneurial decision-making. AI, described as a "prediction technology" by Agrawal et al. (2022), excels at analyzing large datasets and forecasting outcomes (Brynjolfsson et al., 2021; Cockburn et al., 2019). While AI is recognized as effective when opportunities are easily identifiable (Toner-Rodgers, 2024), in novel contexts, it could add value by detecting anomalies and guiding "scientific" entrepreneurs toward breakthroughs (Mullainathan and Rambachan, 2024).

To formalize these dynamics, we propose a simple framework in which entrepreneurs hold an initial set of theories, face uncertainty about their validity, and use AI to detect novel, hidden paths that refine their causal reasoning.

2.1 Framework Setup

We begin by introducing key definitions and assumptions that lay the foundation for our main intuitions. We define $D = \{(x_i, y_i)\}_{i=1}^{n}$ as a dataset where each observation i includes (i) a vector of explanatory variables $x_i \in X \subset \mathbb{R}^k$ and (ii) an outcome of interest $y_i \in Y \subset \mathbb{R}$. We define a *causal theory* as a mapping $T(\cdot)\colon X \to Y$ that attempts to capture how the features $x \in X$ generate the outcome $y \in Y$. Concretely, each theory T specifies both *which* explanatory variables matter and *how* they combine to produce y. We call $\mathcal{C}$ the *theory space*—the set of all theories. Within $\mathcal{C}$, any theory T may be incorrect for two primary reasons: i) *Model misspecification*, the functional form of T may be inappropriate and ii) *Estimation error*, the theory may omit relevant variables or mis-estimate parameters.

Assumption 1 (Existence and Inclusiveness of the True Theory)**.** *There exists a true causal theory T^* within $\mathcal{C}$ that explains a known dataset D and which, if perfectly identified and implemented, yields the maximum possible expected payoff for an entrepreneur's venture.*

One can think of the entrepreneur as aiming to minimize the expected gap between her chosen theory T and T^*, i.e., $\min_{T \in \mathcal{C}} \mathbb{E}\left[(T^*(x) - T(x))^2\right]$, subject to her prior beliefs about which theories are more or less likely to be correct. In practice, $\mathbb{E}[\cdot]$ represents a Bayesian expectation over the parameters and functional forms within $\mathcal{C}$. We posit that, initially, an entrepreneur has access to a baseline known dataset D_0 and holds a baseline theory $T_0 \in \mathcal{C}_0 \subset \mathcal{C}$ on D_0.

Baseline True Theory. $T_0^* \in \mathcal{C}_0^*$ is the *baseline true theory*, such that

$$T_0^* = \arg\max_{T_0 \in \mathcal{C}_0} \Pi(T_0; D_0).$$

T_0^* maximizes the entrepreneur's payoff under the *initial* data D_0. It is noteworthy that the baseline dataset D_0 may include relevant but unobserved features, which we label *anomalies*. Following Mullainathan and Rambachan (2024), we define an *anomaly* as any observation (x_i, y_i) not explained by $T_0 \in \mathcal{C}_0$ or for which $T_0(x_i) = \varnothing$. Such anomalies suggest that T_0 fails to account for certain relevant data aspects, even for T_0^*. Thus, the baseline true theory T_0^* is "true" only for the dataset fully known to the entrepreneur.

Expanded True Theory. Anomalies may arise from new domain points D' that an entrepreneur might not have *recognized* in D_0. Let $D_1 = D_0 \cup D'$ denote an *expanded dataset* that includes anomalies, assuming that the entrepreneur manages to observe them at some point. There may exist an "augmented" or *expanded true theory* $T_1^* \in \mathcal{C}_1^*$ such that:

$$T_1^* = \arg\max_{T_1 \in \mathcal{C}_1} \Pi(T_1; D_0 \cup D'),$$

with $\mathcal{C}_1^* \supset \mathcal{C}_0^*$. If D' contains valuable insights (e.g., new features, new domains, new causal pathways), then

$$\Pi\big(T_1^*; D_0 \cup D'\big) \; > \; \Pi\big(T_0^*; D_0\big),$$

that is, T_1^* yields a strictly higher payoff as it exploits initially unrecognized opportunities.

Assumption 2 (Possibility of Domain Expansion)**.** *The set $\mathcal{C}_1^*$ strictly contains $\mathcal{C}_0^*$ (i.e., $T_1^* \notin \mathcal{C}_0^*$) whenever anomalies reveal a new domain $D' \neq \varnothing$. Thus, the "expanded" theory space admits causal relationships or variables unobserved in D_0.*

AI-based Expansion. Let $\mathcal{AI}$ be an AI-based anomaly detection algorithm that, given (D_0, T_0), identifies $D' \subseteq D_0$ as anomalous:

$$\mathcal{AI} : (D_0, T_0) \mapsto D'.$$

We assume that $\mathcal{AI}$ not only detects anomalies but also translates them into, for instance, neglected covariates or potential functional forms—thereby *expanding* the set of candidate theories from $\mathcal{C}_0$ to a broader set $\mathcal{C}_1 = \mathcal{C}_0 \cup \mathcal{C}'$, where $\mathcal{C}'$ encompasses the newly discovered causal pathways.

Assumption 3 (AI Reveals Potentially Valuable Anomalies)**.** *$\mathcal{AI}$ uncovers anomalies in D_0 which, if incorporated into a revised theory T_1, bring the entrepreneur strictly closer to T_1^*.*

If the entrepreneur adopts $\mathcal{AI}$ ($I = 1$), she observes $D' = \mathcal{AI}(D_0, T_0)$ and refines or replaces T_0 with some $T_1 \in \mathcal{C}_1$. Conversely, if $I = 0$, the entrepreneur continues to rely on $\mathcal{C}_0$. Under the assumption of negligible adoption costs,

$$V^1 \;=\; \max_{T_1 \in \mathcal{C}^1} \Pi\big(T_1; D_0 \cup D'\big) \;\geq\; \max_{T_0 \in \mathcal{C}^0} \Pi\big(T_0; D_0\big) \;=\; V^0.$$

Realistically, not all anomalies lead to the same performance gains. It is reasonable to posit that many simply expose weaknesses in the existing theory without uncovering meaningful opportunities (*'dead-end'* anomalies), while some highlight gaps that reveal previously unexplored mechanisms (*'opportunity-revealing'* anomalies). Only in the latter case does expanding the entrepreneur's theory result in a meaningful increase in payoffs. Thus, we partition the anomaly set $D' = \mathcal{AI}(T_0, D_0)$ into two disjoint subsets: $D'_{\text{dead-end}}$ and $D'_{\text{opportunity}}$.

Novel Ideas. We further model novel ideas or novel theories $T^{\dagger} \in \mathcal{C}_1 \setminus \mathcal{C}_0$ as those that appear *incorrect* or suboptimal when evaluated solely on D_0 given that they incorporate features or causal mechanisms unobserved in D_0. Formally,

Assumption 4 (Novel Ideas)**.** *There exists at least one "novel" theory $T^{\dagger} \in \mathcal{C}_1 \setminus \mathcal{C}_0$ which appears suboptimal on the baseline dataset D_0 but can become strictly superior once relevant anomalies are revealed. Formally, for this $T^{\dagger}$,*

$$\Pi\big(T^{\dagger}; D_0\big) \;<\; \Pi\big(T_0^*; D_0\big) \quad \textit{but} \quad \Pi\big(T^{\dagger}; D_0 \cup D'\big) \;>\; \Pi\big(T_0^*; D_0 \cup D'\big),$$

where $T_0^* \in \mathcal{C}_0^*$ is the baseline true theory and $D' = \mathcal{AI}(D_0, T_0)$ is the set of anomalies detected by the AI. Thus, $T^{\dagger}$ initially underperforms relative to T_0^* on D_0 but outperforms after integrating the newly discovered anomalies.

Human Capital The detection of anomalies does not guarantee a profitable revision of the theory, unless entrepreneurs determine how to integrate these anomalies into a coherent new theory, T_1. This requires causal theorizing, which hinges upon an entrepreneur's cognitive ability and prior knowledge—collectively referred to as *skills.*

Assumption 5 (Human Capital Matters)**.** *Let ω be an index of entrepreneurs' skills. Given a detected anomaly set D', the additional payoff from revising T_0 is $\mathcal{G}\big(D', \omega\big)$, where $\mathcal{G}$ is strictly increasing in ω.*

That is, more-skilled entrepreneurs are better at mapping anomalies into valuable theories by hypothesizing new causal relationships or features, thereby extracting higher payoffs.

2.1.1 Intuitions

From the model setup, we derive the main intuitions that will guide our empirical analysis. Consistent with Assumption 3, if all anomalies lie in $D'_{\text{dead-end}}$, then revising the theory yields only marginal payoff improvements for an entrepreneur:

$$\max_{T_1 \in \mathcal{C}_1} \Pi(T_1; D_1) \gtrapprox \max_{T_0 \in \mathcal{C}_0} \Pi(T_0; D_0).$$

Conversely, if $D'_{\text{opportunity}} \neq \varnothing$ and can be resolved by some $T_1 \in \mathcal{C}_1$ closer to T_1^*, then payoffs can increase substantially:

$$\max_{T_1 \in \mathcal{C}_1} \Pi(T_1; D_1) > \max_{T_0 \in \mathcal{C}_0} \Pi(T_0; D_0).$$

Intuition 1 (Breakthrough Gains from AI)**.** *As long as* $\Pr(D'_{\text{opportunity}} \neq \varnothing)$ *is positive for some entrepreneurs and zero for others, the distribution of payoffs from AI adoption is right-skewed: while many experience no sizeable change, a minority realizes substantial gains.*

Turning the attention to novel ideas, let us consider two types of entrepreneurs: entrepreneurs pursuing incremental ideas whose C_0 is close to T_0^*, and entrepreneurs pursuing novel ideas whose C_0 is far from T_0^*. Given Assumption 4, entrepreneurs pursuing novel ideas hold higher $\Pr(D'_{\text{opportunity}} \neq \varnothing)$ than those pursuing incremental ones and are thus more likely to extract greater value from D', conditional on adopting $\mathcal{AI}$,

$$\Delta V_{\text{Novel}} = \Pi(T^\dagger; D_1) - \Pi(T^\dagger; D_0) \quad > \quad \Delta V_{\text{Incremental}}.$$

Intuition 2 (Novel Ideas Amplify the Effect of AI)**.** *Entrepreneurs pursuing novel ideas experience a larger payoff increase from anomaly detection compared to those pursuing incremental innovations, amplifying the right-tail skew in the distribution of payoffs.*

Finally, based on Assumption 5 regarding entrepreneur human capital, we derive our last intuition.

Intuition 3 (Entrepreneur Skills Amplify the Effect of AI)**.** *Fix a set of AI-detected anomalies D'. Suppose two entrepreneurs with skill levels $\omega^1 > \omega^2$ detect the same D', then:*

$$\Delta\Pi(T_1^1, T_0^1) \; > \; \Delta\Pi(T_1^2, T_0^2),$$

where T_1^i is the revised theory chosen by an entrepreneur with skill ω^i. Thus, entrepreneurs with higher skill levels (ω^1) derive a strictly larger payoff increase upon detecting the same anomalies, reflecting their superior ability to "connect the dots."

The proof of this last intuition is provided in the Appendix. In essence, the largest performance leaps occur only when anomalies *expose fundamental gaps* in the baseline model, and entrepreneurs -particularly those pursuing novel ideas- possess sufficient skills to capitalize on these opportunities. Consequently, the distribution of AI-driven gains is not merely average-improving; it is *tail-inflating*, concentrating substantial benefits among the entrepreneurs who leverage *opportunity-revealing* anomalies and possess the skills to integrate them into novel, high-value theories. In the next sections, we bring our intuitions to the data.

3 Large Scale Data Analysis

In this section, we empirically explore Intuitions 1 to 3 from the previous section, leveraging large-scale data derived from the combination of various datasets. We begin by describing the data, followed by an overview of the empirical methodology and the results.

3.1 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 et al., 2024; 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[4] 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.

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

[4] 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.

(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.[5] Semrush recently partnered with Crunchbase to improve website tracking accuracy for startups, underscoring the growing importance of website visits as a metric for assessing startup 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. Approximately 28% of the startups had raised a financing round as of the end of our sample period, November 2023. The average number of website technologies used by a startup by November 2023 is 93, 50% had used GA4, while 37% had used A/B testing tools.

To measure startup innovativeness, we developed a machine learning-based innovation score. Using a training dataset of 903 startups founded in 2019 to ensure they predated GA4, we had Gemini evaluate their innovativeness by analyzing LinkedIn and Crunchbase descriptions, along with U.S. patent texts when available. We then developed a predictive model to forecast the innovativeness of our sample startups. First, we applied Term

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

Frequency-Inverse Document Frequency (TF-IDF) vectorization to quantify word importance, followed by a Stochastic Gradient Descent Regressor (SGDRegressor) to estimate innovativeness. The final output is a score normalized between 0 (low innovativeness) and 100 (high innovativeness).[6]

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% 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 ⟩

3.2 Empirical Methodology

To assess how the adoption of AI tools impacts 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 enhanced the algorithms used for anomaly detection, identifying

[6] Details are provided in the Appendix.

atypical patterns in the data that could signal key insights. It additionally 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. 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 GA4 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 values, we estimate the following difference-in-differences Poisson Quasi-Maximum Likelihood (PQML) model at the level of startup i observed in year-month t:

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

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

$DevelopApps_i$ 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, 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 assess whether the release of GA4 enabled mobile startups to achieve breakthrough performance or primarily facilitated incremental performance improvements. For this purpose, we will modify Eq. 1 estimating linear probability models for whether a startup's number of website visits received falls in the 90th percentile, serving as a proxy for breakthrough performance, or if the startup attained a number of website visits above the median, serving as a proxy for marginal 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.

$\langle$ Insert Figure 2 about here $\rangle$

3.3 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 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

[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.

2023. With this exercise, we aim to assess whether the potential adoption of AI 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 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 ⟩

3.4 Impact of AI tools on the tails of website visits distribution

Having assessed the average effect of the availability of new AI tools, we next examine their impact on the tails of the website visits distribution. By doing so, we test the prediction in *Intuition 1*, stating that the distribution of gains from AI adoption is right-skewed. Specifically, 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. Here, 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.[9] Overall, these

[9] Results reported in Table A3 show that app developers become 0.0239 (0.0039) less likely to be in the lowest

results are in line with the prediction in *Intuition 1*, stating that the distribution of gains from AI adoption is uneven. Indeed, while the availability of AI tools helps startups boost customer acquisition, the positive premium is driven by the upper tail of the treatment effect distribution. These findings suggest that founders adopting AI tools might produce more valuable innovations and experience breakthrough performance.

⟨ Insert Table 4 and Figure 4 about here ⟩

To verify the robustness of these results, we first show in Table A4 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 A5, 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 the older version of GA4, Universal Analytics

quartile for the number of mobile (total) visits received, equivalent to a 3% (1%) decline relative to the outcome mean.

(UA), 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 UA 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 A6 and A7 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 tools became available.

Finally, to verify that our findings are not specific to the kinds 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 A8 show that the availability of AI 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 A9, 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.

3.5 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. The aim is to test the prediction of *Intuition 2*, that the gains from AI adoption are amplified for entrepreneurs pursuing novel ideas. 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 pursuing innovative ideas 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 pursuing more traditional ideas.

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 A10, 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 A10 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 *Intuition 2* of our conceptual framework, these results suggest that AI tools are particularly beneficial for startups undertaking novel ideas.

⟨ Insert Table 5 and Figure 5 about here ⟩

3.6 Distinguishing By the Human Capital of Startup Founders

In this section, we test the prediction of *Intuition 3*, which states that high-skilled entrepreneurs derive larger gains from AI adoption than their lower-skilled counterparts. To examine this, we categorize startups according to the skill level of their founders. 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 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 A11, 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 align with *Intuition 3*. Since the benefits of AI adoption are unevenly distributed, disproportionately favoring founders with top-tier expertise, our results highlight the importance of human capital, as extracting value from AI tools requires significant cognitive effort.

⟨ Insert Table 6 and Figure 6 about here ⟩

3.7 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 top decile in terms 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 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 top decile 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 A12, where we repeat the same analysis as described earlier, but this time focusing on the number of mobile visits.

⟨ Insert Table 7 about here ⟩

3.8 AI Tools and A/B Testing

To strengthen the results discussed so far, we conclude this section by examining the relationship between AI tools and A/B testing. Since AI tools generate predictions about covariates that may affect startup performance, they should enhance the marginal productivity of A/B testing, which primarily serves as a hypothesis-testing mechanism.

The results from this investigation are reported in Table 8. In column 1, we show that after the release of GA4, mobile startups did not significantly increase 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, in columns 4 and 5, we show that following the release of GA4, app developers who have utilized A/B testing are 1.5 (2.5) percentage points more likely to be in the top decile for the number of mobile (total) visits.

Clearly, the decision to utilize A/B testing tools is an endogenous one; therefore, the results presented in columns 4 and 5 should be interpreted as correlations. Nonetheless, the event studies shown in Figure 7 indicate flat pre-trends prior to the release of GA4 (Panel A and B). Following GA4's launch, there is a significant increase in the probability of a startup being in the top decile for mobile visits among those that employed A/B testing (Panel A). In Panel B, the positive effect of GA4's release on the likelihood of a startup reaching the top decile for total visits is less pronounced. As reported in Table A13, these results remain robust to estimating an instrumental variables (IV) model using a Bartik-like instrument similar to that employed in Table A4. Overall, our results highlight the importance of AI tools in enhancing the value that A/B testing can generate for entrepreneurial decision-making.

⟨ Insert Table 8 and Figure 7 about here ⟩

4 Exploring the Mechanisms

The results from our large-scale data suggest that while adopting AI tools can enhance customer acquisition for startups, the positive premium is driven by the upper tail of the treatment effect distribution. Consistent with this finding, we also demonstrate that the effects are strongest for innovative startups led by highly skilled founders. As we mentioned in Section 2, these results suggest that AI tools may not only enhance startups' operational efficiency but also enable them to implement more valuable strategies in the form of statistical models, helping to define their functional form and relevant features. This unfolds in a context where entrepreneurs act as 'scientific' experimenters, elucidating causal relationships between model features and outcomes, while the identification of anomalies through AI might help them achieve breakthrough performance. In this section, we explore the role of AI tools in defining and updating a startup's strategy, as well as the potential impact of anomalies. We do so first with a descriptive survey administered to startup founders, and then with an ad hoc lab experiment.

4.1 Survey Evidence

As we mentioned above, one possible mechanism driving our results is that AI tools, like GA4, might help founders detect anomalies, which, in turn, could be conducive to discovering new opportunities, leading to breakthroughs. To descriptively explore this mechanism, we administered a survey to software founders in the US, Canada, Europe, and Asia. Given the challenges of getting founders to complete our survey, we leveraged personal connections and partnered with incubators in France and Switzerland, securing participation from 135 founders.[10] Of these founders, 32 were from France, 26 from Switzerland, 29 from the rest of Europe, 14 from North America, and 15 from Asia, while 19 could not be categorized. Among them, 24 held a Ph.D., 53 had a master's degree, and the remainder had a bachelor's degree. A total of 61 founders earned their highest degree in a STEM discipline, 88 had at least some work experience, and 30 had worked for startups in the past. Among the 135 founders' startups, 42 were listed on Crunchbase, and 30 had raised funding. Additionally, 37 founders indicated that their startup's business model was primarily B2C,

[10] The text of the survey is in the Appendix.

while 98 described it as B2B. Furthermore, 86 founders reported having developed apps.

An analysis of the survey questions, detailed in Appendix B, reveals several interesting patterns. First, the vast majority of founders (84%) reported framing at least occasionally a conceptual model to interpret information, using causal logic (e.g., X causes Y) to predict outcomes and guide strategic decisions. Among these founders, 88% develop hypotheses at least sometimes, compared to just 38% of those who never use causal logic. Furthermore, 92% of founders who use causal logic also employ A/B testing at least occasionally, while only 68% of the remainder do so. This provides important initial evidence, as the underlying assumption of the model in Section 2 is that founders act like scientists—forming theories, developing hypotheses, and testing them.

Regarding the use of predictive AI tools (PAI), 66% of startups developing apps reported using them, compared to 45% of other startups. The difference is statistically significant, with a p-value of 0.007. When we specifically consider the use of GA4, the percentages are 44% for app-developing startups and 20% for the others, with a p-value for the difference of 0.003. These findings are consistent with our large-scale data, which show that startups developing apps adopt GA4 more frequently than other software startups.

Figure 8 presents evidence on how founders use predictive AI tools. Panel A shows that startups primarily use these tools to optimize their apps or websites and to forecast market trends and customer behavior, with secondary applications in predicting competitor behavior. The differences in the use of predictive AI tools between startups developing apps and other startups are not statistically significant at conventional levels.

As displayed in Panel B, survey respondents, on average, rated the relevance of predictive AI tools at 53 on a scale from 0 to 100, for exploring potential features of their statistical models, and at 52, and for selecting the most relevant ones. The values are similar across startups developing apps and other startups.

Panel C shows that predictive AI tools are particularly relevant for developing innovative projects, rather than just optimizing operations or working in areas where founders can leverage their expertise. These tools are also used by founders to refine their strategies.

Finally, Panel D summarizes the extent to which founders had to modify their strategies, operations, and marketing teams in response to the adoption of GA4. As shown, founders primarily modified their strategies, with fewer making changes to their operations.

Overall, the evidence presented in Figure 8 suggests that there is potential for predictive AI tools to enrich and enhance the theories founders develop around their business ideas.

⟨ Insert Figure 8 about here ⟩

Moving to the specific use of predictive AI tools for anomaly detection, 54% of founders reported using anomaly detection services at least sometimes. This percentage is 48% among founders using GA4 and 10% among the remainder, with the difference being statistically significant (p-value of 0.00). Moreover, Panel A of Figure 9 shows that only one startup disregarded anomalies in its data upon detecting them with GA4. In contrast, 52% and 47% of startups using these tools reported being likely to use anomalies to explore new opportunities and evaluate potential problems with their business ideas, respectively.

Panel B shows that only 18% of startups reported never using information from data anomalies to update their existing conceptual model or frame a new one. Among those who reported using this information at least once, 41% stated that they would use it most of the time.

Panel C shows that 82% of startups exploring anomalies made incremental changes to their business idea at least once, with 45% doing so more than once. Remarkably, Panel D reveals that a significant percentage of startups exploring anomalies radically changed their business idea at least once (57%), and 26% did so more than once.

⟨ Insert Figure 9 about here ⟩

Overall, this descriptive evidence suggests that the use of AI tools, particularly GA4, is widespread among startups developing apps. These tools not only help optimize operations but also play a crucial role in shaping business strategies. Moreover, consistent with our assumptions in Section 2, AI tools seem to leverage anomalies to help founders refine and enrich their theories, fostering both incremental and radical innovations. These results clarify our large-scale data findings, suggesting that GA4's positive impact on customer acquisition—driven by the upper tail of the treatment effect distribution—possibly arises from certain founders leveraging AI tools to enhance their strategies and potentially uncover novel opportunities by detecting and assessing anomalies.

4.2 Lab Experiment

4.2.1 Experimental Design

To further investigate how the detection of anomalies allows founders to update their theories and the relationship between anomaly exploration and founders' causal reasoning, we designed and conducted an online experiment. In this experiment, participants acted as CEOs of a food delivery startup, analyzed simulated sales data through a custom-built platform, and identified potential business innovations.

The experiment employed a factorial design, with participants randomly assigned to one of four groups in the 2×2 matrix shown in Figure 10. Before beginning the task, all participants watched a generic 5-minute video on data analytics and innovation. Following this, the first treatment group watched a supplementary tutorial to be trained in framing a conceptual model (*theory*) for interpreting data-extrapolated information using causal logic. The second treatment group viewed a video that presented data *anomalies* as hidden patterns that can reveal novel business opportunities. Both videos included clear, real-world examples. The third treatment group received both interventions, while the control group watched only an extended version of the generic video, without any nudges.

⟨ Insert Figure 10 about here ⟩

All four groups had access to a platform we designed and developed as a simplified version of Google Analytics 4. This approach ensured consistency with the study's large-scale data analysis while preventing participants from feeling overwhelmed by GA4's complexity without adequate training. Screenshots of the platform for the group treated with both interventions are provided in Figure 11.

Before accessing the platform, all participants watched a tutorial video explaining the platform's features. The platform allowed them to easily visualize time trends for key performance metrics of the food delivery app, such as the number of users, revenue, the popularity of new recipes, and the number of page views. As shown in Panel A of Figure 11, all participants could view a blue line representing the actual 2024 trend of a given metric, alongside a grey line displaying the expected value of the same metric based on 2023 data, with shaded areas indicating the confidence interval. This representation was designed to help entrepreneurs anchor their reasoning in historical trends.

Additionally, participants could segment each metric by various factors, including device type, weekday versus weekend orders, meal plan preferences, customer group, and meal plan type. To generate the data, we exploited available information on consumption patterns typical of food delivery companies in the US. As an illustrative example, Panel B of Figure 11 shows that user engagement for "Keto" meal plans (a ketogenic diet, which is high in fat and low in carbohydrates) experiences spikes in consumer demand just before summer, particularly among single customers. This trend reflects the app's popularity among young, single individuals seeking a service to manage their diet in preparation for the summer season.

⟨ Insert Figure 11 about here ⟩

Although all groups had access to the platform and could visualize the same trends, there were slight variations in their experiences. Participants in the groups trained to frame causal theories were exposed to a brief sentence reinforcing this concept. Meanwhile, those trained on the importance of anomalies saw anomalies highlighted with colored circles—exactly as Google Analytics 4 does in its anomaly detection section. As shown in Panel B of Figure 11, participants who received both interventions could visualize both the reinforcing sentence and the circles highlighting data anomalies.

To identify anomalies in the data, we implemented the same machine learning model used by GA4, namely the Bayesian state space model. Such a model estimates the hidden (latent) state of a system over time by incorporating prior beliefs, observed data, and uncertainty. It dynamically updates predictions based on new information, identifying anomalies as deviations from the expected state while accounting for noise and temporal dependencies. This approach is particularly useful for detecting sudden changes or irregular patterns in time-series data, which can help uncover potential business opportunities. For example, an anomalous spike in demand for Keto meals by single customers just before summer could prompt participants to develop targeted strategies for this consumer segment.

After visualizing the data, participants were asked to propose innovations to improve the business performance of their food delivery startup. The best innovations would be rewarded with a monetary prize of $100. The estimated completion time for the task was 40 minutes.

We conducted the experiment on Prolific (Palan and Schitter, 2018) in January 2025, following preregistration in the AEA Registry (AEARCTR-0015201). The experiment involved 289 participants from the US, Europe, Australia, and Canada, selected based on the following criteria: some level of entrepreneurial experience, an undergraduate degree or higher in STEM disciplines or social sciences, and expertise in computer programming (see

Tables A14-A17 for details). Participants were required to complete a brief survey before starting the experiment. This survey was designed to gather control variables, in addition to those provided by Prolific, such as participants' expertise in the food delivery industry, their ability to reason using causal logic, and their familiarity with tools like Google Analytics 4, other predictive AI tools, and generative AI. The text of the survey, which closely mirrors the one administered to founders (though shorter), is provided in the Appendix.

To ensure data quality, we excluded participants who failed to answer checkpoint questions correctly after the instructional videos. Additionally, we removed participants who completed the task in an unrealistically short time and failed to provide sufficiently detailed descriptions of their proposed business ideas. Applying these criteria, we retained 275 participants: 68 in the control group, 64 in the *theory* group, 62 in the AI-*anomalies* treatment group, and the remaining 81 in the group treated with both *theory* and AI-*anomalies*.

4.2.2 Experimental Results

Here, we report and discuss the results of our experiment, focusing on: (i) the impact of the interventions on the likelihood that participants generated valuable business ideas, and (ii) the extent to which they leveraged prior knowledge in developing these ideas.

To construct the dependent variable for (i), we employed an advanced Natural Language Processing (NLP) model to systematically evaluate the potential of the business ideas participants proposed for enhancing customer acquisition. Specifically, we first processed each submitted idea using TF-IDF vectorization. Subsequently, we applied K-Means clustering to categorize these vectorized ideas into distinct thematic groups. K-Means clustering partitioned the data into a predefined number of clusters by iteratively assigning each idea to the cluster with the nearest centroid, minimizing intra-cluster variance and maximizing inter-cluster separation.

To estimate the probability that each idea would successfully attract new customers relative to other ideas in the same cluster, we adopted an empirical Bayesian methodology. This methodology combined the insights from the clustering analysis with ChatGPT's prior knowledge of historical benchmark data, past market innovations, and validated best practices within the food delivery sector. We enriched ChatGPT's knowledge by feeding it the 2023 and 2024 datasets displayed in the lab experiment. We then employed the totality of this information to compute our probabilistic estimates of an idea's potential.

To align our experimental analysis with the large-scale data results discussed earlier, we finally retained each participant's idea with the highest predicted probability of enhancing customer acquisition. This probability constitutes our dependent variable, labeled as *Market Success Probability*, and serves as a proxy for the potential effectiveness of a business idea in increasing customer acquisition.

The results of assessing the impact of our interventions on *Market Success Probability* are reported in Table 9. Given that the dependent variable ranges from zero to one, we estimated a Fractional Logistic Regression model with robust standard errors. In this model, we control for participants' demographic characteristics, which were provided by Prolific. Additionally, we account for participants' adoption of causal theories, their use of GA4 and other AI tools, and their prior knowledge of the food delivery industry. This information was obtained from the survey participants completed before the experiment.

We observe an increase in the likelihood of identifying valuable business ideas resulting from the interplay between the theory-based and anomaly-driven approaches. In fact, the coefficient for *Intervention (Theory)* is not statistically significant, suggesting that causal reasoning alone does not lead to a measurable improvement in customer acquisition compared to the control group. Similarly, *Intervention (Anomalies)* alone shows a negative albeit weakly significant effect (-0.19, p-value = 0.079), suggesting that detecting deviations in customer behavior without a structured decision-making framework could result in misleading information, potentially leading to less valuable business ideas. However, the interaction term *Intervention (Theory × Anomalies)* has a positive and statistically significant coefficient of 0.393 (p-value = 0.009). This coefficient corresponds to a marginal increase of 39.3% in the probability of identifying valuable business opportunities, indicating that the combined application of causal theories and anomaly detection generates a synergistic effect. Figure 12 provides a graphical representation of the marginal effect of *Intervention (Theory × Anomalies)*. As shown, the predicted probability that a participant's most valuable idea enhances customer acquisition is highest for those who received both the theory-based and anomaly-driven interventions.

The results from Table 9 and 12 are consistent with our theoretical frameworks in that the detection of anomalous data points helps uncover high-value ideas only to the extent that it enriches a prior theory or facilitates the adoption of a new theory in $\mathcal{C}_1$, obtaining a higher payoff Π. However, in the absence of a theory, decision-makers might derail, finding themselves unable to causally explain unexpected data. Therefore, these findings underscore the importance of decision-making models that integrate structured theoretical reasoning with data-driven anomaly detection, thereby optimizing the process leading to the generation of valuable innovations.

⟨ Insert Table 9 and Figure 12 about here ⟩

To investigate the mechanisms underlying the results in Table 9, we analyze participants' self-assessments at the end of the task, where they rated the extent to which they leveraged their prior knowledge in developing their business ideas on a scale from 0 to 100. In Table 10, we report the results from a Poisson model regressing this outcome on the four groups identified by the interventions, with the control group serving as a reference outcome. Table 10 presents the results from regressing this outcome on the four intervention groups, with the control group serving as the reference category. Notably, the anomaly-driven intervention, when applied in isolation, reduces the extent to which entrepreneurs leverage prior knowledge in generating novel business ideas by 34.9 percentage points (p-value = 0.001). This suggests that entrepreneurs discard their prior knowledge when confronted with surprising information. Conversely, the coefficient for the interaction between the theory-based and anomaly-driven interventions is 0.282 (p-value = 0.043), suggesting that theory-based causal reasoning mitigates the negative effect of anomaly detection on the application of prior knowledge, neutralizing it.

Interpreting these results in light of our framework suggests that entrepreneurs without training in causal theory may be more likely to disregard prior knowledge when encountering an anomaly and instead rely on that anomaly to develop a new business idea. Conversely, when entrepreneurs are also trained to frame a causal theory, they might use the insight of the anomaly they observed to revise their existing theory, without necessarily discarding their prior knowledge. This is because they would know how to explain the anomaly through the lens of their theoretical construct to build a more valuable theory, implying $\mathcal{C}_1 \supset \mathcal{C}_0$).

⟨ Insert Table 10 about here ⟩

5 Concluding Remarks

The impact of AI on entrepreneurial outcomes hinges on how effectively AI tools can offer strategic guidance, uncovering valuable patterns for founders to pursue. This paper takes a first step towards addressing this question, combining large-scale data analysis, a survey of founders, and a lab experiment to provide a number of insights.

Implementing a difference-in-differences empirical approach that exploits the quasi-random release of GA4 in a large dataset of U.S. software startups, we first show that AI tools enhance a startup's customer acquisitions. However, these gains are disproportionately driven by the upper tail of the treatment effect distribution, and are largest when AI tools are employed by skilled founders leading innovative startups. These findings suggest that

AI tools play a key role in producing higher-value innovations, enhancing the capabilities of skilled founders in pushing the innovation frontier forward.

To better understand the mechanisms at play, we surveyed over 100 entrepreneurs, uncovering that AI tools do more than optimize operations—they are also pivotal in shaping business strategies. Entrepreneurs reported using AI to enhance their causal reasoning and develop new projects. Notably, they underscored the importance of anomaly detection, as identifying anomalies enables them to explore new opportunities in addition to diagnosing potential flaws in their business models. This process, in turn, facilitates both incremental and radical innovation in their original business ideas.

These insights are further validated through an experiment with Prolific entrepreneurs, where participants who had access to an anomaly detection tool and causal reasoning training were able to identify more valuable innovations. This suggests that AI tools enable founders to detect hidden patterns that deviate from the norm, which they integrate into their causal reasoning to refine strategies and uncover novel opportunities.

Our findings offer relevant implications for the entrepreneurial literature. While prevailing views suggest that AI tools help entrepreneurs develop low-variance, incremental innovations by extrapolating trends from historical data (Dell'Acqua et al., 2023; Meincke et al., 2024), our study presents an alternative perspective. Echoing the work of Acemoglu and Restrepo (2019) and Acemoglu et al. (2022) on the non-democratic nature of AI tools, our results suggest that these tools empower skilled founders to produce highly valuable innovations, especially in underexplored areas where experience, albeit useful, is not sufficient on its own to make sense of data.

Contrary to concerns that AI tools might amplify a "streetlight effect" (Hoelzemann et al., 2024; Kim, 2023), our findings suggest that AI can enhance founders' causal reasoning by identifying valuable patterns they might have otherwise overlooked. In this regard, our study contributes to the work by Mullainathan and Rambachan (2024) on the role of anomalies in advancing economic research: By uncovering predictive signals that existing theories fail to capture and entrepreneurs might miss, AI has the potential to accelerate breakthrough innovations. In light of this, anomaly detection algorithms, such as those provided by GA, might not only help entrepreneurs diagnose issues in their models but also reveal new opportunities, ultimately leading to richer decision-making frameworks.

Our analysis leaves several important questions unanswered, opening avenues for future research. First, while our focus was on the effects of predictive AI tools, extending the analysis to generative AI tools would be valuable, as it is plausible that the effects we uncovered could be amplified. Second, a deeper exploration of the specific skills founders need to fully leverage anomalies is crucial for understanding how to unlock their potential. Finally, while we concentrated on a key outcome for startups—customer acquisition—future research could assess additional outcomes, such as attracting VC funding or potential acquirers.

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., Camuffo, A., Gambardella, A., Gans, J. S., Scott, E., and Stern, S. (2025). *Bayesian Entrepreneurship: A Framework for Decision-Making in Uncertain Environments*. MIT 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.

Agrawal, A., McHale, J., and Oettl, A. (2023). Artificial intelligence and scientific discovery: A model of prioritized search. *Research Policy*, 53(5):104989.

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*, pages 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., 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*.

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*, pages 115–148. University of Chicago Press.

Conti, A., Peukert, C., and Roche, M. P. (2024). Beefing it up for your investor? engagement with open source communities, innovation and startup funding: Evidence from github. *Organization Science*, (ja).

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.

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

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).

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., Novelli, E., and Zenger, T. (2024). A scientific method for startups. *Journal of Management*, page 01492063231226136.

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. *Organization Science*, (forthcoming).

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.

Hoelzemann, J., Manso, G., Nagaraj, A., and Tranchero, M. (2024). The streetlight effect in data-driven exploration. Technical report, National Bureau of Economic Research.

Kim, S. (2023). Shortcuts to innovation: The use of analogies in knowledge production. *Columbia Business School*.

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

Kuhn, T. S. (1962). *The Structure of Scientific Revolutions*. University of Chicago Press, 1 edition.

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.

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*.

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

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

Ortoleva, P. (2012). Modeling the change of paradigm: Non-bayesian reactions to unexpected news. *American Economic Review*, 102(6):2410–2436.

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*.

Palan, S. and Schitter, C. (2018). Prolific.ac—a subject pool for online experiments. *Journal of Behavioral and Experimental Finance*, 17:22–27.

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*.

Toner-Rodgers, A. (2024). Artificial intelligence, scientific discovery, and product innovation. *arXiv preprint arXiv:2412.17866*.

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

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

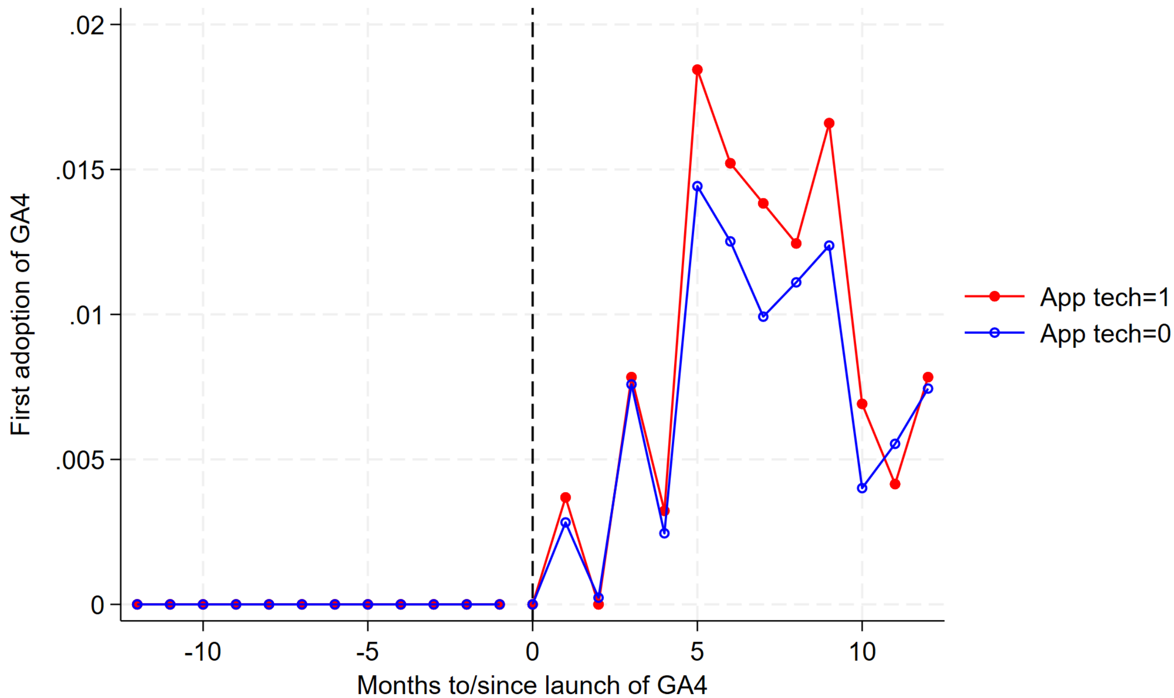


Figure 1: Adoption of Google Analytics 4 Over Time

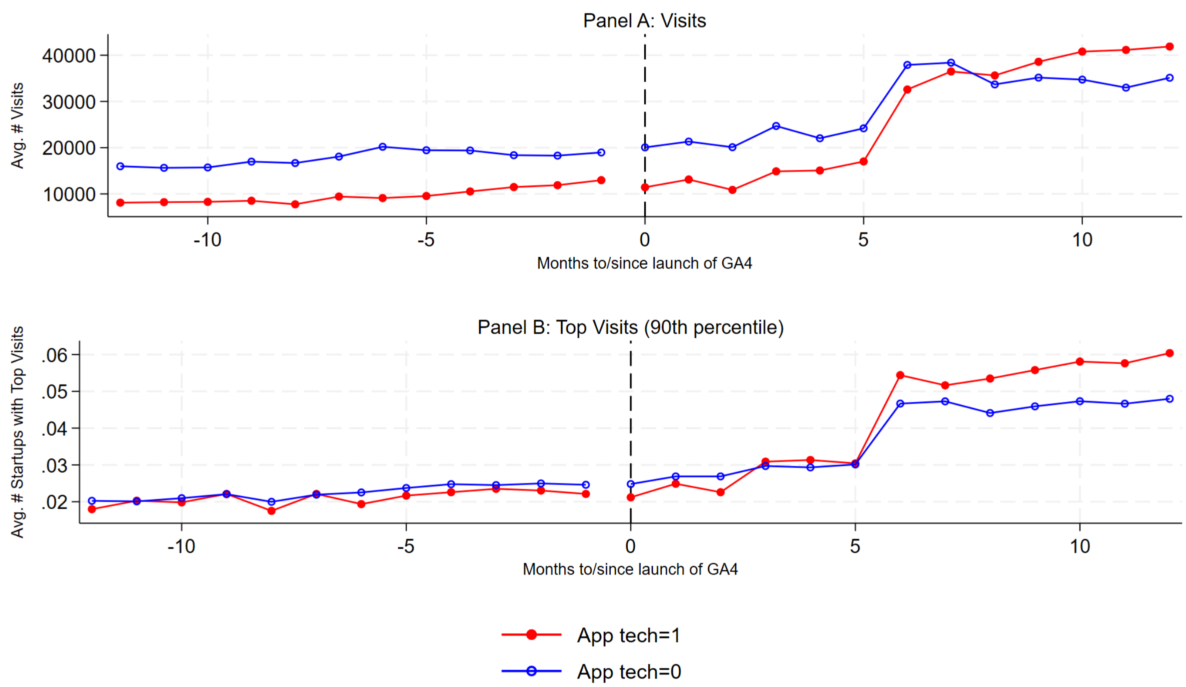


Figure 2: Website Visits Over Time

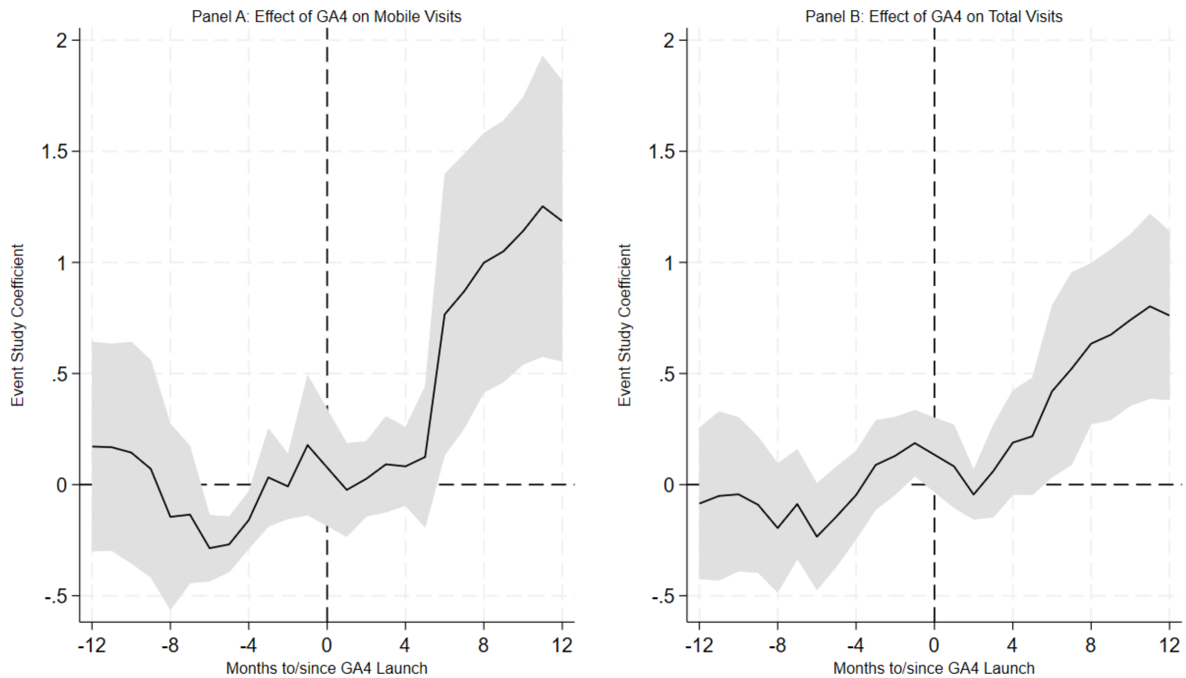


Figure 3: Event Studies for the Number of Total (Mobile) Visits

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.

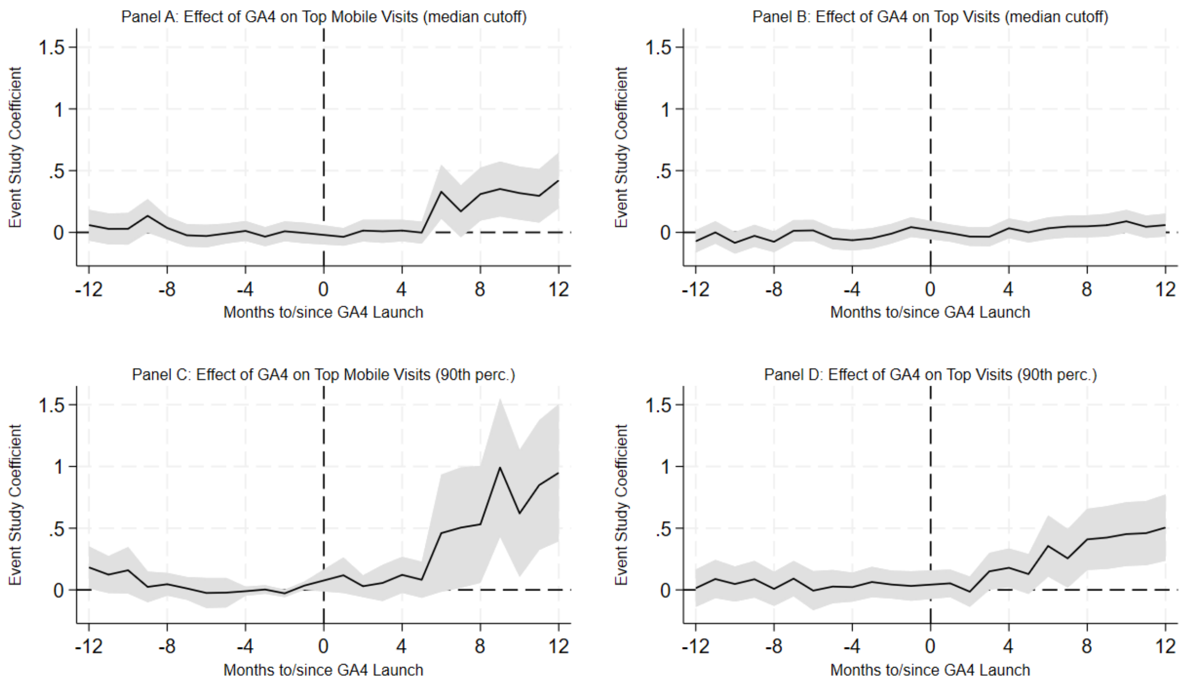


Figure 4: Event Studies for Top Total (Mobile) Visits

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.

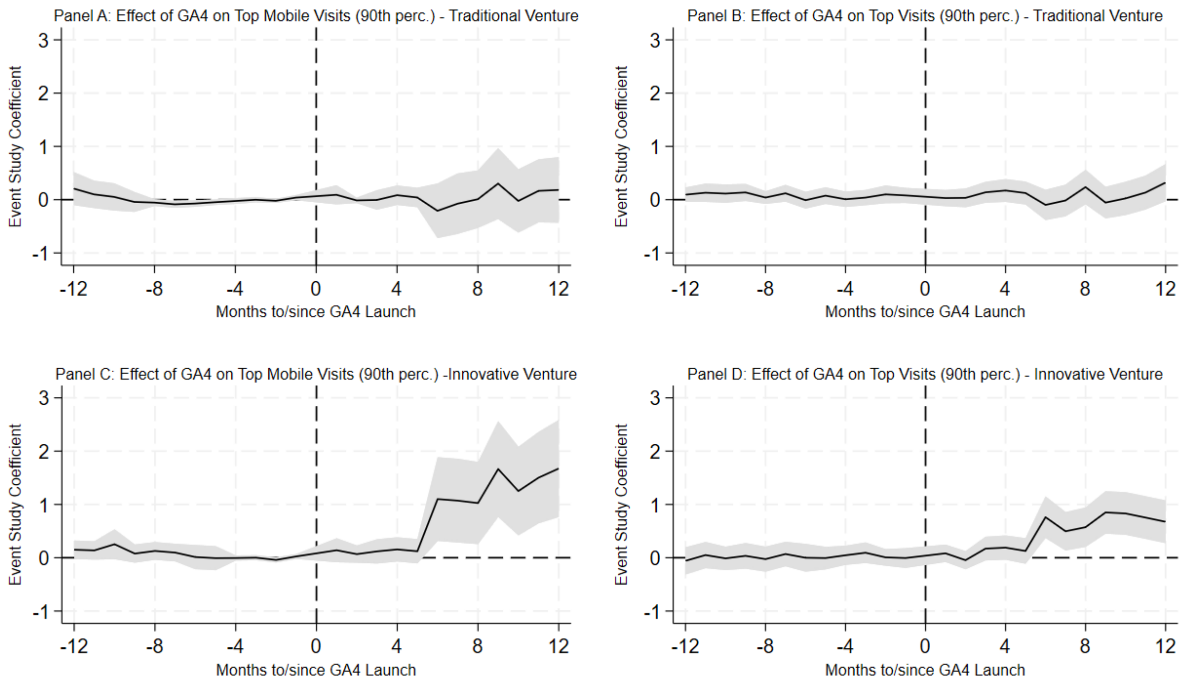


Figure 5: Event Studies for Top Total (Mobile) Visits: By the Innovativeness of a Startup's Venture

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 decile 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.

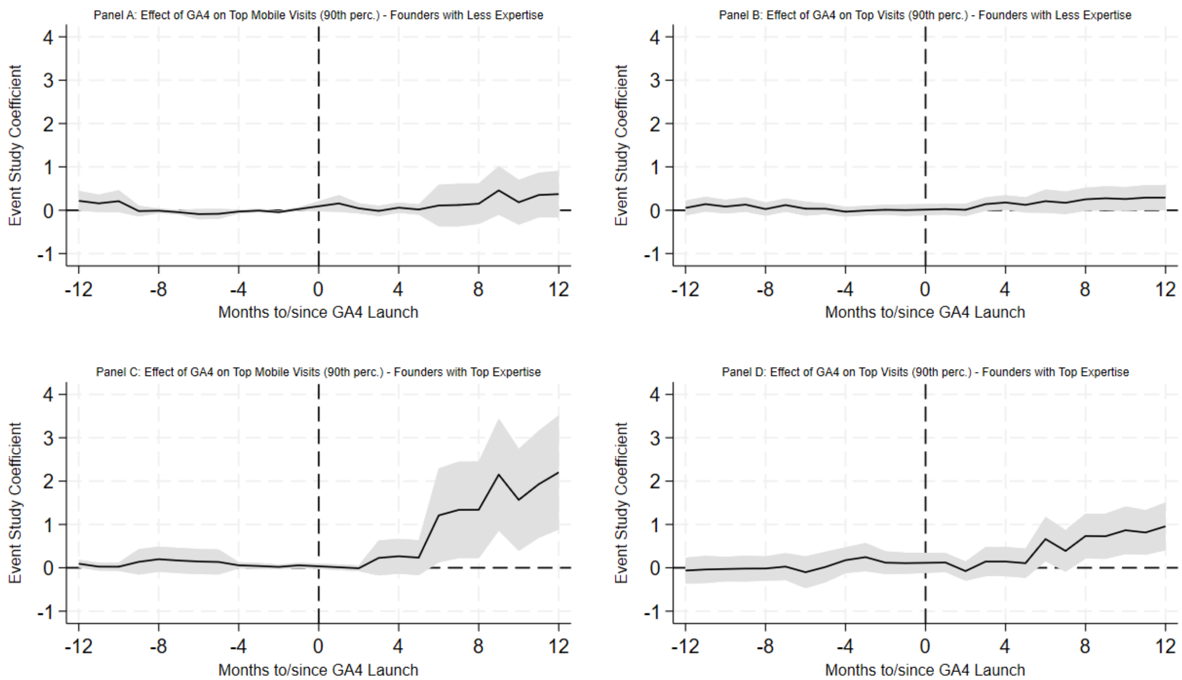


Figure 6: Event Studies for Top Total (Mobile) Visits: By the Expertise of a Founding Team

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 decile 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.

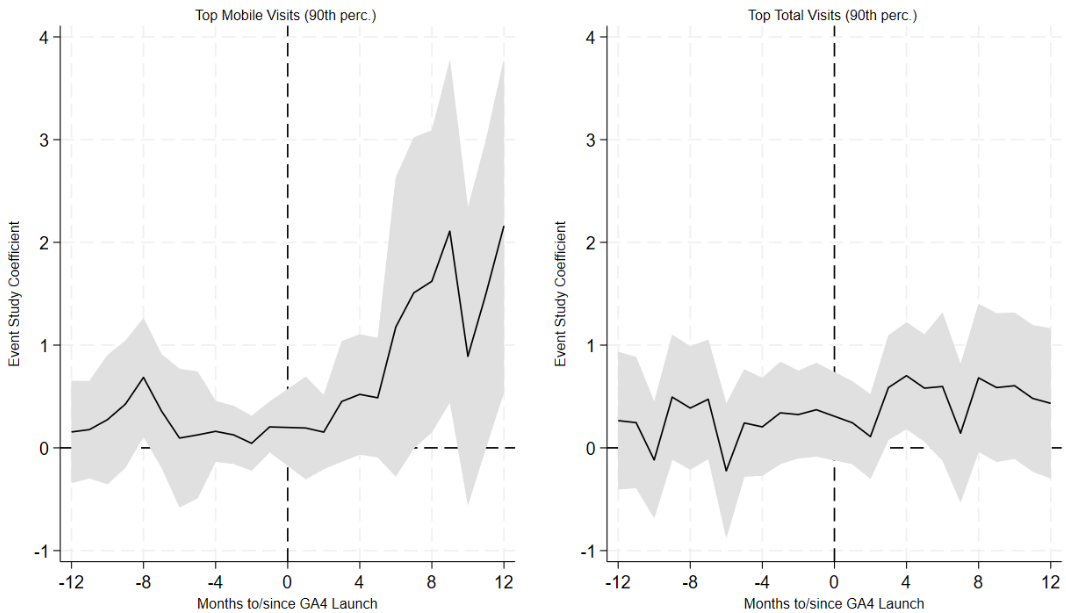


Figure 7: The Joint Effect of A/B Testing and Machine Learning

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 decile 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.

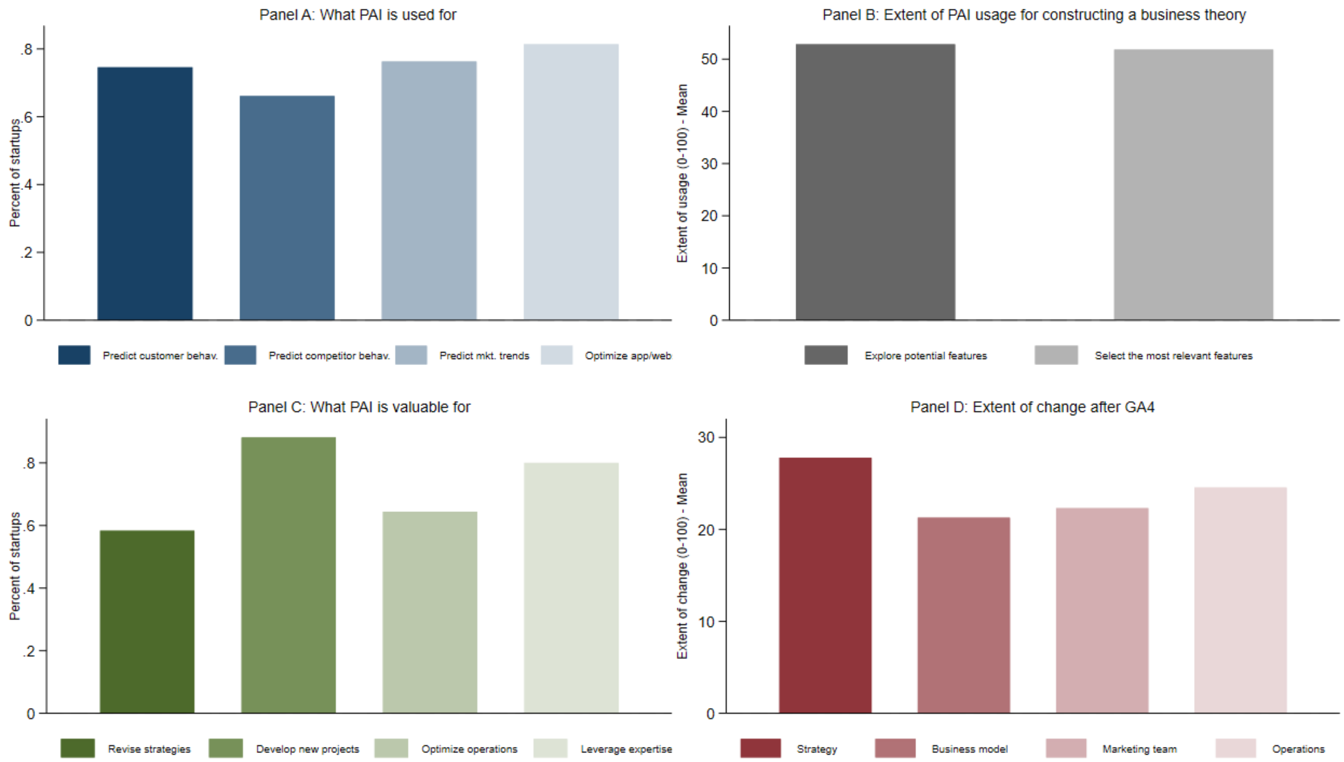


Figure 8: The Role of Predictive AI Tools

Panel A summarizes respondents' answers to the survey question: *'How frequently do you use predictive AI tools for the following tasks?'* The tasks include: predicting customer behavior, predicting competitor behavior, predicting market trends, and optimizing app/website configurations. Respondents were asked to choose from the following options: *'Never'*, *'Sometimes'*, *'About half the time'*, *'Most of the time'*, and *'Always'*. We report the percentage of startups that indicated using predictive AI for the listed tasks at least 'Sometimes.' Panel B summarizes respondents' answers to the question: *'How likely are you to use the predicted insights [...] to perform the following?'*. Respondents were asked to rank the options *'Select potential features impacting sales and other performance metrics'* and *'Determine which of the selected features have the strongest impact on sales and other performance metrics'* on a scale from 0 to 100. We report the mean rankings. Panel C summarizes respondents' answers to the question: *'Based on your experience, what are predictive AI tools most valuable for?'*. The tasks listed are: develop projects in areas where you can leverage your expertise, develop innovative projects, optimize operational efficiency, and revise company strategies. Respondents were asked to choose from the following options: *'Strongly disagree'*, *'Somewhat disagree'*, *'Neither agree nor disagree'*, *'Somewhat agree'*, and *'Strongly agree'*. We report the percentage of startups that at least somewhat agreed. Panel D summarizes respondents' answers to the question: "When Google Analytics 4 (GA4) was released in October 2020, to accommodate the new features of GA4, to what extent did you modify". The options given were: *'Your business model and/or product'*, *'Your marketing team'*, *'Your business operations'*, *'Your business strategy'*. Respondents were asked to provide a value on a scale from 0 to 100. We report the mean rankings.

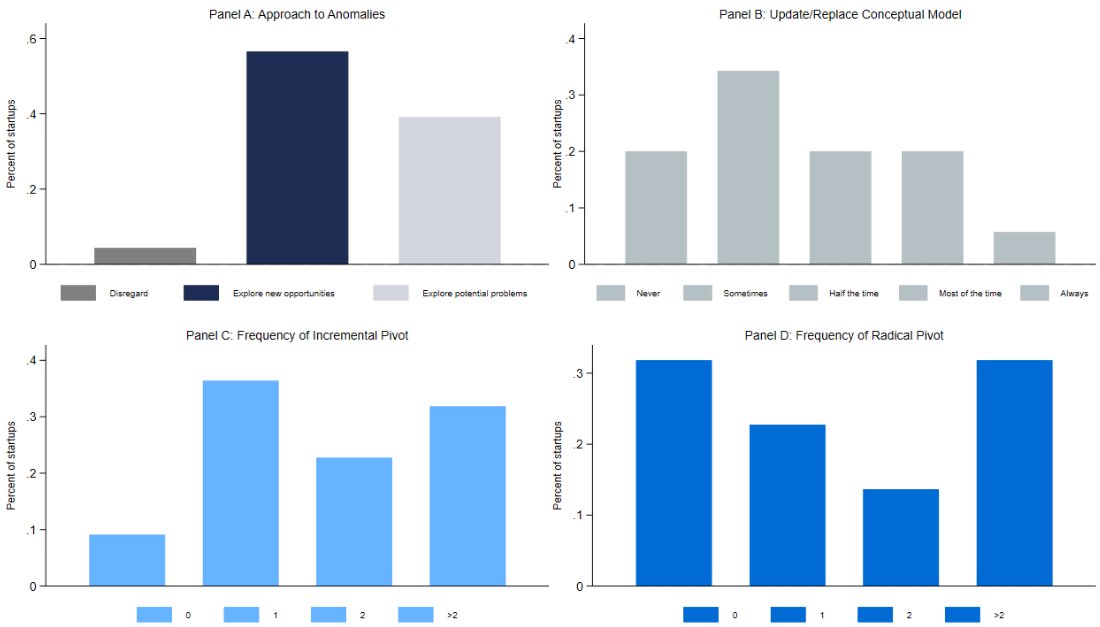


Figure 9: The Role of Anomalies

Panel A summarizes survey respondents' answers to the following survey question: *'If you detect anomalies in your data, how do you usually approach them?'*. The options provided are: *'Likely to disregard them, as they might be random isolated cases'*, *'Likely to explore them, as they might signal potential new opportunities'*, *'Likely to explore them, as they might signal problems with my business idea'*. We report the percentage of startups selecting each option. Panel B summarizes respondents' answers to the question: *'Lastly, we are interested in understanding how you structure your decision making process. When developing your business idea, how do you act and reason?'*. Respondents were asked to rank the option *'I use information from data anomalies to update my existing conceptual model or to frame a new one'* on a scale from 1 (never) to 5 (always). Panels C and D summarize respondents' answers to the question: *'After exploring anomalies, how often have you actually...'*. In Panel C, we summarize respondents' answers to the option *'Incrementally changed your initial business idea (business model and/or product)'*, while in Panel D, we summarize respondents' answers to the option *'Radically changed your initial business idea'*. Respondents could choose between: *'Never'*, *'Once'*, *'Twice'*, and *'More than twice'*. In all panels, the sample is limited to founders having used GA4.

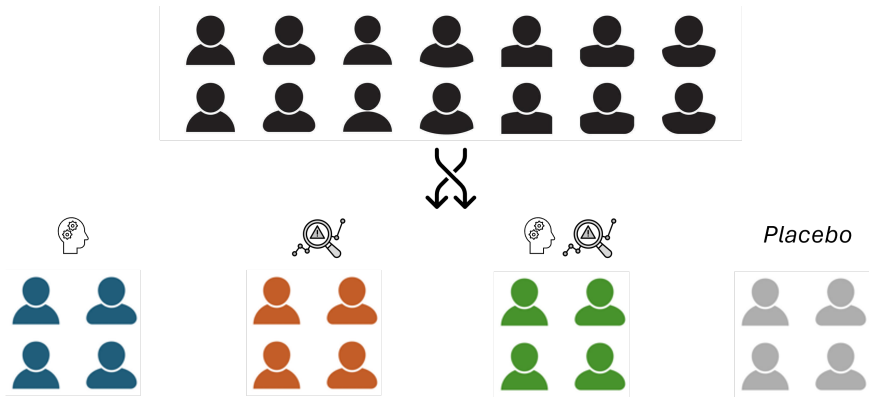
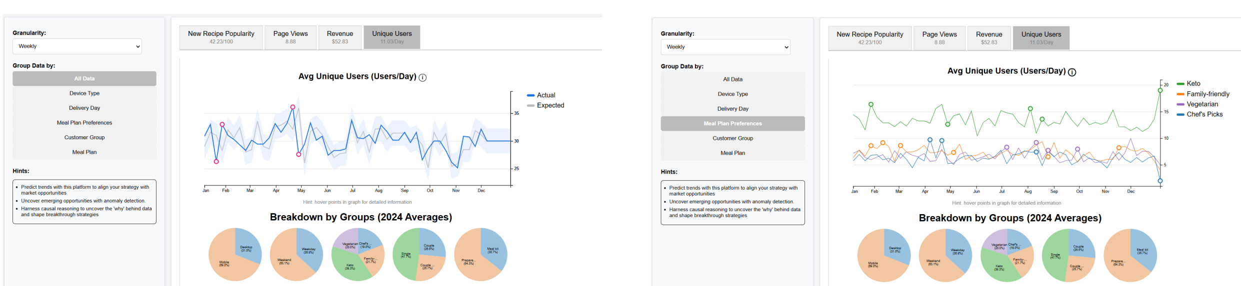


Figure 10: Experimental Design

This figure illustrates the three groups of Prolific entrepreneurs who were exposed to: (1) the theory-based reasoning intervention, (2) the anomaly detection intervention, and (3) both treatments, along with the control group, which received a placebo.

Data Analytics Platform



Panel A: Platform Version with anomalies

Panel B: Platform Version with anomalies and a breakdown example

Figure 11: Panel A shows the standard platform version. Panel B shows a screenshot of the platform version available to the group that received both interventions. This group saw a reminder about the importance of causal reasoning and circles highlighting anomalies.

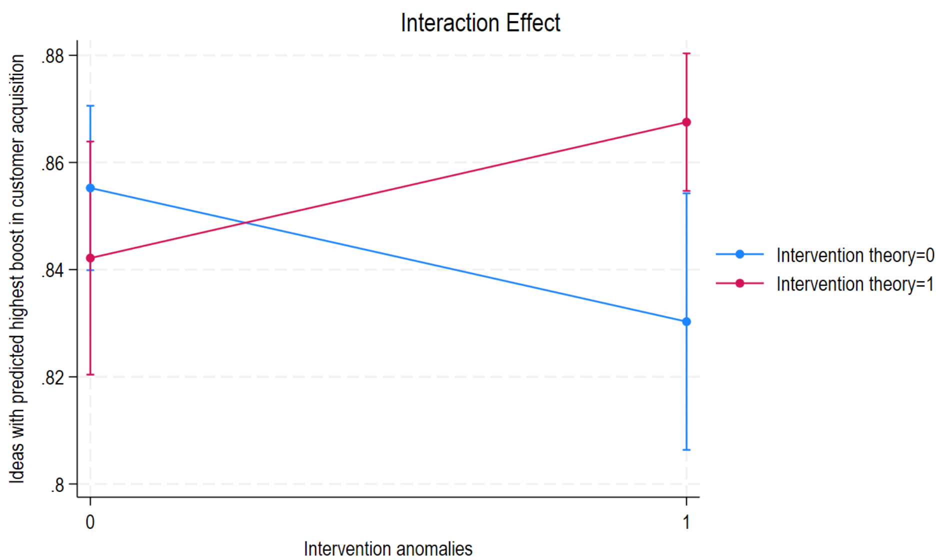


Figure 12: Impact of the Interventions on the Predicted Probability of a Business Idea Leading to Increased Customer Acquisition

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)	(2)	(3)	(4)
	Mobile Visits	Total Visits	Mobile Visits	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)	(2)	(3)	(4)
	Mobile Visits	Total Visits	Mobile Visits	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 decile 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) Top Mobile Visits	(2) Top Total Visits	(3) Top Mobile Visits	(4) 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 decile 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) Top Mobile Visits	(2) Top Total Visits	(3) Top Mobile Visits	(4) 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 decile 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. *PostGA*4 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 decile 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 *PostGA*4 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 decile 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.

Table 9: Experimental Results: Impact of the Interventions on the Generation of Valuable Business Ideas

	Market Success Probability
Interventions:	
Intervention (Theory)	-0.102
	(0.110)
Intervention (Anomalies)	-0.189*
	(0.107)
Intervention (Theory $\times$ Anomalies)	0.393***
	(0.151)
Baseline Controls:	
Education	-0.071
	(0.078)
Usage GA4	-0.013
	(0.078)
Usage Generative AI	0.469***
	(0.157)
Usage Predictive AI	-0.024
	(0.127)
Theory (Causal Logic)	0.009
	(0.042)
Theory (Anomalies)	0.004
	(0.037)
Knowledge (Food Industry)	-0.048
	(0.040)
Startup Experience	0.013
	(0.041)
Entrepreneurship	0.022
	(0.043)
Full-time Job	0.054
	(0.081)
Computer Programming	0.166
	(0.172)
Gender	0.125*
	(0.073)
Other Controls:	
Completion Time	0.000*
	(0.000)
Ideas Count	0.075***
	(0.017)
Constant	0.356
	(0.394)
Observations	275
Prob > χ^2	0.000
Pseudo R-squared	0.009

Notes: We report the results from estimating a fractional logit model, given that the dependent variable is bounded between 0 and 1. The outcome is the probability that a participant's most promising idea enhances customer acquisition. Robust standard errors are in parentheses. Significance noted as: *p<0.10; **p<0.05; ***p<0.01.

Table 10: Experimental Results: Leveraging Prior Knowledge

	Leveraging Prior Knowledge
Interventions:	
Intervention (Theory)	-0.099
	(0.100)
Intervention (Anomalies)	-0.349***
	(0.102)
Intervention (Theory $\times$ Anomalies)	0.282**
	(0.139)
Baseline Controls:	
Education	0.070
	(0.073)
Usage GA4	0.002
	(0.080)
Usage Generative AI	0.067
	(0.272)
Usage Predictive AI	0.224*
	(0.122)
Theory (Causal Logic)	0.010
	(0.043)
Theory (Anomalies)	0.038
	(0.039)
Knowledge (Food Industry)	0.128***
	(0.042)
Startup Experience	-0.059*
	(0.035)
Entrepreneurship	-0.025
	(0.046)
Full-time Job	-0.121
	(0.077)
Computer Programming	-0.184
	(0.222)
Gender	0.000
	(0.073)
Other Controls:	
Completion Time	-0.000**
	(0.000)
Ideas Count	0.021
	(0.016)
Constant	3.010***
	(0.570)
Observations	275
Prob > χ^2	0.000
Pseudo R-squared	0.095

Notes: We report the results from quasi-maximum-likelihood Poisson regression, given that the outcome variable. The outcome is the probability that a participant's most promising idea enhances customer acquisition. Robust standard errors are in parentheses. Significance noted as: *p<0.10; **p<0.05; ***p<0.01.

Appendix A

A1 Theoretical framework: Proof of Intuition 2 in Section 2

Proof Intuition 2 proof When C_0 is far from C_0^*, the baseline theory mis-specifies a larger region of D_0, leading to a higher probability of $D'_{\text{opportunity}} \neq \varnothing$. Integrating these newly exposed mechanisms can yield substantial gains. By contrast, incremental entrepreneurs—whose baseline C_0 is already near C_0^*—are less likely to experience large expansions, making their expected payoff gain from anomalies strictly lower. ■

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) 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. 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) Mobile Visits	(2) Total Visits	(3) Mobile Visits	(4) 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 the Likelihood that a Startup's Number of Website Visits is Below the 25th Percentile Cutoff

	(1) Bottom qrt. Mobile Visits	(2) Bottom qrt. Visits
Post GA4 × Develops Apps	-0.0239***	-0.00388
	(0.00360)	(0.00539)
Technology Stack	-0.00137***	-0.00397***
	(0.0000330)	(0.0000508)
Startup FEs	Yes	Yes
Year-month FEs	Yes	Yes
Observations	2492033	2492033
R2	0.454	0.481
Mean D.V.	0.916	0.750

Notes: We report the results of linear probability models for the likelihood that a startup's number of mobile (column 1) and total (column 2) visits received in month t is below the 25th percentile cutoff. *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 Top Website Visits: Having Winsorized the Top 1% of Observations of the Number of Visits

	(1) Top Mobile Visits	(2) Top Total Visits	(3) Top Mobile Visits	(4) 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 decile 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 A5: 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 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 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 decile 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: 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 decile 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 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 decile 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 A9: 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 decile 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 A10: 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 A11: 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 A12: 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 decile 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 A13: 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 decile 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. To build the Bartik 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. 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.

Table A14: Descriptive Statistics - Treatment 0 (Control)

Variable	Description	Mean	Std. Dev.	Min	Max
Market Success Probability	Max probability that a business idea boosts customer acquisition	0.856	0.064	0.641	0.950
Leveraging Prior Knowledge	Extent to which prior knowledge is used to identify business ideas (0-100)	28.265	15.784	1	57
Education	Holds an undergraduate degree or higher	0.485	0.503	0	1
Usage GA4	Familiarity with Google Analytics 4	0.632	0.486	0	1
Usage Generative AI	Familiarity with Generative AI tools	1.000	0.000	1	1
Usage Predictive AI	Familiarity with Predictive AI tools	0.838	0.371	0	1
Theory (Causal Logic)	Adoption of causal reasoning (1-5)	3.603	0.964	1	5
Theory (Anomalies)	Adoption of anomaly detection (1-5)	3.147	1.083	1	5
Knowledge (Food Industry)	Knowledge of the food delivery industry	2.824	0.929	1	5
Startup Experience	Prior startup involvement (1-4)	2.441	1.042	1	4
Entrepreneurship	Experience with entrepreneurship (1-3)	2.044	0.742	1	3
Job	Currently employed full-time	0.750	0.436	0	1
Computer Programming	Self-reported programming expertise	1.971	0.170	1	2
Gender	Dummy identifying male participants	0.279	0.452	0	1
Time (Minutes)	Time spent on tasks (converted from seconds)	58.105	159.34	8.716	1342.967
Innovation Count	Number of innovative ideas generated	8.544	2.488	2	10

Table A15: Descriptive Statistics - Treatment 1 (Theory)

Variable	Description	Mean	Std. Dev.	Min	Max
Market Success Probability	Max probability that a business idea boosts customer acquisition	0.843	0.087	0.500	0.950
Leveraging Prior Knowledge	Extent to which prior knowledge used to identify business ideas (0-100)	25.641	16.846	2	59
Education	Holds an undergraduate degree or higher	0.609	0.492	0	1
Usage GA4	Familiarity with Google Analytics 4	0.641	0.484	0	1
Usage Generative AI	Familiarity with Generative AI tools	0.969	0.175	0	1
Usage Predictive AI	Familiarity with Predictive AI tools	0.875	0.333	0	1
Theory (Causal Logic)	Adoption of causal reasoning (1-5)	3.594	0.955	1	5
Theory (Anomalies)	Adoption of anomaly detection (1-5)	3.266	0.913	1	5
Knowledge (Food Industry)	Knowledge of the food delivery industry	2.953	0.967	1	5
Startup Experience	Prior startup involvement (1-4)	2.359	0.998	1	4
Entrepreneurship	Experience with entrepreneurship (1-3)	2.109	0.758	1	3
Job	Currently employed full-time	0.766	0.427	0	1
Computer Programming	Self-reported programming expertise	2.000	0.000	2	2
Gender	Dummy identifying male participants	0.344	0.479	0	1
Time (Minutes)	Time spent on tasks (converted from seconds)	117.884	288.274	10.4	1474.017
Innovation Count	Number of innovative ideas generated	8.563	2.436	1	10

Table A16: Descriptive Statistics - Treatment 2 (Anomalies)

Variable	Description	Mean	Std. Dev.	Min	Max
Market Success Probability	Max probability that a business idea boosts customer acquisition	0.826	0.112	0.446	0.946
Leveraging Prior Knowledge	Extent to which prior knowledge used to identify business ideas (0-100)	20.484	12.359	1	46
Education	Holds an undergraduate degree or higher	0.484	0.504	0	1
Usage GA4	Familiarity with Google Analytics 4	0.548	0.502	0	1
Usage Generative AI	Familiarity with Generative AI tools	1.000	0.000	1	1
Usage Predictive AI	Familiarity with Predictive AI tools	0.903	0.298	0	1
Theory (Causal Logic)	Adoption of causal reasoning (1-5)	3.468	0.900	1	5
Theory (Anomalies)	Adoption of anomaly detection (1-5)	3.226	0.982	2	5
Knowledge (Food Industry)	Knowledge of the food delivery industry	2.903	0.970	1	5
Startup Experience	Prior startup involvement (1-4)	2.435	1.002	1	4
Entrepreneurship	Experience with entrepreneurship (1-3)	2.258	0.745	1	3
Job	Currently employed full-time	0.677	0.471	0	1
Computer Programming	Self-reported programming expertise	1.984	0.127	1	2
Gender	Dummy identifying male participants	0.274	0.450	0	1
Time (Minutes)	Time spent on tasks (converted from seconds)	43.157	45.268	7.85	350.333
Innovation Count	Number of innovative ideas generated	8.145	2.598	1	10

Table A17: Descriptive Statistics - Treatment 3 (Theory and Anomalies)

Variable	Description	Mean	Std. Dev.	Min	Max
Market Success Probability	Max probability that a business idea boosts customer acquisition	0.862	0.060	0.640	0.948
Leveraging Prior Knowledge	Extent to which prior knowledge used to identify business ideas (0-100)	24.802	15.462	1	57
Education	Holds an undergraduate degree or higher	0.630	0.486	0	1
Usage GA4	Familiarity with Google Analytics 4	0.617	0.489	0	1
Usage Generative AI	Familiarity with Generative AI tools	0.975	0.156	0	1
Usage Predictive AI	Familiarity with Predictive AI tools	0.889	0.316	0	1
Theory (Causal Logic)	Adoption of causal reasoning (1-5)	3.469	0.896	1	5
Theory (Anomalies)	Adoption of anomaly detection (1-5)	3.309	1.080	1	5
Knowledge (Food Industry)	Knowledge about the food delivery industry	2.938	0.966	1	5
Startup Experience	Prior startup involvement (1-4)	2.346	1.027	1	4
Entrepreneurship	Experience with entrepreneurship (1-3)	2.000	0.837	1	3
Job	Currently employed full-time	0.790	0.410	0	1
Computer Programming	Self-reported programming expertise	1.975	0.156	1	2
Gender	Dummy identifying male participants	0.272	0.448	0	1
Time (Minutes)	Time spent on tasks (converted from seconds)	59.357	128.131	12.85	1176.95
Innovation Count	Number of innovative ideas generated	8.210	2.533	1	10

A4 Lab experiment details

Here below we report the Task that Profilic participants were asked to perform. Each participant, was exposed to short videos, depending on their treatment arm.

Task Instructions

Dear Participant,

You will now watch one or more short videos providing essential information. After watching the videos, you will have 40 minutes to complete your tasks, which will involve writing.

Your attentiveness will be assessed through ad hoc questions based on the instructions provided in the videos.

In this project, you will take on the role of the CEO of **Yellow Apron**, a food delivery startup founded in 2023 and based in California, U.S.A. **Yellow Apron** specializes in delivering pre-portioned ingredients and chef-inspired recipes, making meal preparation easy and enjoyable. The company also offers pre-cooked meals for added convenience, with a menu that includes vegetarian, wellness-focused, and family-friendly options.

All ingredients are sustainably sourced, and customers can subscribe to flexible plans tailored to their dietary preferences.

Your task is to analyze historical data about **Yellow Apron** from 2024, presented on a custom platform that you will be allowed to access. Using basic data analytics skills, you will propose up to 10 innovations (related to business offer(s) and/or technology) that **Yellow Apron** could implement to maximize customer acquisition, revenue, and consumer engagement.

Your proposals might reference specific trends in the data (e.g., increases, decreases, or spikes) and may incorporate your own insights or experience, which you should clearly describe.

Incentive: The two best innovations, as judged by the project team, will each win a $100 prize.