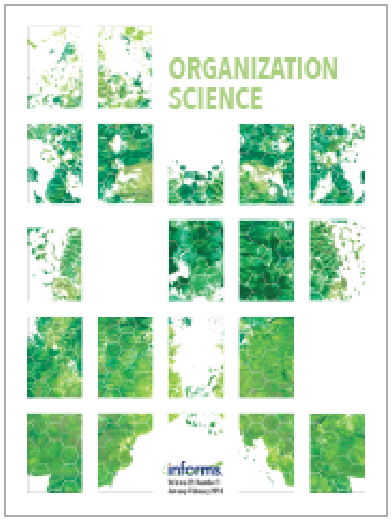


Organization Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

# Beefing IT Up for Your Investor? Engagement with Open Source Communities, Innovation, and Startup Funding: Evidence from GitHub

Annamaria Conti, Christian Peukert, Maria Roche

To cite this article:
Annamaria Conti, Christian Peukert, Maria Roche (2025) Beefing IT Up for Your Investor? Engagement with Open Source Communities, Innovation, and Startup Funding: Evidence from GitHub. Organization Science 36(4):1551-1573. https://doi.org/10.1287/orsc.2023.18348

informs
https://pubsonline.informs.org/journal/orsc

ORGANIZATION SCIENCE
Vol. 36, No. 4, July–August 2025, pp. 1551–1573
ISSN 1047-7039 (print), ISSN 1526-5455 (online)



# Beefing IT Up for Your Investor? Engagement with Open Source Communities, Innovation, and Startup Funding: Evidence from GitHub

**Annamaria Conti,[a] Christian Peukert,[b] Maria Roche[c,]***

[a] IE Business School, 28046 Madrid, Spain; [b] Faculty of Business and Economics (HEC), University of Lausanne, Quartier Chamberonne, 1015 Lausanne, Switzerland; [c] Harvard Business School, Boston, Massachusetts 02163

*Corresponding author

**Contact:** annamaria.conti@ie.edu, https://orcid.org/0000-0002-1888-2449 (AC); christian.peukert@unil.ch, https://orcid.org/0000-0003-3997-8850 (CP); mroche@hbs.edu, https://orcid.org/0000-0003-4941-5402 (MR)

**Received:** November 6, 2023
**Revised:** June 17, 2024; November 19, 2024
**Accepted:** November 19, 2024
**Published Online in Articles in Advance:** March 7, 2025

https://doi.org/10.1287/orsc.2023.18348



**Abstract.** We study the engagement of nascent firms with open source communities and its implications for innovation and attracting funding. To do so, we link data on 160,065 U.S. startups from Crunchbase to their activities on the open source software development platform GitHub. In a matched sample of firms with and without GitHub activities, difference-in-differences models reveal a substantial increase in the likelihood of being funded after early stage startups engage with open source communities on GitHub. This relationship is weaker for firms that employ GitHub for internal development only. Startups developing novel technologies tend to benefit more from engaging with open source communities, unlike those in highly competitive environments. This heterogeneity highlights a potential trade-off between engaging with open source communities and appropriability. To provide insight regarding mechanisms, we classify startups' technology use-cases on GitHub using machine learning and exploit data on product launches. Our results from these additional analyses support the notion that one important channel potentially driving our findings is the access to external knowledge for technology development provided by open source communities. Engaging with these communities may thereby aid startups in innovating and creating a (minimum) viable product.



**Funding:** A. Conti and C. Peukert received financial support from the Swiss National Science Foundation [Project IDs 100013_188998 and 100013_197807]. M. Roche received financial support from the Harvard Business School Division of Research and Faculty Development.
**Supplemental Material:** The online appendix is available at https://doi.org/10.1287/orsc.2023.18348.

**Keywords:** startups • knowledge • open source communities • GitHub • machine learning • venture capital • innovation

> There is a huge open source community of code and developers to work with and draw inspiration from on GitHub. To build better developer tools, we need a direct line to it. GitHub accelerates production, progress, and connections, bringing us closer to our users. (Developer Advocate, Stripe)[1]

## 1. Introduction

Open source software, with its freely accessible and modifiable source code, has become integral to modern technologies, powering operating systems (e.g., Linux, Android), web browsers (e.g., Chrome, Safari), and Artificial Intelligence (AI) (e.g., Spark, TensorFlow). Unlike closed source software, which is controlled by a single entity, open source projects are maintained by communities of contributors, including individual developers, corporations, universities, and government agencies. With the growing importance of open source software, the communities behind it have also become increasingly significant as knowledge suppliers. In fact, recent survey results suggest that firms rely twice as much on open source for external supply of knowledge than patents.[2] Provided how much firms and society depend on open source (Greenstein and Nagle 2014), understanding the dynamics of engagement with open source communities (OSCs)[3] appears critical. Thus far, much of the work examining use of open source has focused on mature firms or those startups that sell open source products or services (Nagle 2018). Moreover, many of the studies in this space rely on small scale or qualitative data (Shah 2006, Stam 2009, Ewens and Marx 2018) and identify how firms organize for open source (Germonprez et al. 2017) or how its use relates to business models (Dahlander and Magnusson 2008), rather than how firms' engagement with OSCs may impact achieving important performance milestones and innovation more broadly.

In this paper, we examine if engaging with OSCs can help nascent firms raise capital and to what extent access to open source code that contributes to firm innovation is a driver of this relationship. These are crucial questions, given that the financing environment fundamentally shapes strategic choices very early in the life of a new venture (Hellmann and Puri 2002, Ewens and Marx 2018, Hochberg et al. 2018, Dushnitsky and Matusik 2019). Prior literature stresses the importance of both the founding team (Bernstein et al. 2017, Gompers et al. 2020) and the underlying technology of a venture (Kaplan et al. 2009) to attract funding. But where the knowledge pieces used to build technologies come from still largely remains an open question.

To contribute new empirical findings on the relationship between nascent digital firms' engagement with OSCs and raising funds, and their innovative output, we create a large data set that combines firm, technology, and performance information. First, we exploit data encompassing 160,065 U.S. startups listed on Crunchbase that were founded between 2005 and 2020 as our initial sample. We also access information on startup activities on open source using data from GitHub, which describes itself as the place "where the world builds software." We use the public record of organization accounts on GitHub from GH Archive, a community-led project that logs GitHub activities since 2011. To shed light on potential innovation outcomes, we augment these data with information on startup product launches available from Product Hunt, a website tracking technology product launches. The combination of these three data sets provides us with information on the industry, investors, and total amount of funds a firm raises, as well as the type and nature of activities a firm engages in on GitHub—if a firm engages on the platform—and the technologies and products it develops.

The first question we address is whether engaging with OSCs makes startups more attractive to potential investors. To this end, we estimate a difference-in-differences model, assessing how the likelihood of being funded changes after startups engage with external repositories on GitHub,[4] relative to a matched sample of startups developing comparable technologies, with similar human capital, founded in the same year and location. We saturate this model with a host of fixed effects to hold constant fixed differences among startups and to account for technology shocks and life cycle differences. Our results suggest that engaging with OSCs on GitHub is associated with an increase in the likelihood of receiving funding by at least 36%.

Concerns about omitted variables and selection are warranted in our setting. To address those, we first assess the potential confound of a startup engaging with GitHub for internal software development rather than for interacting with OSCs. To do so, we augment our regression specification with a measure for a startup's engagement with internal repositories it directly controls. The results from this exercise confirm our baseline findings.

Next, we exploit heterogeneity in the expected payoffs for digital startups from engaging with OSCs (Fosfuri et al. 2008). Two feasible determinants of the expected payoffs in this context may be the type of technology (complex and novel) and market conditions that a startup faces (level of competition). In particular, the benefits of relying on input from open source may offset appropriability concerns when digital startups are developing novel technologies that rely on a very complex set and uncharted space of knowledge. Conversely, when the level of competition is high, and thus either concerns about appropriability are potentially more relevant or technology is more readily available, the relationship should be weakest (Lee and Kim 2024). Our results confirm these conjectures. Consistent with these findings, we further observe that although startups operating in the Platform domain have the highest propensity to engage with OSCs on GitHub, AI and Blockchain startups benefit the most when it comes to attracting funds. As the latter startups develop more novel and complex technologies (Jordan and Mitchell 2015), open source might help them bring these technologies to the market faster and achieve better performance outcomes (Franco et al. 2009). The need for speed and codevelopment may outweigh the costs associated with limited ability to protect intellectual property (IP) because, particularly, in the digital context, available IP protection mechanisms are relatively weak or difficult to execute (Gans and Stern 2003).

Overall, we can establish a robust relationship between startups' engagement with OSCs on GitHub and achieving funding milestones, and the expected payoffs appear highest for firms involved with developing the more complex and novel technologies. From this, we expect that the main correlation between OSC engagement and funding is primarily driven by the enhanced ability of startups to innovate and create a minimum viable product. To assess whether engaging with OSCs is correlated with innovative outcomes, we follow the classic Schumpeterian definition of innovation—commercialization of invention—and use product launches as the closest available proxy in our analysis (Schumpeter 1934). Our results using product launches on Product Hunt suggest that engaging with OSCs on GitHub can, indeed, contribute to completing a minimum viable product.

Theoretically, our motivation to examine the relationship with innovation builds on recent work suggesting that open source software has become evermore critical to innovation and is set to increase in its importance over the next years (Nagle 2019, Conti et al. 2023). One prominent example of a technology where this is visible is machine learning, which has been lauded as a fundamental general-purpose technology with the potential to shape growth and firm productivity over the next decades (Agrawal et al. 2019). However, in contrast to past technology revolutions (e.g., semiconductors, gene-editing), the adoption and advancement of core machine learning technologies are driven, to a large extent, by open source tools.

Naturally, it is also possible that GitHub serves as an endorsing entity to increase visibility (Rysman and Simcoe 2008), as has been shown in other "closed" contexts (Conti et al. 2013a, b; Hsu and Ziedonis 2013), rather than as a means to innovate and create new technologies. To rule out this alternative explanation, we distinguish between different technology use-cases that startups interact with on GitHub. Using natural language processing and machine learning methods (Furman and Teodoridis 2020, Miric et al. 2023, Tranchero 2023), we classify the external repositories that startups interact with according to the following use-cases: Software Development/Backend (SD/BE), Machine Learning (ML), Application Programming Interface (API), and User Interface (UI). We find that startups derive positive gains from interacting with all types of external repositories, although there are important differences across domains. The large spectrum of technology use-cases which startups interact with on GitHub provides a first indication that engaging with OSCs may provide critical access to relevant knowledge. To further support this notion, we investigate the relationship with a specific activity on GitHub that should demonstrate being active, but not alter the actual technology a startup produces: software documentation in "readme files." Here, we provide suggestive evidence that investors do not react to "cosmetic" changes that startups add to the often nontechnical documentation of their codebase. As such, this finding provides further support for a knowledge production explanation by which open source does not merely act as an amplifier of visibility.

Finally, in closing our investigation, we show that the relationship we have uncovered is especially strong when startups attract funds from venture capitalists (VCs) and successful investors. Such investors have been shown by the literature to be particularly inclined to value a startup's technology over other aspects (Conti et al. 2013a) and bring other types of capital (Shane and Stuart 2002).

The totality of our results carries important implications for entrepreneurs and policy-makers alike. For one, our study contributes to increasing our understanding of the role of a particular channel through which startups can access outside knowledge—open source software—and how engagement with OSCs can matter for innovation, which we capture by examining product launches and by the ability to attract funding. Our study thereby extends the literature that analyzes the role of open source for firm productivity (Nagle 2018, 2019; Shah and Nagle 2019), applying it to entrepreneurship and innovation (Wright et al. 2023). Critically, we highlight that engagement with OSCs can help digital startups innovate and attract investors, particularly VCs and successful investors, in early stages. However, adding on prior work suggesting an association between startup valuation and open source software (Lin and Maruping 2022), we find that the extent to which firms can leverage OSCs depends on the business model and use-cases they operate in.

For another, we contribute to the entrepreneurial finance literature that has investigated whether VCs invest in the founding team or the technology (Kaplan et al. 2009, Bernstein et al. 2017, Gompers et al. 2020). Our results suggest that an advantage of OSC engagement lies in accessing knowledge relevant for innovation with implications for financing. Indeed, our findings seem to indicate that startups are "beefing up" their products by relying on inputs from OSCs and, by doing so, are increasing their likelihood of receiving early stage financing. Finally, given increasing shifts toward VC investment in software (Lerner and Nanda 2020), and the difficulties associated with designing suitable support programs for entrepreneurial activity (Lyons and Zhang 2018), our results may provide useful insight for policies aiming at fostering innovation and the success of new ventures.

## 2. Guiding Framework

In this paper, we examine whether engagement with OSCs helps nascent firms raise capital, and whether any relationship might be driven by enhanced innovation. This is a crucial issue, as early funding decisions have long-term impacts on a venture's success (Hellmann and Puri 2002, Hochberg et al. 2018, Ewens and Marx 2018, Dushnitsky and Matusik 2019). Although prior research underscores the importance of a startup's core technology in securing funding (Kaplan et al. 2009), the role of engagement with OSCs, a form of external knowledge, remains underexplored. In our study, we will examine this role in the specific context of digital startups.

### 2.1. Engagement with OSCs and Funding

At a baseline, engagement with OSCs could both increase and decrease a startup's chances of securing funding relative to firms that do not engage in OSCs. On the one end, engagement with OSCs may enable startups to lower development costs and accelerate innovation, leveraging contributions from a diverse talent pool, which can appeal to funders interested in cost-efficient, high-growth opportunities (Kortum and Lerner 2000, Baum and Silverman 2004). On the other hand, startups engaging with OSCs may also face challenges. Some investors perceive open source as a potential risk due to concerns about the appropriation of IP and limited monetization opportunities, which can impact a startup's value capture potential (Bonaccorsi and Rossi 2003). Thus, although engagement with OSCs can enhance a startup's appeal by enhancing the ability to innovate, it may also lead to valuation concerns, creating a nuanced effect on funding likelihood relative to firms that pursue closed, proprietary models.

In what follows, we dive into further detail on the potential link between an enhanced ability to innovate as a driver for raising funds. This is because funding organizations often seek ventures with a strong innovation component, as they typically offer competitive advantages that can lead to sustained market leadership (Hellmann and Puri 2002; Conti et al. 2013a, b). It is, however, possible that this innovation advantage may disappear if obtained through engagement with OSCs provided costs unique to the digital and open context.

## 2.2. Digital Startups and the Role of External Knowledge for Innovation

Over the past decades, the locus of innovation has experienced a shift, moving from big corporate R&D labs to smaller players (Arora et al. 2018) and from offline to online markets. As such, digital entrepreneurship has been gaining in importance for economies across the globe (Yoo et al. 2012). Shaped by the use of digital technologies, the environments in which such entrepreneurial and innovative activities take place have been suggested to differ from more traditional entrepreneurship and approaches to innovation (Foege et al. 2019).

Although traditional theory emphasizes internal knowledge development, which may, for example, aid in developing absorptive capacity (Cohen and Levinthal 1990), digital startups increasingly rely on external, open ecosystems to obtain knowledge (Chesbrough 2006). Recent studies provide compelling empirical evidence that, particularly in the digital technology context, engaging with and learning from others may lead to critical performance improvements (Roche et al. 2024). In fact, digital startups are uniquely positioned to benefit from external knowledge due to their need for rapid experimentation, resource constraints, and reliance on digital tools (Contigiani 2023).

## 2.3. The Role of OSCs for Startup Innovation

One ecosystem that digital startups may rely on is embodied by OSCs—collaborative groups that develop, maintain, and improve software or projects with publicly accessible and modifiable source code—which, for example, provide access to cutting-edge tools such as open source libraries, helping substantially reduce development times and costs (Haefliger et al. 2008, Nambisan 2017). Such access is essential for digital startups' approach of rapid iteration (West and Kuk 2016). The reduction is possible because open source requires that original material is made publicly accessible for others to use freely. OSCs are based on the principles of collaboration, transparency, and community-oriented development and enable a decentralized and collaborative approach to software development (Conti et al. 2023). Whereas traditional closed source software is developed by a single entity—often a commercial firm—open source projects are typically maintained by distributed communities of contributors. These contributors may range from individual developers donating their time and expertise to employees of large corporations who contribute to open source as part of their professional roles. Universities and research institutions also play a significant role in many areas of open source technology. Numerous widely used software projects, including operating systems like Linux, web servers like Apache, and programming languages like Python, have been the result of this work where users collectively solve problems valuable to the OSC and beyond. Large open source software (OSS) projects are often backed by commercial organizations. For instance, Chromium, a critical open source component for modern web browsers, is developed and maintained primarily by Google, Microsoft, and other firms (Haese and Peukert 2024). Additionally, governmental agencies participate in the open source ecosystem by contributing code and projects that serve both public and commercial interests (Hoffmann et al. 2024).

OSCs not only support the technical growth of startups but also their integration into broader innovation ecosystems and thereby potentially help overcome barriers to entry (Gruber and Henkel 2006). OSCs do so by fostering collaboration, knowledge sharing, and accessibility across diverse groups of contributors, including developers, corporations, universities, and government agencies (Wright et al. 2023). These communities lower barriers to entry by providing free access to cutting-edge technologies, which accelerates the development and adoption of new innovations. Open source projects promote modularity, enabling seamless integration with other technologies and encouraging rapid iteration. Furthermore, by operating through transparent, peer-reviewed processes, OSCs facilitate the cross-pollination of ideas and solutions, helping bridge gaps between different sectors and driving collective innovation (Nagle 2018). This open collaboration encourages widespread adoption, contributing to the growth and dynamism of the broader innovation ecosystem (Hoffmann et al. 2024). Today, over 90% of all Fortune 500 companies use open source products, and their role is increasingly relevant in the development of the software industry (Nagle 2019, Conti et al. 2023). Indeed, OSCs appear to have become a more important source than other well-studied traditional channels, such as standards or patents (see Figure A1 in the Online Appendix), across industries and firm sizes.

### 2.4. Potential Risks of Engaging with OSCs

Although OSCs provide valuable resources and opportunities for digital startups, they also introduce a unique set of risks (Almirall and Casadesus-Masanell 2010), particularly related to IP, competitive advantage, and governance. For one, OSS comes with a variety of licenses, such as the General Public License, Massachusetts Institute of Technology, Apache, and others, each with specific legal requirements. Some licenses require that derivative works also be open sourced. If a startup incorporates open source code without understanding the licensing terms, they may face legal challenges or be forced to open source their proprietary code. This could, especially in the context of digital technologies where traditional instruments such as patents are not as effective in protecting technologies as, for example, patents protecting drugs (Conti et al. 2023), seriously inhibit the ability to appropriate value created (Teece 1986, Laursen and Salter 2014, Buss and Peukert 2015), undermining a startup's competitive advantage or even their reason for being. Similarly, OSS is available to everyone, including competitors who may be part of the OSC. Startups that rely heavily on open source code risk losing their competitive edge, as competitors can build similar products using the same open source tools and frameworks (some of the code may have even been coproduced with competitors). This can make it harder for digital startups to differentiate their products in the market and hurt a new venture's ability to capture value. Ultimately, this lack of differentiation can undermine a new venture's appeal to investors, as limited barriers to entry and reduced potential for building a unique moat may deter funding.

### 2.5. The Cost-Benefit Trade-Off(s) of Engaging with OSCs

The associated costs and benefits of engaging with OSCs[5] suggest an important trade-off that digital startups need to make when doing so. For instance, because of their youth and limited resources, startups may lack the internal knowledge pieces required for the recombination process, yielding the ability to access external sources of ideas all the more critical (Roche et al. 2024). We expect these benefits to be more salient, especially when startups are in their earliest stages of technology development—that is, when startups' main focus lies in developing their core technologies and resolving technical issues (Tzabbar and Margolis 2017). In other words, we expect engagement with OSCs to be particularly critical when startups aim to innovate. Exposure to external sources of knowledge (Marinoni and Roche 2024), such as through engagement with OSCs, might thereby help startups solve their technical issues quickly and generate a first viable prototype to showcase to potential investors (Zahra and George 2002). However, new ventures are also more likely to run into adoption issues—integrating external knowledge into their existing technology stack (Roche et al. 2024)—and to suffer from exposure to appropriability concerns, especially provided their lack of necessary complementary assets to reap the rewards from their innovative efforts (Teece 1986).

New entrants are also more likely to face appropriability concerns because IP protection can be difficult to obtain in the context of digital startups. For one, traditional forms of protection, such as patents, are costly and relatively slow to receive and, even then, provide weak protection, especially for such small firms (Veugelers and Schneider 2018). Particularly in the digital context where insight into the backend of a technology can be found relatively easily (Roche et al. 2024), safeguards against reverse engineering are weak (Gans and Stern 2003). Other mechanisms, such as secrecy (Cunningham and Kapacinskaite 2024), are, similarly, difficult to execute. As such, "isolating mechanisms" (Rumelt 1984, p. 568) that contribute to a startup's competitive advantage are more limited. Speed or (offline) complementary assets become increasingly important (Boudreau 2010, Lee and Kim 2024), especially if the technologies being built are intended to be used on two-sided platforms or in winner-takes-all or most markets where network effects and economies of scale are prevalent (Katz and Shapiro 1994, Benzell et al. 2019, Garcia-Swartz and Campbell-Kelly 2019).

Taken together, in the context of digital entrepreneurship, *we expect that, on average, engagement with OSCs will have a positive impact on the entrepreneurial success, specifically the ability to raise earliest stage funding. We conjecture that this relationship is driven by the enhanced ability to innovate, given access to knowledge the OSC provides*.

However, engagement with OSCs introduces distinct risks related to IP and competitive differentiation, which may hamper a startup's ability to attract funds. The goal of this paper is to empirically examine how digital startups navigate this trade-off, as they represent a shift from traditional models of innovation and external knowledge acquisition. Our approach is thereby to first provide empirical evidence on the nature of the relationship between startup engagement with OSCs and achieving funding milestones. Second, we set out to unveil heterogeneity among different types of technologies that may benefit or even be harmed by such engagement. Third, we uncover underlying mechanisms—namely, whether startups participate in OSCs to innovate or just to increase their visibility. Finally, we provide some indication of which investors are more sensitive to these activities.

## 3. Data

To build our data set, we combine data on U.S. startups and their investors from Crunchbase with information on startups' GitHub activities available from GH Archive and the GitHub API and with information on technology products listed on Product Hunt.

### 3.1. Crunchbase

Crunchbase is an online directory that records fine-grained information on a large sample of startups, their founders, and their investors. As described in Conti and Roche (2021), a considerable portion of the data are entered by Crunchbase staff, whereas the remaining part is crowdsourced. Registered members can enter information into the database, which the Crunchbase staff successively reviews. Relative to databases such as VentureXpert and VentureSource, Crunchbase has the advantage of providing larger coverage of technology startups, as it also encompasses startups that did not raise venture capital. From Crunchbase, we extract information pertaining to all the recorded U.S. startups that were founded between 2005 and 2020. This amounts to 160,065 startups, for which we have data encompassing their founding dates, industry group keywords, location, financing rounds, and participating investors, as well as exit outcomes.

As shown in Table 1, approximately half of the startups (46%) are located in California, Massachusetts, and New York, reflecting the comparative advantage of these regions in entrepreneurship. Thirty-six percent of them raised at least one round of financing. Additionally, 8% of the startups were acquired as of December 2020, and 1.2% went public.

**Table 1.** Summary Statistics—Full Sample of Startups

| Variables | Mean | Std. Dev. | Min | Max | p50 |
|---|---|---|---|---|---|
| Raised funds | 0.357 | 0.479 | 0 | 1 | 0 |
| Launched product | 0.061 | 0.239 | 0 | 1 | 0 |
| IPO | 0.012 | 0.107 | 0 | 1 | 0 |
| Acquired | 0.081 | 0.273 | 0 | 1 | 0 |
| GitHub | 0.093 | 0.29 | 0 | 1 | 0 |
| Engaged with external GitHub repos | 0.084 | 0.28 | 0 | 1 | 0 |
| Top Team | 0.002 | 0.039 | 0 | 1 | 0 |
| AI | 0.039 | 0.194 | 0 | 1 | 0 |
| Data Analytics | 0.098 | 0.297 | 0 | 1 | 0 |
| Information Technology | 0.182 | 0.386 | 0 | 1 | 0 |
| Internet Services | 0.200 | 0.400 | 0 | 1 | 0 |
| Software | 0.356 | 0.479 | 0 | 1 | 0 |
| N. Industry Groups | 3 | 2 | 1 | 19 | 3 |
| Software share | 0.464 | 0.387 | 0 | 1 | 0.5 |
| California | 0.296 | 0.457 | 0 | 1 | 0 |
| Massachusetts | 0.045 | 0.207 | 0 | 1 | 0 |
| New York | 0.128 | 0.334 | 0 | 1 | 0 |

*Notes*. In the case of product launches, we restrict the sample to companies founded after 2011, given that Product Hunt started only in 2014. Std. Dev., standard deviation; p50, 50th percentile or median.

Although Crunchbase does not categorize startups into sectors, it provides industry group information for each of them.[6] There are approximately 40 distinct industry groups, and, on average, a startup is assigned three industry group keywords. Using this information, we develop a measure of how much a startup's technology is related to software. This measure is defined as the share of a startup's industry groups that are related to software (the number of software-related industry groups divided by the total number of industry groups describing a startup's technology). The groups related to software are: Apps, Artificial Intelligence, Consumer Electronics, Data and Analytics, Design, Financial Services, Gaming, Information Technology, Internet Services, Messaging and Telecommunications, Mobile, Payments, Platforms, Privacy and Security, and Software. As shown in Table 1, the mean of this index is 0.46.

### 3.2. GitHub Activities

GitHub is our next source of information. With 40 million public repositories in April 2021, GitHub was the largest host of source code[7] at the time of this study and has come to be known as the place "where the world builds software" (https://github.com/). Individuals and organizations use GitHub to host their own projects and upgrade them by using code and information from other existing projects, as well as to contribute to the projects of others. GitHub has a history of being backed by a number of high-profile investors (e.g., through a $100 million investment by Andreesen Horowitz) and was acquired by Microsoft for $7.5 billion in 2018.[8]

Publicly available code on GitHub often incorporates time- and cost-saving functionalities, enabling organizations to significantly accelerate their development processes. For instance, React.js, an open source JavaScript library, optimizes web application development with its virtual Document Object Model feature, allowing for efficient rendering and improved performance. Similarly, Docker offers an environment where applications can be developed, shipped, and run consistently across different setups, simplifying configuration and scaling, and thereby reducing the time to market for technologies. In addition to these functionalities, GitHub offers access to cutting-edge technologies, thanks to its diverse and global community of developers who contribute to open source projects. TensorFlow, for example, represents an open source library for AI and machine learning. The Ethereum blockchain platform enables access to new technologies, which aids in the creation of decentralized applications and smart contracts with enhanced security and transparency. The iterative and collaborative nature of open source development further enhances the value of code developed on the platform. The Linux project exemplifies this, as it is continuously refined, updated, and secured by a global community of developers, ensuring a high-performance operating system kernel.

GitHub offers personal user as well as organization accounts, the latter being the object of our analysis. The source code on GitHub is organized in *repositories*—that is, folders containing projects. The underlying version control system, git, effectively stores historical versions of files within a repository. When a user issues a *commit*, a snapshot of the file's contents is created and associated with a timestamp. The most frequently used way to interact with the repositories of other users is called *forking*. By forking, a user makes a copy of another user's repository, which is then integrated into the initial user's account. Forked repositories can be the foundation for further internal development. Forking plays a significant role in how knowledge is shared, created, and how collaboration occurs in OSCs. Note that all forks of public repositories are public and that it is not possible to change the visibility of a fork.[9]

We use the GitHub API to collect all organization accounts on GitHub. From these accounts, we extract the websites of the account owners. We use this information to link GitHub organization accounts to 14,881 Crunchbase company profiles. Over 60% of the startups with a GitHub organization account are described by Crunchbase's industry group keyword "software," and, perhaps not surprisingly, our software share index is higher for startups with GitHub accounts. We further gather time-variant information on the public events of all startups through GH Archive, which is a community-led project that provides a full record of the public timeline of users on GitHub since February 2011. This archive includes, among others, time-stamped data on events such as commits and forks—activities related to all own or external public repositories with which an organization account interacts.[10]

As reported in Table 1, 9% of the startups in our sample have an organization account on GitHub, whereas 8% have an organization account and have engaged in an *external activity* on GitHub—namely, interacted with an external repository that the startups do not directly control. The latter serves as our primary measure of *engagement with OSCs*. In our analysis, we will make the distinction between activities related to internal and external repositories that can be publicly observed on *github.com*. In contrast to *external* repositories, we define *internal* repositories as those that originate from and are held by the focal firm. In other words, the focal organization created these and can determine what happens to the original source code. Note that startups might interact with internal repositories for developing software without engaging with OSCs.

Among the events through which organizations interact with external repositories, forks are the largest portion (96%), followed by watch events (2%) and push events (1%).[11] This distribution suggests that startups may interact with external repositories mostly to develop their own technologies and less so to provide comments or contribute to other users' projects. As reported in Figure 1, there has been a stark increase in the number of startups' interactions with external repositories over time, which has tracked the increase in startups' overall activities on GitHub.

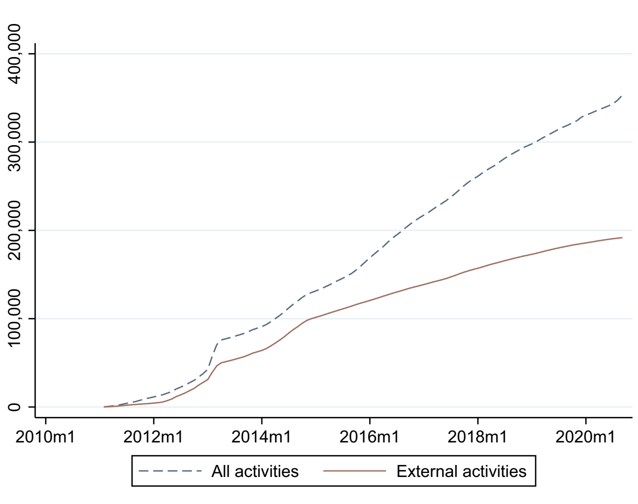
As suggested by recent work, machine learning methods are powerful in detecting patterns in large and complex data (Furman and Teodoridis 2020, Choudhury et al. 2021, Miric et al. 2023, Tranchero 2023). In this paper, we use supervised and unsupervised machine learning methods, which we describe in detail in the Online Appendix, to classify the external repositories with which the organizations interact through commits, pull requests, or forks according to their type. We identify repositories that pertain to SD/BE, ML, API, and UI. By doing so, we consider a comprehensive set of use-cases that are relevant for the development of digital technologies (Yoo et al. 2012).[12]

We report a descriptive representation of the output obtained from the algorithm in Figure A2 in the Online Appendix, displaying the most common words for each of the categories we consider. Additionally, we report in Table 2 descriptive statistics summarizing startups' external activities on GitHub. Here, we show that startups prevalently engage with external repositories related to SD/BE and API.

### 3.3. Product Hunt

We use data from Product Hunt to assess how a startup's activities on GitHub relate to innovation. Product Hunt is a website where innovators can post and share their new products. The website was founded in 2013

**Figure 1.** (Color online) Startups' Engagements with External and Total GitHub Repositories over Time




*Notes*. This figure displays patterns of activity on GitHub over time. We distinguish between external and all activities (which include external activities).

**Table 2.** Summary Statistics—Startups that Engaged with an External Repository on GitHub

| Variables | Obs | Mean | Std. Dev. | Min | Max | p50 |
|---|---|---|---|---|---|---|
| SD/BE | 13,413 | 0.646 | 0.478 | 0 | 1 | 0 |
| ML | 13,413 | 0.135 | 0.341 | 0 | 1 | 0 |
| API | 13,413 | 0.628 | 0.483 | 0 | 1 | 0 |
| UI | 13,413 | 0.277 | 0.448 | 0 | 1 | 0 |

*Notes.* The classification of repositories external to an organization's account was derived from implementing the machine learning algorithm described in the main text and in the Online Appendix. Obs, observations; Std. Dev., standard deviation; p50, 50th percentile or median.

by U.S. entrepreneur Ryan Hoover, admitted to Y Combinator in 2014, and subsequently received funds from several prominent venture capitalists, including Greylock. In 2016, it was acquired by AngelList.

The products that are posted on Product Hunt can be web apps, mobile apps, hardware products, games, and even books and podcasts (Conti and Santaló 2023).[13] We retrieved the entire data set of products featured on Product Hunt from January 2014 to September 2022. For this period, we were able to match startups with their product launches, using the startups' website information. As shown in Table 1, 6% of the startups founded after 2011 had launched a product on Product Hunt as of December 2022.

## 4. Empirical Specification

Ideally, to assess how engaging with OSCs is related to attracting funds, we would randomly assign startups to the treatment of interest. That way, we would address the concern that the relationship between a startup's engagement with OSCs on GitHub and achieving funding milestones may be confounded by technology or geographical shocks, a startup's technology characteristics, or a startup's technology life cycle. Implementing a field experiment that randomizes access to OSCs across startups and over time would be very costly and extremely challenging. In its place, we estimate a difference-in-differences model, where we compare startups that engage with OSCs on GitHub to randomly selected startups with similar observable characteristics. By doing so, we assess how the likelihood of being funded by a given period changes after startups begin to engage with OSCs on GitHub relative to the comparison group. Similar to Arts et al. (2018), we randomly select up to five startups that either do not have an organization account on GitHub or only engage in public activities related to their internal repositories (this occurrence is more rare) and are similar to the focal startup because they were founded in the same state and year and have a similar value of the software share index described above.[14] We additionally impose that a startup that engages with OSCs and its comparison group have similar human capital, as defined by whether (or not) a startup's founder or CXO[15] is highly ranked on Crunchbase's list of top people.[16] We adopt these criteria, given that Table 3 shows that startups engaging with OSCs on GitHub differ along these important dimensions from the other startups. By doing so, we compare, as much as possible, startups engaging with OSCs with (on average, four) comparable startups exposed to similar funding conditions and technology shocks, developing similar technologies, operating in a similar phase of the technology life cycle,[17] and characterized by comparable human capital.[18] We call the focal startup and its comparison startups an observation group $g$. Thus, we estimate:

$$Y_{igt} = \delta PostGitHub_{gt} \times EngagementWithOSC_i + \mu_i + \psi_{gt} + \tau_{it} + \varepsilon_{igt}, \quad (1)$$

where $Y_{igt}$ is an indicator that becomes one starting from the quarter $t$ when startup $i$, belonging to the observation group $g$, raises a first financing round. We focus on a startup's first round, as this is a fundamental funding milestone. Indeed, investors participating in this round typically provide the necessary network for raising subsequent rounds (Conti and Graham 2020). We conduct our analyses at the quarter level to mitigate the concern that even though we have monthly data on the round announcements, there could be delays

**Table 3.** Startups Engaging with OSCs on GitHub

| Variables | (1) Engagement with OSCs (0/1) |
|---|---|
| Accelerator | 0.0342*** |
| | (0.00605) |
| Top Team | 0.376*** |
| | (0.0370) |
| California | 0.0334*** |
| | (0.00353) |
| Massachusetts | 0.0309*** |
| | (0.00312) |
| New York | 0.0336*** |
| | (0.00205) |
| Artificial Intelligence | −0.00604 |
| | (0.0119) |
| Data Analytics | 0.0473*** |
| | (0.00579) |
| IT | 0.0216*** |
| | (0.00352) |
| Internet Services | 0.0206*** |
| | (0.00293) |
| Mobile | 0.00430 |
| | (0.00705) |
| Software | 0.0806*** |
| | (0.00339) |
| Observations | 160,065 |
| $R^2$ | 0.0431 |

*Notes.* We report the results from estimating the likelihood of being engaged with OSCs on GitHub. Standard errors (in parentheses) are clustered by founding year.

$^{***}p < 0.01$.

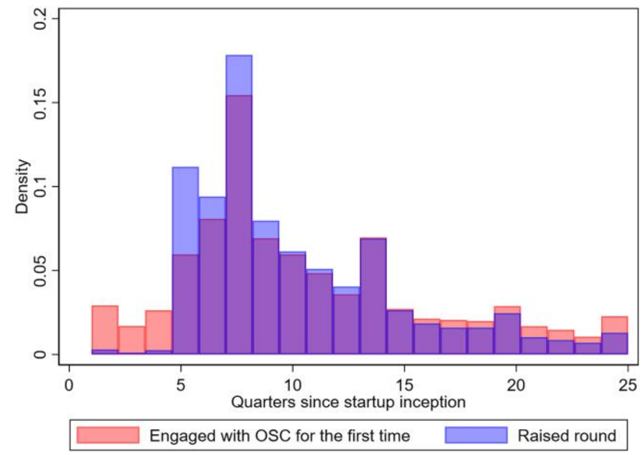
between the announcement and the receipt of funds. Moreover, we censor the sample in the second quarter of 2020, the date at which we retrieved the Crunchbase data set, and following Conti and Guzman (2021) and Amore et al. (2023), we observe each startup for a maximum of six years, as most startups raising a financing round do so within this time window. We use the same specification to assess innovation outcomes, as measured through the launch of a product on Product Hunt.

$EngagementWithOSC_i$ is our main variable of interest, which takes a value of one if startup $i$ engages in at least one public activity with external repositories on GitHub during the period we observe, and zero otherwise. In more fine-grained analyses, we also distinguish a startup's type of engagement with OSCs on GitHub. $PostGitHub_{gt}$ is a time-varying binary indicator that becomes one for all the startups in an observation group $g$ after startup $i$ engages in an external activity on GitHub. The coefficient of interest is $\delta$, which is associated with the interaction between $EngagementWithOSC_i$ and $PostGitHub_{gt}$. This coefficient measures the average change in the likelihood that a startup has raised a first financing round by quarter $t$ after it engages with OSCs on GitHub. In our main specifications, the effects of $EngagementWithOSC_i$ and $PostGitHub_{gt}$ are absorbed by our fixed effects.

Although our empirical approach enables comparison of observably similar startups, it may still suffer from omitted variable bias to the extent that our selection of similar startups based on observables does not allow us to fully net out the impact of a startup's intrinsic characteristics, as well as technology trends. To mitigate this concern as much as possible, we saturate our main specification with relevant fixed effects. Specifically, $\mu_i$ is a fixed effect for startup $i$, which absorbs differences in startups' fixed characteristics. The $\psi_{gt}$ is a group $g$ by (year-)quarter fixed effect, which controls for the possibility that differences across observation groups may change over time. Moreover, $\tau_{it}$ is a startup's industry group by quarter fixed effect that absorbs the effect of technology shocks we might have been unable to capture with our selection of similar startups. As we consider each of the industry groups used to build the synthetic software share index, $\tau_{it}$ allows for more granular control over technology shocks that may affect engagement with OSCs on GitHub and venture outcomes. We also add startup by age fixed effects, where age is defined at the yearly level to provide an even more stringent control of startup technology life-cycle effects. In addition, we estimate a two-way Mundlak regression model (Wooldridge 2021) for robustness—an approach that allows for substantial treatment effect heterogeneity. Using this model, we address concerns of inconsistency due to heterogeneous treatment effects by cohort and time (Callaway and Sant'Anna 2021).

Figure 2 displays the distribution of startups' initial engagement with OSCs on GitHub and their first funding rounds across quarters since inception. As shown, although engagement with OSCs tends to precede the moment a startup raises a first financing round, there is also a large degree of overlap between the two events. When we specifically concentrate on startups that both secure funding and engage with OSCs on GitHub, we find that 80% become engaged with OSCs on GitHub before receiving funding. On average, these startups begin their GitHub activity five months before their first round of funding. Table 4 provides additional descriptive statistics for the matched sample, while Table 5 explicitly distinguishes between startups that engaged with OSCs on GitHub and similar startups that did not. As shown in Table 5, the proportion of startups that have raised a financing round, and have raised a

**Figure 2.** (Color online) Startups' First Engagements with OSCs on GitHub and Rounds Raised




*Note*. This figure illustrates the distribution of startups' initial engagement with OSCs on GitHub and their first funding rounds across quarters since inception.

**Table 4.** Summary Statistics—Matched Sample

| Variables | Mean | Std. Dev. | Min | Max | p50 |
|---|---|---|---|---|---|
| Raised funds | 0.384 | 0.486 | 0 | 1 | 0 |
| Launched product | 0.087 | 0.282 | 0 | 1 | 0 |
| IPO | 0.006 | 0.077 | 0 | 1 | 0 |
| Acquired | 0.077 | 0.267 | 0 | 1 | 0 |
| Engaged with OSCs | 0.21 | 0.407 | 0 | 1 | 0 |
| Top Team | 0.119 | 0.324 | 0 | 1 | 0 |
| AI | 0.071 | 0.257 | 0 | 1 | 0 |
| Data Analytics | 0.16 | 0.367 | 0 | 1 | 0 |
| Information Technology | 0.273 | 0.446 | 0 | 1 | 0 |
| Internet Services | 0.263 | 0.44 | 0 | 1 | 0 |
| Software | 0.545 | 0.498 | 0 | 1 | 1 |
| N. Industry Groups | 3 | 2 | 1 | 14 | 3 |
| Software share | 0.667 | 0.34 | 0 | 1 | 0.714 |
| California | 0.396 | 0.489 | 0 | 1 | 0 |
| Massachusetts | 0.047 | 0.212 | 0 | 1 | 0 |
| New York | 0.155 | 0.362 | 0 | 1 | 0 |

*Notes*. In the case of product launches, we restrict the sample to companies founded after 2011, given that Product Hunt started only in 2014. Std. Dev., standard deviation; p50, 50th percentile or median.

**Table 5.** Descriptive Statistics of Matched Startups

| | (1) Eng. with OSCs | | (2) Not eng. with OSCs | | (3) |
|---|---|---|---|---|---|
| | Mean | S.D. | Mean | S.D. | Diff. ($p$-value) |
| *Raised round* | 0.4976 | 0.5000 | 0.3540 | 0.4782 | 0.1436 (0.00) |
| *Raised 1st round from VC* | 0.2356 | 0.4244 | 0.0994 | 0.2992 | 0.1362 (0.00) |
| *Raised 1st round from successful investor* | 0.1797 | 0.3840 | 0.0683 | 0.2522 | 0.1114 (0.00) |
| *Raised large financing amount* | 0.1392 | 0.3461 | 0.0633 | 0.2436 | 0.0758 (0.00) |
| *Launched a product* | 0.1370 | 0.0046 | 0.0492 | 0.0015 | 0.0878 (0.00) |
| Observations | 7,207 | | 27,091 | | 34,298 |

*Notes.* This table reports descriptive statistics distinguishing between startups that had engaged (eng.) with external repositories on GitHub (our measure for engagement with OSCs) and startups with no such engagement. Successful investors are those with a number of portfolio exits larger than the median. Portfolio exits are IPOs and acquisitions by portfolio startups in the five years prior to an investor's investment in startup $i$. A large financing amount is an amount greater than the median. The *Launched a product* variable is defined for startups founded after 2011. Diff., difference; S.D., standard deviation.

first round from a VC and/or a successful investor, is larger among those that engage with OSCs. Additionally, startups engaging with OSCs raise more money. Finally, the proportion of startups that have launched a product on Product Hunt is larger among startups that engage with OSCs.

# 5. Results

## 5.1. Baseline Results: Engagement with OSCs and Financing

We begin by discussing the results of our difference-in-differences model, where we compare the change in the probability that a startup will have raised a financing round after it starts to engage with OSCs on GitHub, relative to similar startups with no such engagement. The results are reported in Table 6, where we cluster standard errors by observation groups.[19]

Our approach consists of progressively saturating our model with fixed effects to address potential threats from factors we might not have appropriately controlled for. The model in column (1) includes year-quarter and startup fixed effects. The latter control for time-invariant unobservables across startups that may be correlated with the propensity to engage with OSCs. As shown, startups are more likely to have obtained funds after they become active on GitHub. Specifically, startups that engage with OSCs on GitHub are 15 percentage points more likely to have raised a financing round, all else equal. This represents a 65% increase

**Table 6.** Baseline—Raising a Financing Round

| | Raising a first round | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| *Post GitHub$_{gt}$* | 0.0518*** | | | |
| | (0.00244) | | | |
| *Post GitHub$_{gt}$ × EngagementWithOSC$_i$* | 0.147*** | 0.129*** | 0.120*** | 0.0811*** |
| | (0.00521) | (0.00534) | (0.00529) | (0.00460) |
| Startup FE | Y | Y | Y | Y |
| Yr-Quarter FE | Y | | | |
| Yr-Quarter × Observation Group FE | | Y | Y | Y |
| Yr-Quarter × Industry Group FE | | | Y | Y |
| Startup Age × Startup FE | | | | Y |
| Observations | 778,594 | 773,769 | 773,755 | 751,674 |
| $R^2$ | 0.673 | 0.747 | 0.753 | 0.954 |
| Mean D.V. | 0.226 | 0.226 | 0.226 | 0.224 |

*Notes.* This table reports the results from estimating the difference-in-differences model described by Equation (1) for the likelihood that a startup will have raised a first financing round by quarter $t$. We censor the sample on 2020q2. We observe each startup for up to 24 quarters, depending on whether the upper sample limit of 2020q2 is binding or not. *PostGitHub$_{gt}$* is a time-varying binary indicator that becomes one for all the startups in an observation group $g$ after startup $i$ engages in an external activity on GitHub. *EngagementWithOSC$_i$* takes a value of one if startup $i$ engages in at least one public activity across external repositories during the period we observe. It is our measure for engagement with OSCs on GitHub. In column (1), we report the results, having included startup and year-quarter fixed effects. In column (2), we add observation group by year-quarter fixed effects. In column (3), we additionally include industry group by year-quarter fixed effects. In column (4), we add startup by age (measured in years) fixed effects. Standard errors (in parentheses) are clustered by observation groups. D.V., dependent variable; FE, fixed effects; Yr, year.

***$p < 0.01$.

over the overall average, which is 0.23. The results remain similar in column (2), where we add observation group by year-quarter fixed effects, which address the possibility that differences between startups that engage with OSCs and similar startups may change over time. In column (3), we add granular industry group by year fixed effects. These fixed effects address the possibility that technology trends differentially impact startups' ability to attract funds. With this specification, the coefficient of interest slightly decreases to 12 percentage points. Finally, in column (4), the coefficient becomes 0.08—equivalent to a 36% increase in the mean—after adding age (measured in years) by startup fixed effects to account for differences in firms' technology life cycle, which may systematically vary with their engagement with OSCs and might be correlated with the outcome. Taken together, these results suggest that engagement with OSCs can help startups achieve funding milestones, possibly by accelerating technology development.

To further explore these initial findings, Table 7, column (1) replicates the analysis from Table 6, column (4), excluding startups that engaged with OSCs on GitHub before their first financing round, while column (2) does the same for startups that engaged after their first financing round, along with their respective set of similar startups. Normalized by the outcome mean, the effect of becoming active with OSCs on GitHub is, as expected, 27 percentage points larger in column (2) (when we exclude startups that only engage with GitHub after getting any type of first funding) than in column (1) (when we exclude startups that engage with OSCs before receiving funding). This suggests that the startups that benefit from engaging with OSCs on GitHub are those that do so before receiving funding.

In Tables A2–A9 in the Online Appendix, we provide a battery of robustness checks. First, as shown in Table A2, our results are robust to matching startups engaged with OSCs on GitHub with startups that do not have

**Table 7.** Raising a Financing Round: By Timing of Involvement with OSCs

| | Raising a first round | |
|---|---|---|
| | (1) | (2) |
| $Post\ GitHub_{gt} \times EngagementWithOSC_i$ | 0.0722*** | 0.0931*** |
| | (0.00594) | (0.00723) |
| Mean D.V. | 0.27 | 0.17 |

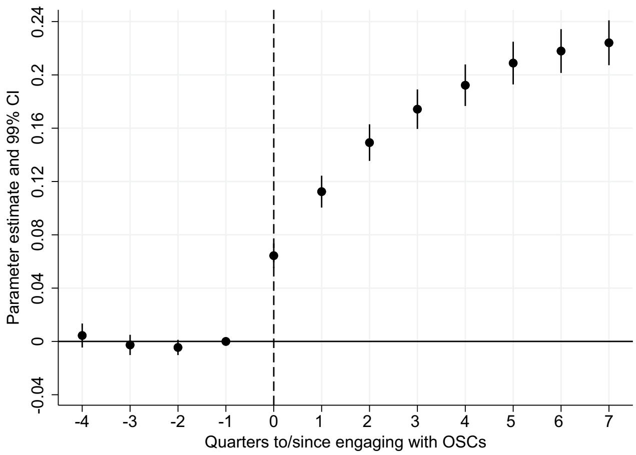
*Notes.* This table reports the results from estimating the difference-in-differences model described by Equation (1) for the likelihood that a startup will have raised a first financing round by quarter $t$. Column (1) (column (2)) replicates the same analysis as in Table 6, column (4), excluding from the sample startups active with OSCs on GitHub that became active before (after) raising their first financing round. Standard errors (in parentheses) are clustered by observation groups. D.V., dependent variable.

$^{***}p < 0.01$.

GitHub accounts. In essence, we exclude startups that had a GitHub account, but only interacted with their own repositories for software development from the pool of comparison startups. Additionally, Table A3 shows that the results are robust to a modified matching procedure where we required the focal and similar startups to share at least two technology keywords, rather than a similar value of the *software share* index. Moreover, as shown in Table A4, these results are robust to not matching startups on *Top Team*, which may not be a strictly predetermined variable. In Table A5, we reproduce the main results having excluded from the sample startups founded before 2014. As GitHub's popularity has grown in recent years, we concentrate on this contemporary time frame wherein startups may have systematically explored leveraging GitHub OSCs to secure funding. The displayed results continue to hold. In Table A6, we reproduce the main results in a sample that only includes startups until their second or third year of age. This suggests that survivor bias is not a big concern for our results. In Table A7, we attempt to address reverse causality concerns to the extent that engagement with OSCs might drive funding, but certain investors might push startups to accelerate the development of their products to obtain funding. To this end, we produce a robustness check that excludes all startups located in California and those issued from an accelerator. In fact, VCs, which are renowned for being hands-on and providing advice on the future technology directions of their investees (Gompers and Lerner 2001), exhibit a notable concentration in both California and accelerators. In Tables A8 and A9, we address the concern that our results might be driven by a startup's more innovative new hires, who may be more likely to engage with OSCs and secure funding for their startups. For this, we collected time-varying data from Crunchbase on new hires and distinguished them based on whether they graduated from top universities. As shown in Table A8, there is a positive association between new hires and active engagement with OSCs on GitHub, as well as with the management of their own internal repositories. Additionally, we show that the association is strongest for new hires who graduated from top universities.[20] The association is also strongest for new hires with a STEM (science, technology, engineering, or math) degree. Encouragingly, Table A9 shows that once we control for a startup's new hires (and whether they graduated from top universities or whether they have a STEM degree) in our main regressions, the coefficient of interest deviates minimally relative to the one reported in Table 6, column (4). Table A10 in the Online Appendix displays that the results—both in terms of magnitude and significance of the main effects—are robust to estimating a two-way Mundlak regression model, which allows for treatment effect heterogeneity by cohort and time.

In Figure 3, we explore possible differences in pretrends between startups that engage with OSCs and

**Figure 3.** Baseline Results—Event Study for Raising a First Round




*Notes*. This figure shows how the probability that a startup will have raised a round changes after a startup starts engaging with OSCs on GitHub. To generate this graph, we modified Equation (1) in the main text by substituting the $PostGitHub_{g,t}$ indicator with binary variables for each year. We control for the same fixed effects as those in Table 6, column (1). The vertical lines represent 99% confidence intervals.

similar startups that do not. For this purpose, we modify Equation (1), substituting the $PostGitHub_{gt}$ indicator with dummies for each quarter before and after a startup becomes engaged with OSCs on GitHub. We restrict the sample to the four quarters preceding a startup's involvement on GitHub and the eight quarters after to focus on the period just before and after such an involvement. We control for the same set of fixed effects as in Table 6, column (1). We infer two important implications from this figure. First, the difference in the probability of having obtained funds is flat around zero in the preperiod, suggesting that startups becoming involved with OSCs on GitHub and those that do not are not meaningfully different, conditional on our set of fixed effects.[21] Second, the relationship between a startup's engagement with OSCs and the likelihood of being funded is strong and persists over time. In Figure A4 in the Online Appendix, we provide evidence that the main trends remain unchanged when we exclude startups (and their respective similar counterparts) that do not have four quarters preceding their engagement with OSCs.

## 5.2. Is This Really Engagement with OSCs?

Ideally, we would want to separate the effect of engaging with OSCs on GitHub from the effect of a startup utilizing GitHub for more efficient internal software development. Indeed, a startup might want to have a GitHub organization account to engage with OSCs, as well as for internal software development purposes, which may become more critical as the technology of a startup matures. To try to separate these two drivers, we modify Equation (1), considering two main variables of interest: (1) whether a startup engages with OSCs on GitHub by interacting with public repositories that it does not directly control, and (2) whether a startup uses GitHub as a public software development tool, engaging with (public) internal repositories that it directly controls. The reference group is represented by startups randomly selected with the same criteria mentioned above and who do not have a GitHub organization account.[22]

The results from this exercise are reported in Table 8. The relationship between startups engaging with OSCs through external repositories and receiving funding is 9 percentage points larger than the relationship between startups focusing on their internal repositories and receiving funding relative to the reference group. The $p$-value for the difference between these two coefficients is 0.00.

## 5.3. Exploiting Variation in Expected Payoffs from Engaging with OSCs: Technology Novelty and Market Competition

Although our analyses are encouraging, we might not have fully addressed concerns of omitted variable bias. To mitigate this concern, we first assess which startups should have higher expected payoffs from becoming engaged with OSCs. Using these startups, we then compare the relationship between their engagement with OSCs and receiving funding to the relationship of those startups with lower expected payoffs from engagement on GitHub. The rationale is to approximate the group of startups that do not engage with OSCs—who arguably derive lower benefits so as to preclude them from interacting with GitHub—with engaged startups on the low end of payoffs associated with interacting with OSCs.

The terms of the expected payoff for startups may vary, depending on the type of technology and market conditions. For example, novel technologies may require codevelopment with other organizations to solve complex problems and access distant knowledge, making access to OSCs crucial. Based on the notion that innovation is a recombinant search process (Fleming 2001), the necessary distant knowledge pieces are unlikely to all be held by one individual organization (Majchrzak et al. 2004, Gans et al. 2021). For instance, ML, a fairly general technology, has advanced largely through open source tools like TensorFlow and PyTorch, contrasting with past reliance on proprietary knowledge (Agrawal et al. 2019, Goldfarb et al. 2023). Speed is also critical for securing a first-mover advantage in developing novel technologies (Lieberman 1987, Franco et al. 2009, Leiblein et al. 2023), which may outweigh the potential risks of disclosing information to others.

In competitive markets, appropriability challenges may be more pronounced, especially for digital startups. Traditional IP protection, such as patents, is costly,

**Table 8.** Adding Engagement with Internal Repositories

| | (1) Raising a first round |
|---|---|
| *Post GitHub*$_{gt}$ × *EngagementWithOSC*$_i$ | 0.157*** |
| | (0.00535) |
| *Post GitHub*$_{gt}$ × *Internal Activity*$_i$ | 0.067*** |
| | (0.0051) |
| Startup FE | Y |
| Yr-Quarter × Observation Group FE | Y |
| Yr-Quarter × Industry Group FE | Y |
| Observations | 946,389 |
| $R^2$ | 0.759 |
| Mean D.V. | 0.261 |

*Notes.* This table reports the results from estimating a variant of the difference-in-differences model described by Equation (1) for the likelihood that a startup will have raised a first financing round by quarter *t*. Here, we consider two events: (1) whether a startup engaged with an external repository (*EngagementWithOSC*$_i$), and (2) whether the startup engaged with an internal repository, which it directly controls (*InternalActivity*$_i$). *PostGitHub*$_{gt}$ is now a time-varying binary indicator that becomes one for all the startups in an observation group *g* after startup *i* engages in an (internal or external) activity on GitHub. The comparison is against startups with no GitHub activity. Standard errors (in parentheses) are clustered by observation groups. D.V., dependent variable; FE, fixed effects.

***$p < 0.01$.

slow, and offers limited safeguards, particularly for small firms (Veugelers and Schneider 2018). In the digital domain, where access to a technology's backend is relatively easy (Roche et al. 2024) and secrecy is difficult to maintain (Cunningham and Kapacinskaite 2024), the risk of reverse engineering is high, leaving startups with limited options for protection (Gans and Stern 2003). As a result, "isolating mechanisms" (Rumelt 1984, p. 568) that can typically sustain a competitive advantage are constrained, making complementary assets more crucial than improvements in the core technology (Teece 1986). Startups may, thus, rely less on and benefit less from engagement with OSCs because the risks are higher than the achievable benefits.

Building on these arguments, we consider the following two determinants of a startup's expected payoffs from becoming involved with OSCs: the novelty of a startup's technology and the level of market competition.

We measure novelty by whether the combination of a startup's industry group keywords at the startup's founding year is relatively new. That is, less than three years should have passed since a keyword combination first appeared on Crunchbase.[23] To operationalize the level of competition in each technology space, we count the number of startups active in a given year-quarter and with the same industry group keywords as the focal one. The market in which a focal startup operates is thus considered competitive if the number of startups possessing the same technology keyword combination is greater than the 90th percentile.

We first estimate a linear probability model where the dependent variable is *EngagementWithOSC*$_i$. As reported in Figure A5 of the Online Appendix, startups developing novel technologies are 9 percentage points more likely to engage with OSCs. Conversely, startups that have experienced high competition for at least one year-quarter are 2 percentage points less likely to engage with OSCs.

Moving to Table 9, the results in column (1) show that startups developing relatively new technologies become 8 percentage points more likely to have attracted funds by a given quarter after engaging with OSCs, relative to other startups that engage with OSCs. These results are obtained after accounting for the differential propensity of novel technologies to receive funding in any given period, which we hold constant by including year-quarter by technology novelty fixed effects. We include these fixed effects in addition to startup and year-quarter by observation group fixed effects.

Moreover, as reported in Table 9, column (2), startups operating in competitive markets are 14 percentage points less likely to have received funding after they become engaged with OSCs, relative to other startups that engage with OSCs. Computing the linear combination of *PostGitHub*$_{gt}$ × *EngagementWithOSC*$_i$ and *Competition*$_{it}$ × *PostGitHighHub*$_{gt}$ × *EngagementWithOSC*$_i$ reveals that the overall impact of engaging with OSCs on the likelihood of getting funding is zero for startups operating in competitive markets relative to the reference outcome. Here, again, we account for the differential propensity of startups in competitive markets to receive funding by including year-quarter by competitive market fixed effects. Overall, these results suggest that engaging with OSCs is especially beneficial when the expected payoffs are relatively high.

### 5.4. Heterogeneity: Technology Use-Cases and Startup Domains

To enrich our analysis, we categorize the technology use-cases of external repositories with which startups interact. In particular, we distinguish between engagement with OSCs with respect to Software Development/Backends, Machine Learnings, Application Programming Interfaces, and User Interfaces. Further, we group a subsample of observed startups into different domains: Software Tools, AI & Blockchain, Platform, Consumer-facing, and the remainder. The goal of this exercise is twofold: (1) to report the association between startup domains and the technology use-cases of external repositories they engage with on GitHub; and (2) to discern which startups gain the most by interacting with (what type of) external repositories on GitHub.

To achieve the first goal, we produce cross-sectional correlations in Figure 4 between the use-cases of the external repositories that GitHub-active startups engage

**Table 9.** Heterogeneity in Raising a Financing Round—By Technology Novelty and the Level of Market Competition

| | Raising a first round | |
|---|---|---|
| | (1) | (2) |
| $Post\ GitHub_{gt} \times EngagementWithOSC_i$ | 0.0717*** | 0.138*** |
| | (0.00774) | (0.00563) |
| $Novel_i \times Post\ GitHub_{gt}$ | 0.0140*** | |
| | (0.00527) | |
| $Novel_i \times Post\ GitHub_{gt} \times EngagementWithOSC_i$ | 0.0777*** | |
| | (0.0109) | |
| $High\ Competition_{it} \times EngagementWithOSC_i$ | | 0.0558* |
| | | (0.0321) |
| $High\ Competition_{it} \times Post\ GitHub_{gt}$ | | −0.0358*** |
| | | (0.00658) |
| $High\ Competition_{it} \times Post\ GitHub_{gt} \times EngagementWithOSC_i$ | | −0.143*** |
| | | (0.0157) |
| Startup FE | Y | Y |
| Yr-Quarter × Observation Group FE | Y | Y |
| Yr-Quarter × Novel | Y | |
| Yr-Quarter × High Competition | | Y |
| Observations | 773,769 | 773,767 |
| $R^2$ | 0.751 | 0.749 |
| Mean D.V. | 0.226 | 0.226 |

*Notes*. This table reports the results from estimating a variant of the difference-in-differences model described by Equation (1) for the likelihood that a startup will have raised a first financing round by quarter *t*. In column (1), we assess whether relationships uncovered in Table 6 vary depending on the novelty of a startup's technology. We measure technology novelty by whether the combination of a startup's industry group keywords at the startup's founding year is relatively new. That is, less than three years should have passed since a keyword combination first appeared on Crunchbase. In column (2), we assess whether the relationships uncovered in Table 6 vary by the level of market competition. We measure market competition by the number of startups active in a given year-quarter and with the same industry group keyword combination as the focal one. The market in which a focal startup operates is thus considered competitive if the number of startups possessing the same industry group keyword combination is greater than the 90th percentile. D.V., dependent variable; FE, fixed effects; Yr, year.

*$p < 0.10$; ***$p < 0.01$.

with and the domains in which these startups specialize, having controlled for observation group and startup founding year fixed effects and clustered standard errors by observation group. We standardize the coefficients with the means of the corresponding outcomes. As shown in Figure 4, platform startups are most actively engaged with external repositories across all use-cases. This is expected, given that their business model rests on cooperating with developers outside a startup (Benzell et al. 2019), and therefore may have a better product-market fit. Platform startups are followed by Software Tools and AI & Blockchain startups. The latter category especially engages with external ML repositories. Finally, Consumer-facing startups appear the least active on the GitHub platform.

Next, we assess which startups gain the most from engaging with OSCs on GitHub. We do so by interacting the different terms on the right-hand side of Equation (1) with the startup domain indicators generated. We add startup fixed effects, as well as year-quarter by startup domain fixed effects. The domains we consider are Software Tools, AI & Blockchain, Platform, and Consumer-facing.

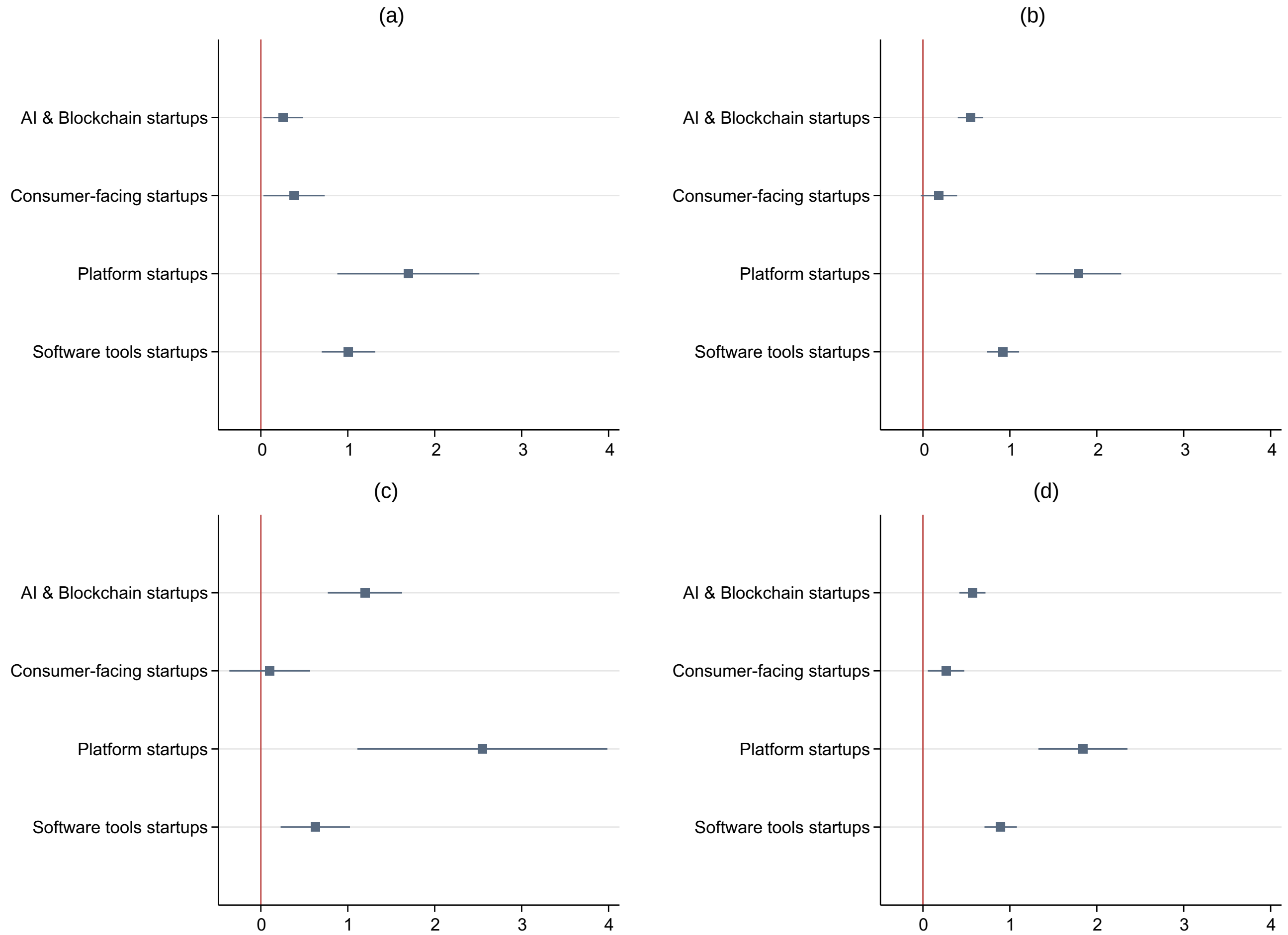
The results reported in Figure 5 show that AI & Blockchain startups derive the largest benefits, in terms of improved likelihood of attracting funds, from engaging with OSCs. As AI & Blockchain are relatively more novel and complex technologies, these results are consistent with our earlier findings that novel technologies derive the largest financing gains from the interaction with OSCs on GitHub.

We investigate this last result in Figure 6 in more detail, distinguishing the different technology use-cases with which our startups engage on GitHub. In panel (a), we find that startups benefit from interacting with the full spectrum of repository use-cases. Zooming in on AI & Blockchain startups in panel (b), we show that they do not benefit as much by engaging with external repositories related to UI.

### 5.5. Toward a Feasible Mechanism: Innovation or Merely Visibility?

Having provided suggestive evidence that engaging with OSCs on GitHub is related to attracting funds, we next evaluate whether startups engage with OSCs on GitHub to innovate or merely to enhance their visibility.

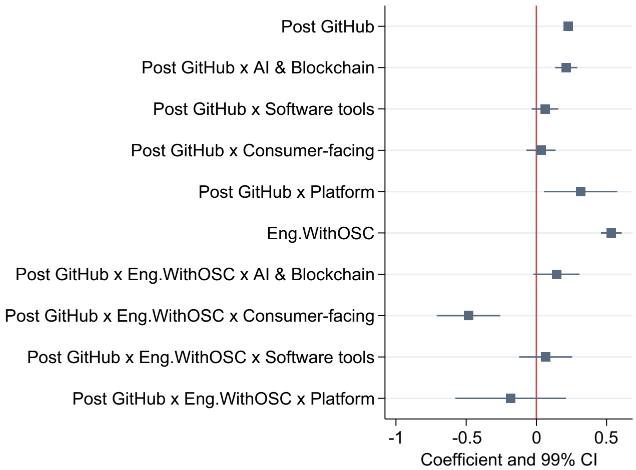
**Figure 4.** (Color online) Correlations Between Startup Domains and Technology Use-Cases on GitHub




*Notes*. In this figure, we report the correlation coefficients and their 99% confidence intervals we obtain from estimating four cross-section models for the following outcomes: (1) a (0/1) indicator for whether a startup engaged with external repositories related to UI (panel (a)); (2) a (0/1) indicator for whether a startup engaged with external repositories related to SD (panel (b)); (3) a (0/1) indicator for whether a startup engaged with external repositories related to ML (panel (c)); and (4) a (0/1) indicator for whether a startup engaged with external repositories related to API (panel (d)). The dependent variables of interest are (1) an indicator identifying Platform startups; (2) an indicator identifying AI & Blockchain startups; (3) an indicator identifying Software Tool startups; and (4) an indicator identifying Consumer-facing startups. These are mutually exclusive categories; startups developing neither of these technologies represent the reference outcome. The reported coefficients are standardized by the means of the outcomes. In the regressions, we include observation group and startup founding year fixed effects. Standard errors are clustered by observation group. (a) Engagement with external UI repositories. (b) Engagement with external SD repositories. (c) Engagement with external ML repositories. (d) Engagement with external API repositories.

To investigate this, we first assess whether a startup's engagement with OSCs on GitHub is related to innovation that potential investors may value. For this scope, we use Product Hunt data on product launches. As Product Hunt only started in 2014, we restrict our sample to those startups founded after 2011 for this analysis, given that older startups might have already launched their products, but without relying on the Product Hunt platform.[24] Building on these data, we modify Equation (1), considering whether a startup will have launched a product by quarter $t$ as the dependent variable. The results are reported in Table 10, reproducing the same structure as in Table 6. For example, in columns (1)–(3), we provide evidence that after startups begin to engage with OSCs, their likelihood to have launched a product increases by 8 percentage points, which represents an increase by a factor of 1.8 in the outcome mean.[25] In Table A12 in the Online Appendix, we show that the results remain similar when we only consider higher-quality products—that is, those whose number of ratings or upvotes received as of September 2022 is greater than the 75th percentile. Moreover, the results in Table A13 in the Online Appendix indicate that the association is stronger when we examine the likelihood that a startup will have launched the first version of a product. This last result aligns with our findings on technology novelty and suggests that the gains from interacting with OSCs are more substantial when startups can address the market with novel products. Overall, our results using data from Crunchbase, as well as Product

**Figure 5.** (Color online) Relationship Between Engaging with OSCs and Funding: By Startup Domain




*Notes.* In this figure, we report the coefficients and respective 99% confidence intervals obtained from estimating a variant of Equation (1) for the likelihood that a startup will have raised a first financing round by quarter *t*. Specifically, we interact $PostGitHub_{gt}$ and $EngagementWithOSC_i$ and $PostGitHub_{gt}$ with (0/1) indicators identifying (1) platform startups; (2) AI & blockchain startups; (3) software tools startups; and (4) consumer-facing startups. In the regression, we include startup and year-quarter by domain fixed effects. The startup domains are Software Tools, AI & Blockchain, Platform, and Consumer-facing. The reported coefficients are standardized by the outcome mean. Standard errors are clustered by observation group.

Hunt, may be viewed as suggestive evidence that startups make use of others' code to innovate and gain access to financing.

To complement this finding, we successively make use of a specific activity on GitHub that should demonstrate being active on GitHub, but not alter the technology that a startup produces: the creation or modification of a repository's readme file, which provides documentation to a repository, often including a nontechnical summary of the codebase. In Table 11, we estimate a similar model as reported in Table 8, decomposing internal activities into creation or modification of readme files and the remainder. As shown, creating or modifying readme files is not significantly related to receiving funding. This provides some indication that investors do not simply react to "cosmetic" activities that startups engage with on GitHub.[26]

To bolster these results, Table A14 in the Online Appendix captures how the relationship between the likelihood of making a new repository public (column (1)), of making a previously private repo public (column (2)), and of making a repo public that will eventually be forked by others (column (3)) changes over time. In particular, we model a linear time trend for the 12 months before and after raising the first round and allow for a different intercept at the time of raising the first round. In column (1), we detect an increasing trend in making at least one repository public prior to raising a first financing round. Importantly, we do not identify any significant change in the trend after receiving financing. In column (2), we zoom in on the likelihood that a startup turns a previously private repository public and find similar patterns as in column (1). Because the dynamics of startups making their existing repositories public do not change before and after the first round, we infer that increasing the visibility of their technology to attract investors is unlikely the startups' sole motive for becoming involved with the GitHub OSCs. Finally, the results reported in column (3) indicate that the likelihood of publishing repositories that are eventually forked increases as the startup approaches its first financing round, and its slope remains unchanged thereafter. We interpret these results such that the trajectory of publishing relevant repositories is similar before and after raising a first round.

### 5.6. Investor Heterogeneity

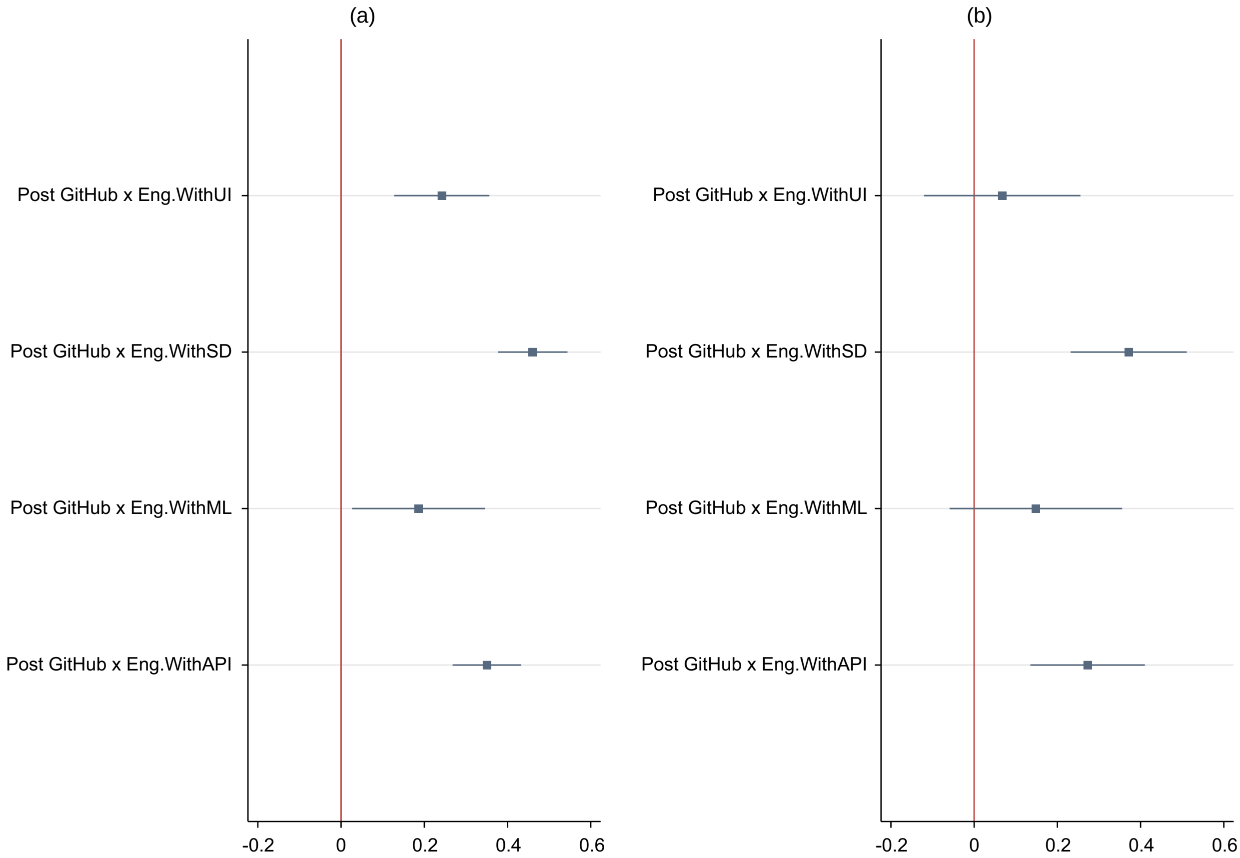
In the last part of our empirical analysis, we aim to assess which investors are most responsive to startups that engage with OSCs on GitHub. The analyses are reported in Table 12. We begin by considering whether a startup's engagement with OSCs is related to raising large round amounts (column (1)) or smaller ones (column (2)).[27] Large financing amounts are those in the upper quartile for first rounds. As displayed, after startups engage with OSCs, their likelihood of being funded through a large first round increases by 7 percentage points, equivalent to an increase of a factor of 1.4 from the mean. Conversely, their likelihood of being funded through a smaller round increases by 5 percentage points, relative to a mean of 0.18.

Further, we report in columns (3) and (4) that the relationship between startups engaging with OSCs and the likelihood that they attract VC funds is larger by a factor of 1.5 relative to the mean, whereas their likelihood of attracting non-VC investors increases by 1.6 percentage points relative to a mean of 0.15.[28] Finally, we find similar differences in the last two columns of Table 12, where we distinguish between having raised a first round from successful investors (column (5)) and having raised a first round from less successful investors (column (6)). We measure investor success by the number of portfolio startups that were acquired or went public in the five years prior to investing in the focal startup. Successful investors are those with a number of portfolio exits in the upper quartile. Overall, these results suggest that by engaging with OSCs, startups can raise relatively larger financing rounds from investors that provide large nonfinancial capital in addition to money.

## 6. Discussion and Conclusions

In 1990, Apple hit peak profitability, thanks to the success of the Mac personal computer. In the next years,

**Figure 6.** (Color online) Relationship Between Engaging with OSCs and Funding: By Technology Use-Cases




*Notes*. In this figure, we report the coefficients and respective 99% confidence intervals obtained from estimating Equation (1) for the likelihood that a startup will have raised a first financing round by quarter *t*. We distinguish startups' engagements with external repositories related to: UI, SD, ML, API, and other. In panel (a), we consider all startups, whereas in panel (b), we zoom in on AI & Blockchain startups. In the regression, we include startup and year-quarter by startup domain fixed effects. We only report the coefficients of interest, which are standardized by the outcome mean. Standard errors are clustered by observation group. (a) All startups. (b) AI & Blockchain.

following competition by IBM and entry by other low-cost competitors, the oft-marveled-at competitive advantage Apple had achieved fell apart and brought the company to the verge of bankruptcy in 1997. One of the core reasons for this course of events has been attributed to the rise and use of an "open regime" by Apple's competitors, which was largely powered by open source. As a consequence, contributions by developers increased and helped amplify offerings and, thereby, the willingness of customers to pay for non-Apple products (Oberholzer-Gee 2021). In contrast to its competitors, Apple had notoriously stuck to a "closed" regime. Although Apple was no longer a startup at that point in time (Apple was founded in 1976), there were many new entrants who served as a useful illustrative example of the power of engaging with OSCs for early innovation strategy; the idea that lies at the core of this study.

In this paper, we investigate the role of engaging with OSCs among nascent digital firms in raising funds and to what extent this relationship is driven by the enhanced ability to innovate. We link Crunchbase profiles to accounts on the open source software development platform GitHub and to product launches on Product Hunt. Our results suggest that engagement with OSCs can be important in achieving funding milestones. Applying a difference-in-differences approach to estimate the likelihood of being funded as a function of engaging with OSCs on GitHub, we find that such engagement is associated with a substantial increase in the likelihood that a startup will have raised a first financing round. The estimated coefficients indicate at least a 36% increase in the mean. These findings appear most pronounced when startups raise large rounds and when funds are obtained from VCs and successful investors. These results, taken together, indicate that a startup's involvement with OSCs may be crucial for attracting earliest-round valuable investors that bring critical capital.

Our results further reveal that the relationship is strongest when startups are developing novel technologies, for which the benefits of relying on knowledge

**Table 10.** Innovation: Product Launches

| | Launching a product | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| *Post GitHub*$_{gt}$ | −0.0102*** | | | |
| | (0.00142) | | | |
| *Post GitHub*$_{gt}$ × *EngagementWithOSC*$_i$ | 0.0800*** | 0.0766*** | 0.0753*** | 0.0218*** |
| | (0.00423) | (0.00424) | (0.00425) | (0.00260) |
| Startup FE | Y | Y | Y | Y |
| Yr-Quarter FE | Y | | | |
| Yr-Quarter × Observation Group FE | | Y | Y | Y |
| Yr-Quarter × Industry Group FE | | | Y | Y |
| Startup Age × Startup FE | | | | Y |
| Observations | 574,519 | 571,572 | 571,572 | 556,579 |
| $R^2$ | 0.631 | 0.700 | 0.703 | 0.958 |
| Mean D.V. | 0.044 | 0.044 | 0.044 | 0.043 |

*Notes*. This table reports the results from estimating the difference-in-differences model described by Equation (1) for the likelihood that a startup will have launched a product on Product Hunt by quarter *t*. As Product Hunt only started in 2014, we restrict our sample to those startups founded after 2011 for this analysis, given that older startups might have already launched their products, but without relying on the Product Hunt platform. *PostGitHub*$_{gt}$ is a time-varying binary indicator that becomes one for all the startups in an observation group *g* after startup *i* engages with OSCs on GitHub. *EngagementWithOSC*$_i$ takes a value of one if startup *i* engages in at least one public activity across external repositories during the period we observe. It is our measure for engagement with OSCs. In column (1), we report the results having included startup and year-quarter fixed effects. In column (2), we add observation group by year-quarter fixed effects. In column (3), we additionally include industry group by year-quarter fixed effects. In column (4), we add startup by age (measured in years) fixed effects. Standard errors (in parentheses) are clustered by observation groups. D.V., dependent variable; FE, fixed effects; Yr, year.
***$p < 0.01$.

inputs from OSCs (e.g., speed, access to existing integration packages, coproduction to make the technology work) may offset costs (Jordan and Mitchell 2015). Conversely, the relationship is weakest when the level of competition is high, and, thus, the expected payoff from engaging with OSCs is relatively lower. These findings also translate when we more closely examine the precise domain in which a startup is active. Using information from Crunchbase to group firms into AI & Blockchain, Consumer-Facing, Platform, and Software Tools, our results suggest that startups active in AI & Blockchain—the, on average, most novel domain—are those who achieve the highest gain. These firms typically rely on input from others to create a functioning product (e.g., state-of-the-art ML algorithms). Conversely, firms operating in the Consumer-facing domain experience the lowest benefits from engaging with OSCs.

**Table 11.** Technology Development or *Simply* Enhancing the Visibility of a Technology?

| | (1) Raising a first round |
|---|---|
| *Post GitHub*$_{gt}$ × *EngagementWithOSC*$_i$ | 0.157*** |
| | (0.00535) |
| *Post GitHub*$_{gt}$ × *Created/Modified Readme*$_i$ | 0.008 |
| | (0.0271) |
| *Post GitHub*$_{gt}$ × *Other Internal Activity*$_i$ | 0.068*** |
| | (0.0051) |
| Startup FE | Y |
| Yr-Quarter × Observation Group FE | Y |
| Yr-Quarter × Industry Group FE | Y |
| Observations | 946,389 |
| $R^2$ | 0.800 |
| Mean D.V. | 0.261 |

*Notes*. This table reports the results from estimating a similar model as the one reported in Table 8, decomposing internal activities on GitHub into creation or modification of readme files and the reminder. As in Table 7, *PostGitHub*$_{gt}$ is a time-varying binary indicator that becomes one for all the startups in an observation group *g* after startup *i* engages in an (internal or external) activity on GitHub. The comparison is against startups with no GitHub activity. Standard errors (in parentheses) are clustered by observation groups. D.V., dependent variable; FE, fixed effects; Yr, year.
***$p < 0.01$.

To delve deeper into the potential roles of engagement with OSCs for innovative outcomes, we examine the relationship between engagement on GitHub and the launching of new technology products, especially high-quality ones. Our results provide suggestive evidence that open source inputs enable startups to commercialize their inventions. From this, and paired with our earlier findings, we garner that although engagement with OSCs may not be the solution when appropriability concerns are heightened, due to competitive pressures, it can provide critical inputs for the development of the most novel and complex technologies. This ultimately helps firms to commercialize their inventions through product launches and achieve funding milestones.

We further provide evidence suggesting that startups do not merely engage with OSCs to increase their visibility vis-à-vis investors. It is possible that startups rely on OSCs and a public platform such as GitHub to signal the quality of their technology, similar to individuals who contribute to online question-and-answer communities to capture recruiters' attention (Xu et al. 2020).

**Table 12.** Investor Heterogeneity

| | (1) Large amount | (2) Smaller amount | (3) VC | (4) Non-VC investor | (5) Successful investor | (6) Less successful investor |
|---|---|---|---|---|---|---|
| *Post GitHub*$_{gt}$ × *EngagementWithOSC*$_i$ | 0.0721*** | 0.0483*** | 0.105*** | 0.0155*** | 0.0892*** | 0.0312*** |
| | (0.00381) | (0.00495) | (0.00435) | (0.00463) | (0.00403) | (0.00485) |
| Startup FE | Y | Y | Y | Y | Y | Y |
| Yr-Quarter × Observation Group FE | Y | Y | Y | Y | Y | Y |
| Yr-Quarter × Industry Group FE | Y | Y | Y | Y | Y | Y |
| Observations | 773,755 | 773,755 | 773,755 | 773,755 | 773,755 | 773,755 |
| $R^2$ | 0.701 | 0.754 | 0.732 | 0.745 | 0.739 | 0.743 |
| Mean D.V. | 0.0502 | 0.1756 | 0.0731 | 0.1527 | 0.0572 | 0.1685 |

*Notes.* This table reports the results from estimating the difference-in-differences model described by Equation (1). In column (1), we examine the likelihood that a startup will have raised a large financing round by quarter $t$. In column (2), we examine the likelihood that a startup will have raised a smaller financing round by quarter $t$. A large funding amount equals one if the amount raised falls in the last quartile for the amount raised, a *Smaller amount* equals one for all others. In column (3), we examine the likelihood that a startup will have raised a first VC-led round. In column (4), we examine the likelihood that a startup will have raised a first non-VC-led round. In column (5), we examine the likelihood that a startup will have raised a first round from successful investors. In column (6), we examine the likelihood that a startup will have raised a first round from less successful investors. *Successful investors* are those with a number of exits in the five years prior to investing in startup $i$ that falls in the last quartile of the distribution. Standard errors (in parentheses) are clustered by observation groups. D.V., dependent variable; FE, fixed effects; Yr, year.
***$p < 0.01$.

However, taking the totality of our findings, the interpretation that startups may be using code repositories available on GitHub to engage with external knowledge critical for scaling, integrating into the ecosystem, and producing a minimal viable product appears at least as relevant. Consistent with that interpretation, we show that investors do not react to "cosmetic" changes that startups add to the readme files associated with their repositories, which would primarily enhance the startups' visibility. Indeed, these results appear to indicate that VCs and other investors do not rely on mere code updates in their due diligence process.

Overall, our findings indicate that startups are feasibly "beefing up" their products by relying on inputs from OSCs and, by doing so, have a higher likelihood of receiving early stage financing. Qualitative data from interviews with developers and blog posts confirm the results of our quantitative analyses, suggesting that access to OSCs is, indeed, viewed as a critical input to launching a successful venture via its impact on innovation. As stated by a developer:

> Starting a new venture is like embarking on an epic quest, and open source is the magic elixir for startups. Here is why: …
>
> Community Magic: Open source projects thrive on collaboration. Tap into the wisdom of the crowd and build upon battle-tested code. It's like having a fellowship of developers at your side.
>
> Innovation Accelerator: With open source, you stand on the shoulders of giants. Leverage existing projects and focus on what makes your startup unique.[29]

Our study contributes to increasing our understanding of the role of a particular channel through which startups can access outside knowledge—engagement with OSCs—and how engagement with OSCs matters in attracting funding. As such, our findings extend the literature that analyzes the role of open source for firm productivity (Nagl 2018, 2019; Shah and Nagle 2019). We highlight a novel channel through which startups benefit from using and actively engaging with OSCs. Namely, in our context, technology startups rely on open source to attract investors, particularly VCs and successful investors, during startups' early stages. Further, we contribute to the entrepreneurial finance literature investigating whether VCs invest in the founding team or the technology (Kaplan et al. 2009, Bernstein et al. 2017, Gompers et al. 2020). We provide evidence that an advantage of open source community engagement lies in enabling innovation, particularly of the most novel technologies. This, in turn, has implications for raising the first financing round. Finally, using machine learning algorithms to classify startups' activities on GitHub builds on an emerging line of research that applies sophisticated data techniques to categorize firm strategies (Conti et al. 2020, Guzman and Li 2023).

Naturally, this study is not without limitations. First, the external validity of our approach may be constrained, given that we focus on a specific open source platform. However, GitHub is the largest host of source code, with over 40 million public repositories to date, and anecdotal evidence suggests that investors consider public GitHub activities in their due diligence efforts (Jain 2018). Although our results are based on particular activities on a specific online platform, we believe they have broader implications, especially for early stage ventures. Second, we acknowledge that startups can use OSS without forking repositories, via package managers like pip or npm, or through private OSS modifications, which could lead to underestimating OSS usage. However, our focus is on the impact of engaging with OSCs, and this should remain unaffected by such concerns.[30] Third, there is a clear selection into using OSCs at play. We acknowledge and address this with the econometric tools at our disposal. As our results suggest, those that engage with OSCs differ along dimensions such as team composition and technologies developed. Although such differences make estimating a truly causal relationship difficult, they do inform our understanding of the type of ventures that actively engage in OSCs. To us, this is a feature of our study, as it furthers our knowledge about which population of organizations these activities play a role in.

In conclusion, this paper provides important insight into the role of engaging with external knowledge through open source platforms for startup performance outcomes. In particular, our findings contribute to our understanding of the impact of early stage tech-stack investments on achieving funding milestones. By opening the technology "black-box," we reveal important nuances that have been largely overlooked in the literature—namely, that using open source to build an at least minimum viable product can help firms attract funding. Given the importance of entrepreneurship and innovation for economic growth (Agarwal et al. 2007, 2010; Adelino et al. 2017), increasing shifts toward VC investment in software (Lerner and Nanda 2020), and the difficulties associated with devising appropriate support systems (Lyons and Zhang 2018, Yu 2020), these findings not only carry important implications for founders and firms, but also for policy-makers alike.

## Acknowledgments

The authors thank the editor and anonymous reviewers for their valuable insight and advice. The authors thank Ilia Azizi, Kimon Protopapas, and Richie Zitomer for excellent research assistance. Many thanks go to Bruno Cassiman, Celine Fei, Matt Higgins, David Hsu, Jorge Guzman, Harsh Ketkar, Rem Koning, Tobias Kretschmer, Frank Nagle, Melissa Perri, and Toby Stuart for advice on drafts. This manuscript benefited from many helpful comments provided at CEAR, the Digital Economy Workshop, the Digital Initiative Workshop, MAD Conference, the Munich Summer Institute, SCECR, the Strategy Science Conference, and the West Coast Research Symposium. The authors are grateful for the suggestions provided by Karim Lakhani and members of the Laboratory for Innovation Science at the onset of this work, as well as participants of seminars at the CAS Platform Seminar Munich, Copenhagen Business School, Cornell University, HEC Paris, the Intellectual Property & Innovation Virtual seminar series, LUISS Guido Carli University, Max Planck Institute for Innovation and Competition, the National Bureau of Economic Research Productivity Seminar, the Strategy Unit Seminar at Harvard Business School, Universitat Pompeu Fabra, Warwick University, and the Workshop for Entrepreneurial Finance and Innovation. The authors are listed in alphabetical order.

## Endnotes

[1] https://github.com/enterprise/startups (accessed March 1, 2024).

[2] For example, data from a representative survey of about 5,000 German firms (Mannheim Innovation Panel) show that close to 20% of firms use open source to access external knowledge, whereas only 10% use patents of others to do so. See Figure A1 in the Online Appendix for details.

[3] For the purpose of this study, we define engagement with OSCs as actively participating in the collaborative open source ecosystem where software, tools, or projects are developed openly and transparently. Specifically, this entails interacting with external repositories that the focal startup does not directly control.

[4] We make the distinction between activities related to internal and external repositories that can be publicly observed on https://github.com. Specifically, we define those repositories as internal that originate from and are held by the focal firm. In other words, the focal organization created these and can determine what happens to the original source code. In contrast, external repositories are those created and controlled by other organizations. This means that the focal organization can only alter the code by requesting changes to the organization that created the code (through pull requests or by raising issues) or by forking and then altering the code (copying). Private repositories, accessible only to the focal organization, are not included in our data set. For our study, we focus on public repositories to analyze engagement with the open-source community.

[5] For the purpose of this study, we conceptually consider engagement with OSCs as so-called "outbound openness," as suggested by Lin and Maruping (2022).

[6] The full list of Crunchbase industry groups is available at https://support.crunchbase.com/hc/en-us/articles/360043146954-What-Industries-are-included-in-Crunchbase- (accessed August 4, 2023).

[7] See https://GitHub.com/search?q=is:public (accessed April 25, 2021).

[8] See https://techcrunch.com/2018/06/04/microsoft-has-acquired-GitHub-for-7-5b-in-microsoft-stock/ (accessed April 25, 2021).

[9] Please refer to https://docs.github.com/en/pull-requests/collaborating-with-pull-requests/working-with-forks/about-permissions-and-visibility-of-forks (accessed November 10, 2024).

[10] One may be worried that startups disclosing their website information on their GitHub profile are more outward-facing and visible to investors in unobserved ways. However, in analyses available upon request, we find no significant differences in the number of public repositories or followers between GitHub accounts with and without website information. Moreover, accounts that have external activities are even less likely to disclose a website.

[11] A watch event is the bookmarking ("starring") of another user's repository. A push event is the "pushing" of commits to a given repository.

[12] As a robustness check, we collected additional data on startups' web technology usage using the *Wappalyzer* technology profiler accessed through the HTTPArchive. We then manually classified the 50 most used web technologies with respect to whether they are open source and which of the use-cases—SD/BE, ML, API, UI—they fall under. This exercise shows that the web technologies on which startups rely are 70% open source, and cover the entire range of use-cases we observe on GitHub. We exclude from the categorization approximately 18% of the total repositories, for which we could not identify a coherent category.

[13] https://www.reddit.com/r/SideProject/comments/qhpjau/how_to_launch_on_product_hunt_a_detailed_guide/ (accessed November 10, 2024).

[14] Specifically, we constructed 10 bins for the software share index with cutoff values at 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1 and derived

comparison startups among those whose index belongs to the same software share bin as the focal startup.

[15] By CXOs, we refer to Chief Executive Officers, Chief Technology Officers, Chief Financial Officers, and Chief Marketing Officers.

[16] Refer to: https://www.crunchbase.com/discover/people (accessed March 4, 2022). We consider the first 1,000 entrepreneurs on Crunchbase's list as top-ranked.

[17] By technology life cycle, we refer to the different phases of the development of a startup's technology (Conti and Graham 2020). It delineates all the stages from the conception of an idea to the prototype phase and further to the mature development and commercialization stages of a startup's technology.

[18] Table A1 in the Online Appendix presents a balance table, demonstrating that the differences in the means of various predetermined startup characteristics between startups engaged with OSCs on GitHub and the comparison group are significantly reduced after applying our matching algorithm. For robustness, we implement a host of different variants of our matching algorithm, including matching on industry group keywords.

[19] Standard errors do not substantially change if we cluster them at the startup level. However, we prefer to cluster standard errors at the level of the observation group because this method tends to produce more conservative (that is, larger) confidence intervals.

[20] Graduates from top universities are (1) more likely to be early adopters of new tools like GitHub (Catalini and Tucker 2017); and (2) more adept at generating innovative products or more connected with potential investors.

[21] In Figure A3 of the Online Appendix, we provide further robustness tests suggesting that the parallel trends assumption holds within large bounds, using the approach of Rambachan and Roth (2023).

[22] Encouragingly, Table A11 in the Online Appendix shows that startups engaging solely with their internal repositories are very similar across a wide range of observable characteristics to those startups that also engage with OSCs on GitHub.

[23] We observe 7,516 unique combinations of industry group keywords in our data set. Using alternative cutoffs does not change the results.

[24] Using more stringent cutoffs, such as considering startups that were founded after 2013, does not change the results.

[25] As reported in Table A9, columns (5) and (6), in the Online Appendix, the results remain qualitatively unchanged when we control for the cumulative number of new hires and when we distinguish between hires who graduated from top universities and other hires.

[26] Unreported analyses unveil that in the average month where we observe a change to the readme files, we also observe other activities, most often push events—that is, changes to code in the repo. This again suggests that startups do not only change the documentation of their codebase to increase visibility to funders, but that changes to readme files reflect changes in the codebase.

[27] The use of the amount of funding raised as a performance outcome is fairly established in the literature (Nanda and Rhodes-Kropf 2013).

[28] Note that the fact that we observe a positive impact of engagement with OSCs on attracting non-VC investors is somewhat encouraging, considering that concerns about reverse causality may be more pronounced when startups pursue VC investments.

[29] https://github.com/enterprise/startups (accessed April 9, 2024).

[30] To address potential unobservable OSS usage, we analyzed a random 10% sample of internal repositories, finding that 60% used OSS packages without observable forking or contributions. However, only 8% of forked repositories were packages, whereas most involved deeper collaboration. Thus, although startups use OSS in unobservable ways, our analysis captures the more intensive, collaborative activities, like forking and contributing, meaning that any measurement error likely understates the true impact of OSS on startup innovation and performance.

## References

Adelino M, Ma S, Robinson D (2017) Firm age, investment opportunities, and job creation. *J. Finance* 72(3):999–1038.

Agarwal R, Audretsch D, Sarkar M (2007) The process of creative construction: Knowledge spillovers, entrepreneurship, and economic growth. *Strategic Entrepreneurship J.* 1(3–4):263–286.

Agarwal R, Audretsch D, Sarkar M (2010) Knowledge spillovers and strategic entrepreneurship. *Strategic Entrepreneurship J.* 4(4):271–283.

Agrawal A, Gans JS, Goldfarb A (2019) Artificial intelligence: The ambiguous labor market impact of automating prediction. *J. Econom. Perspect.* 33(2):31–50.

Almirall E, Casadesus-Masanell R (2010) Open versus closed innovation: A model of discovery and divergence. *Acad. Management Rev.* 35(1):27–47.

Amore MD, Conti A, Pelucco V (2023) Micro venture capital. *Strategic Entrepreneurship J.* 17(4):886–924.

Arora A, Belenzon S, Patacconi A (2018) The decline of science in corporate R&D. *Strategic Management J.* 39(1):3–32.

Arts S, Cassiman B, Gomez JC (2018) Text matching to measure patent similarity. *Strategic Management J.* 39(1):62–84.

Baum JAC, Silverman B (2004) Picking winners or building them? Alliance, intellectual, and human capital as selection criteria in venture financing and performance of biotechnology startups. *J. Bus. Venturing* 19(3):411–436.

Benzell S, Hersh JS, Van Alstyne MW, Lagarda G (2019) How APIs create growth by inverting the firm. Preprint, submitted August 16, https://dx.doi.org/10.2139/ssrn.3432591.

Bernstein S, Korteweg A, Laws K (2017) Attracting early-stage investors: Evidence from a randomized field experiment. *J. Finance* 72(2):509–538.

Bonaccorsi A, Rossi C (2003) Why open source software can succeed. *Res. Policy* 32(7):1243–1258.

Boudreau K (2010) Open platform strategies and innovation: Granting access vs. devolving control. *Management Sci.* 56(10):1849–1872.

Buss P, Peukert C (2015) R&D outsourcing and intellectual property infringement. *Res. Policy* 44(4):977–989.

Callaway B, Sant'Anna PH (2021) Difference-in-differences with multiple time periods. *J. Econometrics* 225(2):200–230.

Catalini C, Tucker C (2017) When early adopters don't adopt. *Science* 357(6347):135–136.

Chesbrough H (2006) New puzzles and new findings. Chesbrough H, Vanhaverbeke W, West J, eds. *Open Innovation: Researching a New Paradigm* (Oxford University Press, Oxford, UK), 15–34.

Choudhury P, Allen RT, Endres MG (2021) Machine learning for pattern discovery in management research. *Strategic Management J.* 42(1):30–57.

Cohen WM, Levinthal DA (1990) Absorptive capacity: A new perspective on learning and innovation. *Admin. Sci. Quart.* 35(1):128–152.

Conti A, Graham SJ (2020) Valuable choices: Prominent venture capitalists' influence on startup CEO replacements. *Management Sci.* 66(3):1325–1350.

Conti A, Guzman JA (2021) What is the US comparative advantage in entrepreneurship? Evidence from Israeli migration to the United States. *Rev. Econom. Statist.* 105(3):528–544.

Conti A, Roche MP (2021) Lowering the bar? External conditions, opportunity costs, and high-tech start-up outcomes. *Organ. Sci.* 32(4):965–986.

Conti A, Santaló J (2023) The hidden costs of fairness in platform markets: The dynamics of lowering developer royalty rates. Preprint, submitted April 12, https//dx.doi.org/10.2139/ssrn.4403845.

Conti A, Guzman J, Rabi R (2020) Information frictions in the market for startup acquisitions. Preprint, submitted August 25, https://dx.doi.org/10.2139/ssrn.3678676.

Conti A, Thursby M, Rothaermel FT (2013a) Show me the right stuff: Signals for high-tech startups. *J. Econom. Management Strategy.* 22(2):341–364.

Conti A, Thursby J, Thursby M (2013b) Patents as signals for startup financing. *J. Indust. Econom.* 61(3):592–622.

Conti A, Gupta V, Guzman J, Roche MP (2023) Incentivizing innovation in open source: Evidence from the GitHub sponsors program. NBER Working Paper 31668, National Bureau of Economic Research, Cambridge, MA.

Contigiani A (2023) Experimentation and appropriability in early-stage ventures: Evidence from the US software industry. *Strategic Management J.* 44(9):2128–2174.

Cunningham C, Kapacinskaite A (2024) Keeping invention confidential. Technical report.

Dahlander L, Magnusson M (2008) How do firms make use of open source communities? *Long Range Planning* 41(6):629–649.

Dushnitsky G, Matusik SF (2019) A fresh look at patterns and assumptions in the field of entrepreneurship: What can we learn? *Strategic Entrepreneurship J.* 13(4):437–447.

Ewens M, Marx M (2018) Founder replacement and startup performance. *Rev. Financial Stud.* 31(4):1532–1565.

Fleming L (2001) Recombinant uncertainty in technological search. *Management Sci.* 47(1):117–132.

Foege JN, Lauritzen GD, Tietze F, Salge TO (2019) Reconceptualizing the paradox of openness: How solvers navigate sharing-protecting tensions in crowdsourcing. *Res. Policy* 48(6):1323–1339.

Fosfuri A, Giarratana MS, Luzzi A (2008) The penguin has entered the building: The commercialization of open source software products. *Organ. Sci.* 19(2):292–305.

Franco AM, Sarkar M, Agarwal R, Echambadi R (2009) Swift and smart: The moderating effects of technological capabilities on the market pioneering–firm survival relationship. *Management Sci.* 55(11):1842–1860.

Furman JL, Teodoridis F (2020) Automation, research technology, and researchers' trajectories: Evidence from computer science and electrical engineering. *Organ. Sci.* 31(2):330–354.

Gans JS, Stern S (2003) The product market and the market for "ideas": Commercialization strategies for technology entrepreneurs. *Res. Policy* 32(2):333–350.

Gans JS, Kearney M, Scott EL, Stern S (2021) Choosing technology: An entrepreneurial strategy approach. *Strategy Sci.* 6(1):39–53.

Garcia-Swartz DD, Campbell-Kelly M (2019) Openness as a business strategy: Historical perspectives on openness in computing and mobile phones. *Inform. Econom. Policy* 48:1–14.

Germonprez M, Kendall JE, Kendall KE, Mathiassen L, Young B, Warner B (2017) A theory of responsive design: A field study of corporate engagement with open source communities. *Inform. Systems Res.* 28(1):64–83.

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

Gompers P, Lerner J (2001) The venture capital revolution. *J. Econom. Perspect.* 15(2):145–168.

Gompers PA, Gornall W, Kaplan SN, Strebulaev IA (2020) How do venture capitalists make decisions? *J. Financial Econom.* 135(1): 169–190.

Greenstein S, Nagle F (2014) Digital dark matter and the economic contribution of Apache. *Res. Policy* 43(4):623–631.

Gruber M, Henkel J (2006) New ventures based on open innovation—An empirical analysis of start-up firms in embedded Linux. *Internat. J. Technol. Management* 33:356–372.

Guzman J, Li A (2023) Measuring founding strategy. *Management Sci.* 69(1):101–118.

Haefliger S, von Krogh G, Spaeth S (2008) Code reuse in open source software. *Management Sci.* 54(1):180–193.

Haese J, Peukert C (2024) Open at the core: Moving from proprietary technology to building a product on open source software. *Management Sci*. Forthcoming.

Hellmann T, Puri M (2002) Venture capital and the professionalization of start-up firms: Empirical evidence. *J. Finance* 57(1): 169–197.

Hochberg YV, Serrano CJ, Ziedonis RH (2018) Patent collateral, investor commitment, and the market for venture lending. *J. Financial Econom.* 130(1):74–94.

Hoffmann M, Nagle F, Zhou Y (2024) The value of open source software. Harvard Business School Strategy Unit Working Paper 24-038, Harvard Business School, Boston.

Hsu DH, Ziedonis RH (2013) Resources as dual sources of advantage: Implications for valuing entrepreneurial-firm patents. *Strategic Management J.* 34(7):761–781.

Jain V (2018) Investor due diligence: Beyond the obvious. Accessed June 29, 2021, https://startupflux.com/blogs/investor-due-diligence-beyond-the-obvious/.

Jordan MI, Mitchell TM (2015) Machine learning: Trends, perspectives, and prospects. *Science* 349(6245):255–260.

Kaplan SN, Sensoy BA, Strömberg P (2009) Should investors bet on the jockey or the horse? Evidence from the evolution of firms from early business plans to public companies. *J. Finance* 64(1):75–115.

Katz ML, Shapiro C (1994) Systems competition and network effects. *J. Econom. Perspect.* 8(2):93–115.

Kortum S, Lerner J (2000) Assessing the contribution of venture capital to innovation. *RAND J. Econom.* 31(4):674–692.

Laursen K, Salter AJ (2014) The paradox of openness: Appropriability, external search and collaboration. *Res. Policy.* 43(5):867–878.

Lee SR, Kim JD (2024) When do startups scale? Large-scale evidence from job postings. *Strategic Management J.* 45(9):1633–1669.

Leiblein MJ, Chen JS, Posen HE (2023) Uncertain learning curves: Implications for first-mover advantage and knowledge spillovers. *Acad. Management Rev.* 48(1):123–148.

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

Lieberman MB (1987) The learning curve, diffusion, and competitive strategy. *Strategic Management J.* 8(5):441–452.

Lin Y-K, Maruping LM (2022) Open source collaboration in digital entrepreneurship. *Organ. Sci.* 33(1):212–230.

Lyons E, Zhang L (2018) Who does (not) benefit from entrepreneurship programs? *Strategic Management J.* 39(1):85–112.

Majchrzak A, Cooper LP, Neece OE (2004) Knowledge reuse for innovation. *Management Sci.* 50(2):174–188.

Marinoni A, Roche MP (2024) You've got mail! The late 19th century us postal service expansion, entrepreneurship, and firm performance. *Management Sci.* Forthcoming.

Miric M, Jia N, Huang KG (2023) Using supervised machine learning for large-scale classification in management research: The case for identifying artificial intelligence patents. *Strategic Management J.* 44(2):491–519.

Nagle F (2018) Learning by contributing: Gaining competitive advantage through contribution to crowdsourced public goods. *Organ. Sci.* 29(4):569–587.

Nagle F (2019) Open source software and firm productivity. *Management Sci.* 65(3):1191–1215.

Nambisan S (2017) Digital entrepreneurship: Toward a digital technology perspective of entrepreneurship. *Entrepreneurship Theory Pract.* 41(6):1029–1055.

Nanda R, Rhodes-Kropf M (2013) Investment cycles and startup innovation. *J. Financial Econom.* 110(2):403–418.

Oberholzer-Gee F (2021) *Better, Simpler Strategy: A Value-Based Guide to Exceptional Performance* (Harvard Business Press, Boston).

Rambachan A, Roth J (2023) A more credible approach to parallel trends. *Rev. Econom. Stud.* 90(5):2555–2591.

Roche MP, Oettl A, Catalini C (2024) Proximate (co-)working: Knowledge spillovers and social interactions. *Management Sci.*, ePub ahead of print February 14, https://doi.org/10.1287/mnsc.2022.03555.

Rumelt RP (1984) Towards a strategic theory of the firm. *Competitive Strategic Management* 26(3):556–570.

Rysman M, Simcoe T (2008) Patents and the performance of voluntary standard-setting organizations. *Management Sci.* 54(11): 1920–1934.

Schumpeter JA (1934) *The Theory of Economic Development; An Inquiry into Profits, Capital, Credit, Interest, and the Business Cycle* (Harvard University Press, Cambridge, MA).

Shah S (2006) Motivation, governance, and the viability of hybrid forms in open source software development. *Management Sci.* 52(7):1000–1014.

Shah S, Nagle F (2019) Why do user communities matter for strategy? Harvard Business School Strategy Unit Working Paper 19-126, Harvard Business School, Boston.

Shane S, Stuart T (2002) Organizational endowments and the performance of university start-ups. *Management Sci.* 48(1):154–170.

Stam W (2009) When does community participation enhance the performance of open source software companies? *Res. Policy* 38(8):1288–1299.

Teece DJ (1986) Profiting from technological innovation: Implications for integration, collaboration, licensing and public policy. *Res. Policy* 15(6):285–305.

Tranchero M (2023) Finding diamonds in the rough: Data-driven decisions and pharmaceutical innovation. Working paper, University of Pennsylvania, Philadelphia.

Tzabbar D, Margolis J (2017) Beyond the startup stage: The founding team's human capital, new venture's stage of life, founder–CEO duality, and breakthrough innovation. *Organ. Sci.* 28(5):857–872.

Veugelers R, Schneider C (2018) Which IP strategies do young highly innovative firms choose? *Small Bus. Econom.* 50:113–129.

West J, Kuk G (2016) The complementarity of openness: How MakerBot leveraged Thingiverse in 3D printing. *Technol. Forecasting Soc. Change* 102:169–181.

Wooldridge JM (2021) Two-way fixed effects, the two-way Mundlak regression, and difference-in-differences estimators. Preprint, submitted August 18, https://dx.doi.org/10.2139/ssrn.3906345.

Wright NL, Nagle F, Greenstein S (2023) Open source software and global entrepreneurship. *Res. Policy* 52:104846.

Xu L, Nian T, Cabral L (2020) What makes geeks tick? A study of stack overflow careers. *Management Sci.* 66(2):587–604.

Yoo Y, Boland RJ Jr, Lyytinen K, Majchrzak A (2012) Organizing for innovation in the digitized world. *Organ. Sci.* 23(5): 1398–1408.

Yu S (2020) How do accelerators impact the performance of high-technology ventures? *Management Sci.* 66(2):530–552.

Zahra SA, George G (2002) Absorptive capacity: A review, reconceptualization, and extension. *Acad. Management Rev.* 27(2):185–203.

**Annamaria Conti** professor of strategy at IE Business School, is Associate Editor of *Management Science* and on the Editorial Review Board of *Strategy Science*. Formerly at HEC Lausanne and Georgia Tech, her research on entrepreneurship and innovation examines startups and venture capital strategies. Published in top journals, she's earned awards like the 2022 Strategy Science Best Paper. Before academia, she was an economist at Caterpillar S.a.r.l and international organizations.

**Christian Peukert** is an associate professor at HEC Lausanne, University of Lausanne, specializing in digitization, innovation, and intellectual property. His research focuses on how digital technologies and regulation impact markets, firms, and consumers, with emphasis on data, AI, and intellectual property. Published in top journals, his work has earned awards from WISE, INFORMS, and *Strategy Science*. Before academia, he cofounded a rap-focused record label.

**Maria Roche** assistant professor at Harvard Business School, specializes in strategy and innovation, focusing on microgeography's role in knowledge commercialization. Her work appears, a.o., in *Management Science, The Review of Economics and Statistics*, and *Organization Science*, and has been featured in *The Economist*, and *The WSJ*. She serves on the editorial review boards of this journal and Strategy Science and received dissertation awards from AOM and EGOS. Maria earned her PhD at Georgia Tech.