

Opt-in to Innovation? The Effects of Apple’s App Tracking Transparency Policy on Complementor Innovation

Robin Nowak (University of Passau), Patrick Figge (University of the Bundeswehr Munich),
Rebecca Janßen (ZEW – Leibniz Centre for European Economic Research)*

Abstract

Platform operators face a continuous trade-off in governance decisions between addressing user interests and preserving incentives for complementor innovation. One highly debated area of platform governance concerns protecting user privacy by restricting third-party tracking, which, however, is vital to many complementors’ value capture models. In this paper, we seek to understand the implications of stricter privacy governance on complementor innovation. We examine how mobile app developers adjust their innovation activities in response to Apple’s App Tracking Transparency (ATT) policy. Using a quasi-experimental difference-in-differences framework, we compare the innovation outcomes (feature updates) for apps affected by the policy with those unaffected, before and after the policy announcement. Our results indicate a positive effect of ATT on affected apps’ innovation relative to unaffected apps after the policy announcement. Further analyses are consistent with an adaptive response to reduced third-party tracking: the innovation response is stronger among more exposed apps, and affected developers innovate to support alternative ways to capture value. We discuss that while stricter privacy governance may initially seem to negatively affect complementor innovation, its nuanced effects can also serve as strategic catalysts and enable platform operators to reshape complementors’ innovation activities.

Keywords: Innovation, Platform Governance, Economics of Privacy, Targeted Advertising, Mobile Apps

* Please send correspondence to robin.nowak@uni-passau.de.

We are grateful for very helpful comments from Néstor Duch-Brown, Jens Foerderer, Grace Gu, Reinhold Kesler, Jan Krämer, Tobias Kretschmer, Thomas Kude, Maria Luisa Mancusi, Raveesh Mayya, Milan Miric, Daniel Obermeier, Bernd Skiera, and participants of the Platform Strategy Research Symposium and the 10th ZEW Conference on the Economics of Innovation and Patenting as well as seminars at LMU Munich and the University of Passau. Robin Nowak gratefully acknowledges funding by the German Research Foundation—project number 443732464. Generative AI tools were used to assist with coding and language editing; we thoroughly checked all AI-assisted work and take full responsibility for the manuscript’s content.

1. Introduction

Innovation by complementors is central to value creation in platform-based ecosystems (Parker and Van Alstyne 2005, Yoo et al. 2010, Tiwana 2015). Typically, the platform operator defines the operational scope for complementors through a set of governance policies (Tiwana et al. 2010, Wareham et al. 2014), which shape how innovation occurs within the ecosystem (e.g., Boudreau 2010, Conti and Santaló 2026). A frequently observed development is that initially, while the platform operator is focused on growth, governance supports a broad range of complementors and encourages their innovation activities. Over time, as the platform operator's focus shifts to value capture, they often become more selective and focused on meeting user needs (Rietveld et al. 2020). At this stage, complementors find it difficult to leave, even if they experience unfavorable changes to platform governance (Rietveld and Schilling 2021, Cutolo and Kenney 2021, Rahman et al. 2024). This has sparked great interest in exploring the behavior of complementors following such governance changes and the implications for ecosystem value.

One set of user-focused governance policies that has become the subject of much debate centers on user privacy and targeted advertising (Goldfarb and Que 2023). Targeted advertising involves personalizing ads displayed to users based on their behavioral, demographic, and location data. This form of advertising is central to many complementors' value capture models, especially in platform-based ecosystems like Apple's iOS and Google's Android, where products are often offered for free and users' willingness to pay is generally low (e.g., Casadesus-Masanell and Hervas-Drane 2015). With stricter privacy governance, the availability of user data is limited, reducing the effectiveness of targeted advertising. A broad literature documents that such privacy restrictions can impose economic costs on data-dependent business models (Goldfarb and Tucker 2011, Acquisti et al. 2016). Mobile app developers in particular are affected if they monetize user data, and they respond strategically to changes in platform privacy policies (Kummer and Schulte 2019, Mayya and Viswanathan 2025). Recent findings by Kircher and Foerderer (2024) point to a substantial drop in overall innovation activities following Google's 2019 ban on targeted advertising in certain app categories on the Android platform. These results raise important questions on how privacy governance can be effectively designed without stifling innovation. Other governance changes where

complementors' value capture was threatened, for instance by the selective promotion of competitors (Rietveld et al. 2019, 2021, Foerderer et al. 2021, Agarwal et al. 2023) or the platform operator's market entry (Foerderer et al. 2018, Wen and Zhu 2019, Shi et al. 2023), led to a reorientation or even an increase in innovation efforts.

Our study seeks to contribute to these questions by investigating the implications of Apple's App Tracking Transparency (ATT) policy for complementors and the iOS ecosystem. In June 2020, Apple announced that it would introduce a privacy-preserving feature that requires explicit consent from users to track them across different apps in the next large iOS 14 update. If enough users deny consent, this could impair the personalized targeting and attribution of advertising, with negative effects for the app's advertising revenue and ad user acquisition. After strong opposition from developers, Apple delayed the implementation until the iOS 14.5 update in April 2021. The ATT policy is intriguing for three reasons. First, this setting allows us to study complementors' reactions to privacy governance that reduces their value-capture opportunities across all app categories in the iOS ecosystem. Many apps rely on cross-app tracking to generate revenue from targeted advertising or to acquire new users through targeted campaigns. Requiring users to opt in to activity tracking makes both directly dependent on user consent. In fact, the ATT policy has reduced revenue streams from app advertising significantly (Kraft et al. 2024). Second, ATT is not a full ban of (targeted) advertising. Instead, it represents an intermediate, consent-based restriction that constrains but does not eliminate targeted advertising, providing insights into the implications of such moderated privacy governance. Third, the policy announcement created a plausible exogenous governance shock that developers could respond to before implementation, allowing us to examine anticipatory adjustment as well as subsequent responses.

We use comprehensive app- and developer-level data from the iOS App Store to track complementor innovation over time. Our primary analyses focus on apps that appeared at least once in the App Store's weekly Top-100 rankings during the observation period. We identify exposure to the ATT policy change based on apps' pre-announcement integration of ad-related software development kits (SDKs). Specifically, we posit that apps integrating SDKs with advertising functionality are likely designed

to capture value through the monetization of advertising spaces or rely on cross-app tracking for advertising campaigns. Conversely, apps without such SDKs are presumed to be unaffected. We model complementor innovation as the monthly frequency of updates that introduce new features and content to an app. To operationalize this measure, we analyze each update’s release notes using a GPT-4.1 Mini model that is fine-tuned for this purpose (Nowak et al. 2025). To examine how developers innovate, we further classify the content of feature updates by the type of feature they introduce. We then estimate difference-in-differences models and event-study specifications that compare apps affected by the policy with those unaffected, before and after the policy announcement.

Our results show that the ATT announcement is associated with a decline in innovation activity. At the same time, this aggregate decline masks substantial heterogeneity across apps. The decline is significantly smaller for ATT affected apps, implying a relative increase in innovation among these apps. Event-study estimates show that this divergence sets in after announcement and widens after the technical implementation of the policy. The response is stronger where exposure to the policy is larger: in app categories with lower tracking opt-in rates, among apps whose in-app monetization is limited to advertising, and among developers with small app portfolios. We further find evidence consistent with innovation as an adaptation to reduced third-party tracking. Affected apps do not change their monetization model. The content of the additional feature updates points to improvements in core app functions and design, extensions of paid offerings within the existing model, and features that help attract and retain users and support the collection and use of first-party data. The effect extends to the full population of iOS apps at a smaller magnitude.

Our study contributes to three strands of literature. First, we extend research on platform governance (Tiwana et al. 2010, Wareham et al. 2014, Rietveld et al. 2019, Conti and Santaló 2026) by showing that privacy governance can reshape complementor innovation by inducing adaptive responses among exposed complementors. Second, we contribute to the economics of privacy (Goldfarb and Tucker 2012, Campbell et al. 2015, Burtch et al. 2015, Acquisti et al. 2016, Gopal et al. 2018) by showing that the consequences of privacy restrictions for innovation activities depend on their design. Compared with outright bans (Kircher

and Foerderer 2024), consent-based restrictions may preserve room for adaptation and thereby sustain or redirect complementor innovation. Third, we add to the emerging empirical literature on ATT and the timing of adaptive reactions (Kraft et al. 2024, Cheyre et al. 2026, Kesler 2026) by showing that developer responses began already at the announcement stage and varied systematically with exposure to ATT, which speaks against treating apps across the iOS ecosystem as equally affected. Further, responses likely took the form of product extensions or improvements within developers' existing value capture models.

2. Related Literature

Digital platforms are systems in which a wide range of complementors independently contribute to a generative infrastructure provided and orchestrated by a platform operator (Tilson et al. 2010, Parker and Van Alstyne 2018, Bhargava 2022). Platform operators compete with other platforms by attracting new complementors to their platform and by encouraging existing complementors to continuously improve their value propositions in ways that benefit platform users and the entire ecosystem.

Research on platform governance has studied various decisions that platform operators make to shape complementors' incentives and activities. Platform governance includes pricing choices (Parker and Van Alstyne 2005, Hagiu 2006), partnership programs (Ceccagnoli et al. 2012), boundary resources (Ghazawneh and Henfridsson 2013, Miric et al. 2023), access control (Casadesus-Masanell and Halaburda 2014, Zhang et al. 2022), performance investments (Anderson et al. 2014), opportunities for interfirm exchange (Foerderer 2020), and property rights (Miric and Jeppesen 2020). At first, these decisions often focus on building an attractive offering of complements. As the platform matures, the operator's focus often shifts to value capture and governance decisions are primarily concerned with users' perception of the platform. Promotion of complements may become more selective (Claussen et al. 2013, Rietveld et al. 2019, 2021, Foerderer et al. 2021, Agarwal et al. 2023), algorithmic management may be used to steer complementor behavior toward platform objectives and to engage in forms of self-preferencing through ranking and recommendation systems (Curchod et al. 2020, Moehlmann et al. 2021), and platform operators may even decide to enter attractive markets for some complements themselves (Gawer and Henderson 2007, Huang et al. 2013, Zhu and Liu 2018, Foerderer et al. 2018, Wen and Zhu 2019, Shi et al. 2023). Although

these governance decisions may threaten complementors' value capture potential and reduce their rents from their contributions to the platform, they may find it difficult to leave the platform (Cutolo and Kenney 2021, Rahman et al. 2024). Complementors often continue to contribute to the platform and may reorient or try to enhance their contributions. Platform operators use such governance policies strategically to steer the value proposition in ways that they consider important for overall ecosystem value (Rietveld et al. 2019).

We relate to this literature by focusing on a widely debated but less studied platform governance decision: enforcing stricter user privacy. Empirical work in this area is still nascent. In the research that relates closest to ours, Kircher and Foerderer (2024) investigate the impact of Google's ban of targeted advertising in children's games on Android, revealing that the ban reduced the release of feature updates, particularly for games of young, undiversified, and ad-dependent firms. An increase in development efforts has only been detected for high-quality and high-demand games. The drop in development efforts is intriguing as it raises questions about how such governance decisions should be introduced in broader roll-outs without stifling the ecosystem's innovation activities and overall viability. We address this issue by investigating the effects of Apple's ATT policy. The ATT policy differs from the ban in three ways that allow us to contribute additional insights to this line of research: (i) ATT is not a full ban of (targeted) advertising: it requires users to actively opt-in to tracking and thereby reduces returns from targeted advertising, but does not prohibit it; (ii) ATT does not only affect a niche in a specific app category but a large share of the iOS platform as many apps rely on cross-app tracking to monetize advertising and to acquire users; (iii) ATT applies to all apps that rely on tracking, so developers cannot escape it by shifting efforts to other tracking-based apps; adaptation has to take place within the affected apps or through apps with a different monetization model.

Notably, several papers have provided first empirical insights into the ATT policy's impact from a complementor perspective. Kesler (2026) examines the shift in app monetization models due to ATT, revealing a modest increase in paid apps and a trend towards more in-app payments. Cheyre et al.'s (2026) findings indicate a rapid recovery in the number of apps on the Apple App Store following an initial decline after ATT's implementation. They also observe shifts in app components, with a decrease in the use of

monetization- and ad mediation-components and an increase in authentication- and payment-related components. This suggests developers adapted their strategies to align with the new privacy framework, rather than withdrawing from the market. Kraft et al. (2024) analyze the ATT policy using ad impression data across various countries. In line with conceptual arguments of Sokol and Zhu (2021), they demonstrate that ATT significantly reduced the trackable traffic in the U.S. and led to a decrease in ad revenue for publishers. Our study differs from this research. We focus on developers' product development choices to understand how they adjust their innovation activities in response to the ATT policy. Furthermore, we argue that a differentiated perspective within the iOS ecosystem might be necessary, as many apps do not primarily monetize through advertising and may therefore be less affected by the policy than is often assumed when treating the entire ecosystem as uniformly impacted (Gao et al. 2025). This distinction also matters for timing: because ATT was announced well before its implementation, developers already had reason to anticipate the policy change and begin adjusting their strategies after the announcement rather than only after implementation. By offering this more nuanced view, our study aims to deepen the understanding of the policy's implications on innovation and to shed light on the potential mechanisms that influence the behavior of affected developers.

Last, given that we study a platform's privacy-enforcing initiative, our study is also related to the research that is concerned with understanding the economic consequences of privacy initiatives. This strand of literature considers data to be valuable for firms, for instance, by leveraging it as an input into innovation, enabling targeted advertising, or creating a direct revenue stream through selling the data to third parties (Kummer and Schulte 2019). Initiatives to regulate access to these data by increasing user privacy have been studied across a variety of contexts, revealing their impact on market structure and competition (Campbell et al. 2015, Casadesus-Masanell and Hervas-Drane 2015, Peukert et al. 2022, Johnson et al. 2023), diffusion of technology (Adjerid et al. 2016, Miller and Tucker 2018), product use (Goldberg et al. 2024, Sun et al. 2024), advertising effectiveness (Goldfarb and Tucker 2012, Johnson et al. 2020), and innovation (Jia et al. 2021, Janßen et al. 2022), among others. It is important to note that these studies have been concerned with government-imposed privacy regulations. But firms might also have reason to protect

consumer privacy, even in the absence of public regulation (Tsai et al. 2011, Goldfarb and Que 2023). For instance, addressing users' privacy concerns can affect business model choices of firms (Gopal et al. 2018) and be a lever to mitigate competition (Lee et al. 2011), aspects that are also in the interest of platform operators. Investigating the effects of non-regulatory privacy initiatives empirically remains an underexplored question, particularly with first empirical work pointing to unintended consequences from strategic complementor responses (Mayya and Viswanathan 2025) and negative consequences on firm innovation (Kircher and Foerderer 2024).

3. The iOS App Platform and Apple's App Tracking Transparency Policy

We empirically study complementors' responses to stricter privacy governance in the context of Apple's iOS app platform and the introduction of the ATT policy. The iOS app platform is integral to mobile application development and distribution. Developers build their apps on top of the infrastructure provided by Apple and thus contribute value to the platform. To keep users engaged with an app, developers improve their apps continuously, and introduce new features, fix bugs, and improve user experience.

To capture value from an app and its extensions, developers have several options. Direct value capture involves charging an upfront price for the app download or integrating in-app purchases. In-app purchases enable users to buy content, features, or digital goods either through single purchases or subscriptions within the app. Indirect value capture involves collecting user data such as usage patterns, purchase history, identifiers, and location, and leveraging this information to sell in-app advertising spaces through networks like Google AdMob or Facebook Ads (Audience Network). These ad networks aggregate digital ad spaces from various apps and broker them to advertisers, with personalized advertising delivering the most significant returns (Sharma et al. 2021). Hybrid or so-called freemium approaches, in which a basic version is offered for free and additional features or content can be purchased, can combine both approaches. As digital goods such as apps are experience goods and thus charging an upfront price can be challenging (e.g., Casadesus-Masanell and Hervas-Drane 2015), freemium approaches and ad-based models are predominant in app markets.

On June 22, 2020, Apple announced at its developer conference that the App Tracking Transparency policy would be an integral part of the following iOS 14 update (scheduled for release in September 2020). The announcement specified that apps would be required to obtain user permission for any tracking across apps, explicitly including advertising measurement and data sharing with third parties.¹ It received broad attention among complementors and in the advertising industry. Major advertising-dependent complementors such as Facebook (now Meta) even publicly opposed the policy (Online Appendix A1). In September 2020, Apple postponed enforcement until “early next year” to give complementors time “to make necessary changes” (Apple 2020). The policy was ultimately implemented at short notice on April 26, 2021, with the launch of iOS 14.5.

This policy mandates that app developers integrate an opt-in consent prompt into their apps, if they want to track users’ activities outside their own app (e.g., Figure 1). The consent prompt appears when the user opens an already downloaded or newly downloaded app for the first time after updating to the new iOS version, unless they have already opted out of tracking in their general settings. Importantly, apps cannot base their functionality on the user’s consent decision. Users can continue utilizing the app regardless of their consent decision. If consent is denied, the app is unable to obtain the user’s Identifier for Advertisers (IDFA). The IDFA is a random device identifier that works as an accurate means to track user behavior. Without consent, complementors cannot ascribe advertising success to an individual user or enhance user ad profiles with personalized information from app usage across apps.

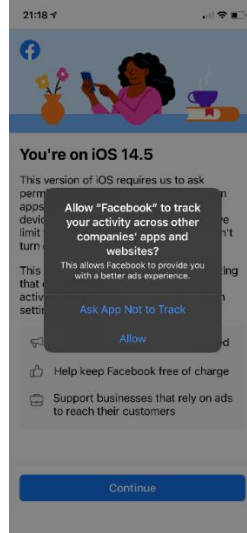
The success of many apps’ value capture strategy is directly dependent on the opt-in rate. The price for trackable ad impressions in the United States is 51% higher than for non-trackable ad impressions. The share of trackable impressions fell from 73% in April 2021, before ATT, to 18% in March 2022, after ATT.

¹ The press release from Apple published that day provided developers with detailed information about the upcoming tracking requirement, including the following: “All apps will now be required to obtain user permission before tracking. [...] This includes connecting information collected separately by other companies for targeted advertising, for advertising measurement, or via data brokers.”

See: <https://www.apple.com/newsroom/2020/06/apple-reimagines-the-iphone-experience-with-ios-14/> (Retrieved October 14, 2026).

Combined with the price premium, this reduction of 55 percentage points reflects a 21% decrease in advertising revenue (see Kraft et al. 2024 for a comprehensive study of the economic effects of ATT).

Figure 1. Prompt for App-Tracking Authorization Request



Notes: This figure shows an exemplary consent prompt for the Facebook app.

4. Data and Methods

4.1 Research Framework

We analyze the effects of Apple's ATT policy in a quasi-experimental difference-in-differences approach.

In our analysis, we compare the innovation outcomes of complementors' apps affected and unaffected by the ATT policy, both before and after the policy's announcement. Our identification strategy involves a close examination of the SDKs the apps are based on. An SDK comprises a suite of software tools and libraries provided by vendors to developers, facilitating the development of applications for specific platforms or devices. SDKs typically include documentation, code samples, processes, and guides necessary to develop and test applications for a specific software framework, hardware platform, or operating system. Apps that rely on SDKs with advertising functionality are likely to collect user data for targeted advertising purposes and are thus affected by the ATT policy. Apps that do not rely on such SDKs are likely unaffected.²

Although some advertising functionalities or access to advertising networks can technically be integrated

² ATT restricts IDFA access for all apps on the platform. As the identifier also underpins data practices beyond advertising, individual apps in the comparison group may have used it for secondary purposes. To the extent such apps were partially exposed, the estimated difference between affected and unaffected apps is attenuated and our estimates are conservative.

server-side without an SDK, the substantial cost savings from using the SDKs encourage most developers to adopt them directly.

Our observation period spans from June 2019, one year before the announcement of Apple’s ATT policy, to September 2021, 15 months after the announcement. We use the announcement on June 22, 2020 as a reference date for the pre- and post-periods. With the announcement, complementors learned about the scope and the consent mechanism of the policy. Strategic responses such as adjusting the monetization model or developing new features require lead time, and Apple itself framed the period until implementation as time for complementors to adapt. Complementors thus had reason to adjust their strategies from the announcement onwards. They could also expect the policy to be enforced, effective, and binding for most of their users. Apple enforces its policies strictly, the ATT framework prevents IDFA access without consent, and major iOS versions reach most users within months (Online Appendix A2). By implementation, they had been aware of the policy and its scope for ten months. The observation period also covers responses to the implementation in April 2021. We end the observation period in September 2021 to limit overlap with the rollout of iOS 15, which introduced additional privacy protections and other platform changes.

4.2 Data Sources and Sample

This study draws on two data sources. First, we obtained app-level data from the app store analytics provider MightySignal. The data contains monthly snapshots of all apps listed on the Apple App Store during our observation period, including app name, developer, category, monetization information, description text, and release date. Second, we obtained complementary SDK-level data as well as app update and ranking information from MixRank, a data provider for mobile apps and app SDKs. The SDK data contains all SDKs that free apps³ from the app-level data have installed or uninstalled, covering 96,876 unique SDKs. For each SDK, we observe its developer, installation and uninstallation dates, description, and tags

³ SDK integrations are detected by downloading an app’s binary and scanning it for the software libraries it contains. This is feasible at scale only for apps that are free to download, so SDK information is not available for paid apps. Paid apps typically do not offer advertising spaces, as their value capture rests on upfront prices. Some paid apps may still integrate advertising-related SDKs to complement their revenues, yet the ATT policy affects their value capture model far less directly.

indicating its usage purpose. These records allow us to retrieve more detailed information about an app’s connection to advertising networks. The app update histories include release notes and changes in installed SDKs, which provide detailed information about update activity.

Our focal sample comprises apps that appeared at least once in the App Store’s weekly Top-100 rankings during our observation period. Activity in the App Store is highly skewed, with a large share of apps receiving only a few downloads and very infrequent updates. For these apps, incentives to innovate and exposure to advertising-related privacy shocks are comparatively weak. In contrast, top-ranked apps account for the bulk of user attention and revenue (Garg and Telang 2013, Ghose and Han 2014) and thus constitute the environment in which the innovation effects of platform governance materialize. We therefore follow prior research that studies app-market outcomes using top-ranked panels as the economically relevant set of interest (e.g., Lee and Raghu 2014, Comino et al. 2019). To construct the sample, we use the index of all apps listed on the App Store as of January 1, 2020. Since our main variables depend on text-based measures, we restrict the sample to apps with descriptions in English.⁴ Of the 1,294,110 apps meeting these criteria, 29,597 appeared at least once in the weekly Top-100 rankings and form the basis of our focal sample. We exclude 3,543 free apps for which no SDK information is observed before the announcement, as their exposure to ATT cannot be determined. The remaining 26,054 apps constitute our focal sample. We additionally draw on the full English-language sample to assess the scope of our findings.

4.3 Metrics of Interest

Ad-based Monetization. To identify apps that are affected by the ATT policy, we need to know whether an app collects the IDFA for individual users and engages in data-collecting and -sharing for targeted advertising purposes. As this information is not included in our main data we build our identification variable based on the SDK dataset. We classified an app as affected if it had integrated an SDK used specifically for advertising purposes before the announcement of the ATT policy (*ATT Affected*). We

⁴ Our classifier is fine-tuned and validated on English release notes. Classification quality for other languages cannot be assured at the same standard. English-language apps also constitute the majority of the App Store index and dominate the Top-100 rankings that define our focal sample. The restriction further ensures that treatment and control apps are drawn from the same language environment so that differences in release note conventions across languages do not affect our main variables. Across all apps observed by MightySignal between January 2018 and September 2021, the restriction to English descriptions excludes 38% of apps. Of the 1,698,532 English-language apps observed during our observation period, 1,294,110 were released before January 1, 2020.

identify such SDKs based on MixRank’s functional tagging of each SDK; overall 1,736 SDKs qualify. These include SDKs of ad networks (e.g., Google AdMob, Meta Audience Network), ad mediation platforms, attribution providers (e.g., AppsFlyer, Adjust), and advertising-related analytics, as well as Apple’s AdSupport framework, which provides apps access to the IDFA. Online Appendix B describes these functions and lists representative SDKs.

Innovation. The main outcome in our analysis is complementor innovation. Previous literature focusing on mobile app platforms used various metrics to measure innovation like entry of new apps (Janßen et al. 2022), frequency of app updates (Foerderer et al. 2018, Wen and Zhu 2019), or the appearance in sales top rankings (Yin et al. 2014). Following Foerderer et al. (2018), we model innovation as complementors’ efforts to further develop their app through new features or content. Hence, we focus on the updating frequency of apps and decompose the type of update. That is, we decompose the number of updates of an app per month into the number of feature updates and maintenance updates, whereby we understand feature updates as the introduction of new features or an extension of the app content. Maintenance updates such as pure bug fixes and performance improvements, simple adjustments to new iOS standards or security-related adjustments, are disregarded for the focal analyses.

To create the corresponding variables, we use data on all updates in the observation period and classify each update based on the corresponding release note. On the App Store, these release notes are displayed in a section titled “What’s New” and limited to 4,000 characters. Since 2018, it is mandatory for developers to specify update details and avoid generic texts. For classification, we follow the methodological framework of Nowak et al. (2025). We fine-tune a GPT-4.1 Mini model on a human-coded training sample of 2,000 randomly selected release notes (cf. Figure 2 for the exact prompt).⁵ The classification includes seven subcategories: (1) novel features; (2) content extensions; (3) support of new OS versions and hardware devices; (4) specific fixes constraining user experience; (5) privacy and security;

⁵ The framework was developed and validated with release notes of iOS apps and thus matches our empirical setting. The training and validation samples are available in the replication package of Nowak et al. (2025).

(6) general bug fixes and performance improvements; (7) marketing and branding, that were derived inductively and validated in interviews with app developers.

Subcategories one and two indicate a *Feature Update* and sub-categories three to seven indicate a *Maintenance Update*. If several subcategories were present, the highest-priority category was selected. The subcategories are ordered by their relevance to our construct of interest. As shipped updates frequently bundle several changes to an app, this rule ensures that an update is classified as a feature update whenever its release note describes new features or content, regardless of accompanying maintenance adjustments. On the corresponding holdout sample of 1,000 randomly selected release notes the classifier reaches an F1-score of .914 for the identification of feature and maintenance updates (temperature = 0, seed = 94032). Three independent runs with the same deterministic setting achieved a 99.3% consistency rate across runs. Online Appendix C provides representative examples of the classification and typical errors.

Figure 2. Prompt for LLM-based Update Classification

ONLY provide a number (1-7) in response.

Determine the category for the following app update text.

If multiple categories seem applicable, always choose the lowest category number (1<2<3<4<5<6<7).

- (1) Novel features – Introducing or enhancing significant functionalities that modify user experience
- (2) Content extensions – Adding or expanding curated or user-facing content
- (3) Platform and device support – Enabling compatibility with new OS versions, development kits, device types, or hardware features
- (4) Specific fixes of bugs – Addressing distinct known issues
- (5) Privacy & Security – Strengthening user data protection, permissions, or security protocols
- (6) General Improvements – Tweaks addressing undisclosed bug fixes, performance improvements, minor changes, or UI/UX polish
- (7) Marketing & Branding – Changes purely about promotional or visual branding assets

Update text: {*release note*}

Feature-based Monetization. We construct variables to track how apps directly monetize the content and features they offer. First, we construct an indicator that takes the value 1 if a price for downloading is charged (*Download Price*) and another indicator that takes the value 1 if at least one in-app purchase element is offered in a given month (*In-App Purchases*).

Monetization Switches. To capture changes in apps’ value-capture strategies, we construct two indicators that identify shifts in the use of in-app purchases. *IAP Entry* equals 1 in the month in which an app offers in-app purchases for the first time after having offered none in the previous month. *IAP Exit* equals 1 in the month in which an app removes all in-app purchase options. Both indicators are 0 in all other months. We do not track switches between paid and free downloads or in ad-based monetization. Changes in download pricing are rare in our sample. Paid status and the presence of advertising-related SDKs further define the comparison groups in our identification strategy, which fixes both at their pre-announcement values.

Other App and Developer Characteristics. We use additional information from our dataset to construct a set of control variables capturing app popularity and developer size. Specifically, we measure app popularity using the monthly number of *Reviews*, the average *Review Rating* and its monthly presence and position in the *Toprank*. To proxy for developer size, we construct the cumulative number of apps released by the developer (*Portfolio Size*), and the cumulative number of distinct categories in which the developer has released apps (*Portfolio Breadth*). Table 1 provides summary statistics for the full sample.

4.4 Matching

A key condition for our quasi-experimental difference-in-differences design is the comparability of ATT affected and unaffected apps. Because treatment exposure, i.e., apps’ reliance on ad-monetization, is not randomly assigned, the groups can differ in observable characteristics. This concern can be particularly relevant in our Top-100 sample as apps that reach top rankings are not a random subset of the App Store. Within this economically relevant subset, ATT affected apps are more likely to be free, scale-oriented apps that attract large user bases and rely on advertising, whereas unaffected apps are more likely to rely on direct monetization. As a result, the groups may already differ before ATT in their popularity, user engagement, and baseline propensity to release updates.

To reduce such pre-ATT heterogeneity, we preprocess the sample using propensity score matching. We estimate each app’s propensity to be ATT affected using a logistic regression model in which the dependent variable indicates ATT exposure. All matching covariates are constructed as averages over the pre-announcement period. Our matching covariate set includes baseline update activity (*Updates*) and the

Table 1. Summary Statistics

Variable	Description	Mean	SD	Min	Median	Max	N
Feature Updates (Count)	Number of feature updates app i releases in month t	.174	.553	0	0	12	721,232
Feature Updates (Binary)	One, if app i releases at least one feature update in month t , else zero	.118	.323	0	0	1	721,232
Maintenance Updates	Number of maintenance updates app i releases in month t	.191	.586	0	0	10	721,232
Updates	Number of all updates app i releases in month t	.367	.835	0	0	13	721,232
Download Price	One, if app i charges an upfront price	.524	.499	0	1	1	721,232
In-App Purchases (IAP)	One, if app i has in-app-purchasing element(s) implemented in month t	.375	.484	0	0	1	721,232
IAP Entry	One, if app i introduces any in-app-purchasing element(s) in month t	.003	.053	0	0	1	721,232
IAP Exit	One, if app i removes all in-app-purchasing elements in month t	.002	.045	0	0	1	721,232
Reviews	Number of reviews app i receives in month t	3.593	27.589	0	0	4,048	721,232
Review Rating	Average rating of the reviews app i receives in month t	3.234	1.271	1	3.375	5	98,268
Toprank	One, if app i is listed in Top-100 ranking in month t	.270	.444	0	0	1	721,232
Toprank Position	Ranking position app i has in month t	54.505	27.622	1	56	100	194,449
Portfolio Size	Number of apps released by developer j up to month t	48.293	168.330	1	11	6,203	687,989
Portfolio Breadth	Number of distinct categories of apps released by developer j up to month t	3.640	3.620	1	2	22	687,989

Notes: Summary statistics are provided on the app-month level for apps that appear at least once in the App Store's weekly Top-100 rankings during our observation period (June 2019 to September 2021). The sample comprises 26,054 apps: 10,668 affected apps, 13,632 paid apps, and 1,754 free apps without advertising-related SDKs in May 2020. Updates without a release note are counted as containing no feature update. Review Rating is observed in months with at least one review, Toprank Position in months in which the app is listed in the top charts. Developer variables are missing for apps without developer-level information.

popularity measures including *Toprank Position*, *Reviews*, and *Review Rating*. We include baseline update activity because an app’s prior update frequency is a strong predictor of its subsequent innovation activity (e.g., Wen and Zhu 2019) and therefore of our main outcome variable. We include the popularity measures because user attention is closely related both to the attractiveness of ad-based monetization and to developers’ incentives and resources to continue feature development (Leyden 2025).

Table 2. Propensity Score Matching and Pre-Announcement Differences of Groups

Matching Variable	Pre-Matching			Post-Matching		
	Mean Affected	Mean Unaffected	SMD	Mean Affected	Mean Unaffected	SMD
Updates	.779	.158	1.023	.377	.329	.101
Reviews	9.888	1.118	.316	7.346	2.461	.160
Review Rating	3.213	3.374	−.137	3.238	3.135	.085
Toprank Position	62.240	63.412	−.050	63.879	63.468	.018

Notes: The pre-matching sample comprises 10,668 affected and 15,386 unaffected apps, the post-matching sample 6,701 affected and 6,701 unaffected apps. Matching covariates are averaged over the pre-announcement period. SMD denotes the standardized mean difference.

After estimating propensity scores, we perform one-to-one nearest-neighbor matching without replacement, so that each unaffected app serves as the match for at most one affected app. Affected apps are matched in descending order of their propensity score with a caliper of .05 to exclude poor matches. For the remaining 3,967 affected apps, no unmatched unaffected app with a sufficiently similar propensity score is available.

These are mainly apps with very high update and review activity, for which comparable apps are scarce. Table 2 reports pre- and post-matching balance. To ensure that results do not depend on the excluded apps, later robustness checks repeat the analysis with a sample matched with replacement that retains all 10,668 affected apps. In this sample, each unaffected app is weighted by the number of affected apps it is matched to.

5. The Effect of ATT on Complementor Innovation

5.1 Econometric Specification

We estimate the effect of the ATT policy on complementor innovation in a difference-in-differences (DiD) framework that compares affected apps with unaffected apps before and after the announcement. Our specification takes the following form:

$$I_{i,t} = \beta_0 + \beta_1 ATT Affected_i \times After ATT_t + X_{i,t} + \eta_i + \theta_t + \epsilon_{i,t}, \quad (1)$$

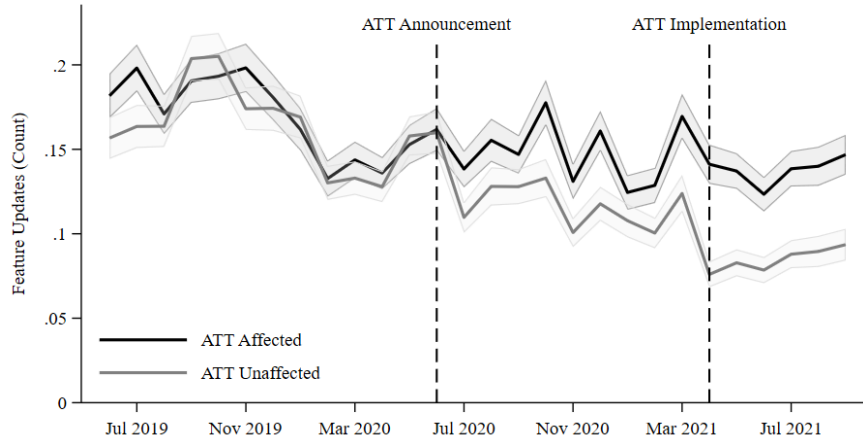
where $I_{i,t}$ denotes *Feature Updates* of app i in month t . $ATT Affected_i$ indicates whether app i had integrated an advertising-related SDK before the announcement. $After ATT_t$ equals 1 for months following the announcement. β_1 captures the differential change in innovation activity of affected apps relative to unaffected apps after the announcement. App fixed effects η_i absorb time-invariant differences between apps. Month fixed effects θ_t absorb shocks common to all apps. $X_{i,t}$ is a vector of time-varying controls.

5.2 Descriptive Trends

Before turning to the estimation, we examine the aggregate development of released feature updates in both groups. Figure 3 plots the monthly average number of feature updates per app for ATT affected (black) and unaffected apps (grey) with 95% confidence intervals. Two descriptive patterns stand out. First, innovation activity declines in both groups during the pre-announcement period. It fluctuates between .15 and .20 in both groups throughout 2019 and drops markedly in early 2020, overlapping with the outbreak of the COVID-19 pandemic. The two series move closely together throughout this period and are virtually identical in June 2020 (first dashed vertical line). Second, the trajectories diverge after the ATT announcement. Feature updates of affected apps stabilize at the level reached in spring 2020 and show repeated pronounced upticks. Unaffected apps show smaller upticks and continue to decline. The gap widens further with the implementation of ATT in April 2021 (second dashed vertical line). The uptick for both groups just before the announcement coincides with Apple’s WWDC timing, when Apple typically presents new developer technologies and tools, which may prompt developers to time releases around the event (Online Appendix D).

Before-after regressions within each group are in line with the descriptive pattern. Relative to the pre-announcement means, feature updates decline by about 15% among affected apps and by about 36% among unaffected apps (Online Appendix E). However, these within-group changes also capture the decline in early 2020 that both groups share and do not isolate the effect of ATT.

Figure 3. *Feature Updates (Count) Trends for Affected and Unaffected Apps*



Notes: Descriptive trends over time are provided for the panel of matched apps ($N = 369,509$). Shaded bands indicate 95% confidence intervals around monthly group means. Apple announced the App Tracking Transparency Policy (ATT) at its annual Worldwide Developer Conference on June 22, 2020. The policy was implemented with the iOS 14.5 update on April 26, 2021.

5.3 Average Effects of the App Tracking Transparency Policy

The difference-in-differences estimates in Table 3 isolate this effect. Pooling affected and unaffected apps, the specification nets out the shared decline through month fixed effects and time-invariant differences between apps through app fixed effects. Across specifications, the interaction *Affected* $\times$ *After ATT* is positive and statistically significant, indicating that affected apps experienced a smaller post-announcement decline (or relative increase) in innovation activity compared to unaffected apps. In the baseline count specification (Column 1), the estimate implies .033 additional feature updates per app-month for affected apps relative to unaffected apps after the announcement. This corresponds to about .40 additional feature updates per app-year ($.033 \times 12$). When adding controls (Column 2), the interaction remains very similar at .035, or about .42 additional feature updates per app-year ($.035 \times 12$). Relative to the pre-announcement mean of feature updates among affected apps (.170), this corresponds to a relative effect of about 20%.

The same conclusion holds when measuring innovation activity at the extensive margin. Using a binary indicator for whether an app releases any feature update in a given month (Column 3), the interaction estimate (.022) implies a 2.2 percentage point higher likelihood of a monthly feature update for affected apps relative to unaffected apps in the post-announcement period. This reflects a relative increase of about 17% over the affected apps' pre-announcement likelihood of a monthly feature update (13%). The

magnitude is in line with complementor reactions to platform governance documented in prior research (Foerderer et al. 2018, Wen and Zhu 2019, Kircher and Foerderer 2024).⁶

We subject these estimates to a set of checks. The results hold in a pseudo-Poisson model that accounts for the count nature of the dependent variable, in a sample restricted to free apps, in alternative matched samples, when accounting for category-specific trends and for potential interference between affected and unaffected apps, and in placebo tests with randomly reassigned treatment status and false announcement date (Online Appendix F).

Table 3. Effect on *Feature Updates*

	DV: Feature Update (Count)		Feature Update (Binary)	
	(1) Baseline OLS	(2) Baseline + Controls OLS	(3) Baseline LPM	(4) Baseline + Controls LPM
Affected × After ATT	.033*** (.004)	.035*** (.004)	.022*** (.003)	.023*** (.003)
Toprank		.027*** (.003)		.019*** (.002)
Reviews		.001*** (.000)		.000*** (.000)
Developer Portfolio Size		−.001** (.000)		−.000 (.000)
Developer Portfolio Breadth		−.006 (.007)		−.005 (.004)
In-App Purchases		−.019+ (.011)		−.005 (.006)
Constant	.133*** (.001)	.177*** (.027)	.101*** (.001)	.119*** (.015)
App Fixed Effects	Yes	Yes	Yes	Yes
Month Fixed Effects	Yes	Yes	Yes	Yes
Clusters	13,402	12,830	13,402	12,830
Observations	356,107	340,865	356,107	340,865
Adj. R ²	.307	.313	.268	.272

Notes: Observations are app-months. The controlled specifications exclude 572 apps with missing developer-level information. The announcement month June 2020 is excluded from the estimation. Clustered standard errors at the app level in parentheses. +, *, **, *** denote significance at the 10%, 5%, 1%, and .1% levels, respectively.

⁶ Prior studies of complementor responses to platform governance in the mobile app context report effects of comparable magnitude. Foerderer et al. (2018) report that Google’s entry into the photography category raised the probability of a major update among affected apps by 9.6 percentage points relative to unaffected apps. Wen and Zhu (2019) find that Google’s entry threat lowered updates on affected apps by 5.1% relative to apps of unaffected developers, and by 7.9% after actual entry. Kircher and Foerderer (2024) document a 16.7% decline in the likelihood of a feature update following Google’s ban on targeted advertising in children’s games.

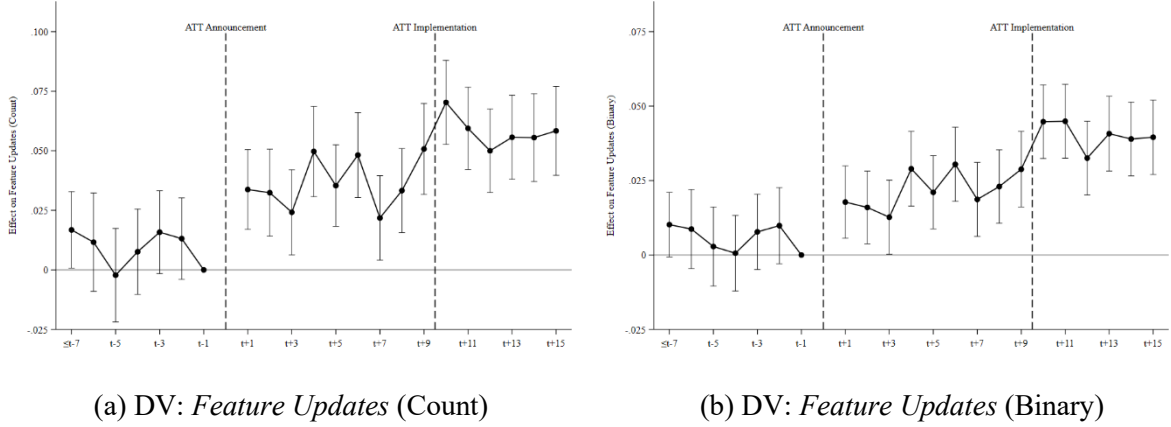
5.4 Dynamic Effects of the App Tracking Transparency Policy

Figure 4 reports event-study estimates of innovation activity over time, distinguishing the intensive margin (Panel a: number of feature updates) and the extensive margin (Panel b: indicator for any feature update in a month). These estimates help assess the parallel-trends assumption underlying the DiD design and show when the effect emerges. Before the announcement, estimates in both panels fluctuate closely around zero without systematic drift and are not jointly statistically significant ($p = .20$ and $p = .32$). These patterns are consistent with affected and unaffected apps following parallel trends before the announcement.

In the post-announcement period, both panels show a divergence that sets in directly after the announcement. Panel (a) shows that the number of feature updates is higher for affected apps relative to unaffected apps from the first post-announcement month and remains higher throughout the announcement phase. Panel (b) shows the same pattern for the likelihood of shipping a feature update. In both panels, the estimates rise again markedly around the implementation of ATT in April 2021 ($t+10$) and remain at a higher level afterwards. Read jointly, the dynamic estimates suggest that the relative post-announcement effect reflects both a higher propensity to ship feature updates and an increase in overall update intensity, with the divergence becoming most pronounced after the implementation. The fact that the divergence strengthens later in the post-period is consistent with responses that materialize with delay due to development cycles. Given the gap between announcement in June 2020 and implementation in April 2021, the pattern is also consistent with responses accumulating as implementation approached.

To formalize this pattern, we replace the post-announcement indicator (Equation 1) with two separate indicators for the period between announcement and implementation (July 2020–March 2021) and the period after implementation (April–September 2021). Relative to the pre-announcement means, the number of feature updates is about 15% higher for affected apps in the first phase and about 27% higher in the second. The likelihood of a feature update is about 11% and 26% higher, respectively. Further details are reported in Online Appendix G.

Figure 4. Effect on *Feature Updates* over Time



Notes: The figure plots coefficients on the interaction of the affected indicator with month indicators, estimated on the matched sample with app and month fixed effects and without controls ($N = 356,107$; clusters = 13,402). The reference month is May 2020 ($t-1$), the month before the ATT announcement. The announcement month June 2020 (t) is excluded from the estimation. Months before December 2019 are pooled into the first coefficient ($\leq t-7$). Bars indicate 95% confidence intervals based on standard errors clustered at the app level.

6. Heterogeneity in Exposure to the App Tracking Transparency Policy

6.1 Category Opt-in Rates

ATT's economic impact is driven by a lack of user consent to tracking, implying that the strength of the shock differs across apps depending on how many of their users consent to tracking (Godinho de Matos and Adjerid 2022, Kraft et al. 2024). When opt-in rates are lower, a larger share of users can no longer be tracked across apps, reducing the data available for targeted advertising. This can make it harder for developers to attract users and reduce the value of advertising displayed within their apps (Goldfarb and Tucker 2011, Rafieian and Yoganarasimhan 2021). The expected loss in advertising effectiveness and incentives to adapt should therefore be greater where fewer users consent. If the documented increase in feature updates reflects a response to ATT, it should therefore be larger where opt-in rates are lower.

To operationalize exposure to ATT, we use category-level tracking opt-in rates observed in later periods (2023–2025) as a proxy for the pressure from low consent rates that apps in different categories face (Adjust 2025). Categories capture meaningful variation in consent behavior as apps within a category share the usage context and data sensitivity that shape users' privacy expectations (Xu et al. 2012). We treat these opt-in rates as informative about ATT exposure because we see little change over time and the relative ordering of app categories by opt-in propensity is stable (Online Appendix H1). Hence, post-2023 opt-in

data still helps distinguish high- vs. low-consent categories that plausibly faced larger expected losses when ATT was announced and later implemented. Importantly, developers could anticipate exposure differences across app categories. Even before the ATT policy, users could restrict tracking through the comparatively hidden ‘Limit Ad Tracking’ device setting, and apps have long needed user permission to access location, notifications, or other device data. In addition, industry benchmarks that report permission opt-in rates and user demographics by app category circulate widely (e.g., Airship 2021). Developers could thus observe how willingly users share data, both for their own apps and relative to other categories, well before the announcement.

Empirically, we implement two complementary tests. First, we estimate a continuous interaction between ATT and the opt-in proxy to assess whether the innovation response varies with exposure intensity. Second, we use a high/low opt-in rate split to provide a robust comparison that is less sensitive to functional-form assumptions.

Table 4. Innovation Response by Category Opt-in Rate

		DV: Feature Update (Count)	
		(1)	(2)
Affected	× After ATT	.034*** (.005)	.054*** (.008)
Category Opt-in Rate	× After ATT	.043 (.043)	
Affected × Category Opt-in Rate	× After ATT	-.285*** (.061)	
High Opt-in Rate	× After ATT		.006 (.007)
Affected × High Opt-in Rate	× After ATT		-.036*** (.010)
Constant		.133*** (.002)	.130*** (.003)
App Fixed Effects		Yes	Yes
Month Fixed Effects		Yes	Yes
Clusters		8,220	8,220
Observations		218,057	218,057
Adj. R ²		.341	.341

Notes: Opt-in rates are available for 13 of the 26 app categories (education, entertainment, finance, food & drink, games, health & fitness, lifestyle, magazines & newspapers, news, shopping, social networking, travel, and utilities), which cover 8,220 of the 13,402 matched apps (61%). See Online Appendix H2 for details on coverage. The Category Opt-in Rate is mean-centered, so that the coefficient on Affected × After ATT in Column (1) gives the effect at the average opt-in rate. High Opt-in Rate indicates categories at or above the median opt-in rate. Observations are app-months. The announcement month June 2020 is excluded from the

estimation. Clustered standard errors at the app level in parentheses. ⁺, *, **, *** denote significance at the 10%, 5%, 1%, and .1% levels, respectively.

Table 4 reports a consistent pattern across both tests. In Column (1), the negative coefficient on *Category Opt-in Rate × Affected × After ATT* indicates that the relative post-announcement increase in feature updates among affected apps is attenuated in higher opt-in categories and stronger in lower opt-in categories where consent frictions and thus the expected shock to targeted advertising are larger. A ten percentage point lower opt-in rate is associated with about .029 additional feature updates per app-month, relative to an effect of .034 at the average opt-in rate. Column (2) corroborates this implication using the high/low opt-in rate splits. Put differently, the estimates align with the exposure prediction that the strongest innovation response should occur where opt-in is low and the incentives to adapt are greatest. Online Appendix H2 complements these estimates with category-level analyses.

6.2 Dependence on Targeted Advertising

A second source of variation concerns the role of advertising in an app’s business model. While reduced advertising effectiveness can make user acquisition more difficult regardless of how an app is monetized, a decline in the value of ads displayed within the app should have different implications across business models. Apps that rely solely on advertising have no alternative source of income within the app to cushion this decline. By contrast, apps that also offer in-app purchases can benefit from user engagement through direct payments that do not primarily depend on tracking (Ghose and Han 2014). This alternative may partly cushion the disruption caused by ATT, even if attracting users becomes more difficult. We therefore expect a stronger innovation response among apps that rely solely on advertising than among apps that combine advertising with in-app purchases.

We test this prediction by comparing the innovation response of apps with and without in-app purchases. Table 5 shows a stronger response among apps whose monetization within the app is limited to advertising. For apps without in-app purchases, the estimated effect is .053 additional feature updates per app-month (Column 1), compared with .033 for apps with in-app purchases (Column 2). Relative to the respective pre-announcement means of affected apps (.143 and .183), these estimates correspond to

increases of about 37% and 18%. Column (3) confirms that the difference between the groups is statistically significant.

Table 5. Innovation Response by In-App Purchase Status

		DV: Feature Update (Count)		
		(1) w/o IAP	(2) IAP	(3) All
Affected	× After ATT	.053*** (.006)	.033*** (.008)	.053*** (.006)
In-App Purchases	× After ATT			-.015* (.007)
Affected × In-App Purchases	× After ATT			-.020* (.010)
Constant		.106*** (.001)	.162*** (.003)	.135*** (.001)
App Fixed Effects		Yes	Yes	Yes
Month Fixed Effects		Yes	Yes	Yes
Clusters		7,425	5,977	13,402
Observations		195,911	160,196	356,107
Adj. R ²		.264	.334	.307

Notes: In-App Purchases indicates apps that offered in-app purchases in May 2020. Columns (1) and (2) compare affected apps with unaffected apps of the same in-app purchase status. Observations are app-months. The announcement month June 2020 is excluded from the estimation. Clustered standard errors at the app level in parentheses. +, *, **, *** denote significance at the 10%, 5%, 1%, and .1% levels, respectively.

6.3 Developer Portfolio

A third source of variation in exposure follows from the design of the ATT policy itself. The policy requires consent for tracking users across apps and websites owned by other companies, while developers can continue to link user data across their own apps without consent.⁷ Developers with multiple apps can therefore draw on data from across their portfolio for analytics, cross-promotion, and targeting, provided that users engage with more than one of their apps. Larger portfolios offer more opportunities for such overlap and may thus reduce developers' reliance on third-party data. Their expected losses from ATT and incentives to adapt should consequently be smaller. If the increase in feature updates reflects a response to ATT, it should therefore be weaker for apps whose developers have larger portfolios.

⁷ Apple's ATT policy distinguishes tracking across properties owned by other providers from analytics across apps owned by the same provider. Apple notes that the ID for Vendors (IDFV) may be used across apps from the same provider without ATT permission, but not for tracking across apps or websites owned by other providers. See <https://developer.apple.com/app-store/user-privacy-and-data-use/> (Retrieved October 14, 2026).

To capture this variation, we use the pre-announcement average of Portfolio Size. Here, this variable serves as a proxy for the potential scope of first-party data, which depends on users' activity across the developer's apps. Developer-level information is available for 12,830 of the 13,402 matched apps, covering 8,906 developers. The median app belongs to a developer with about ten apps, while 13% belong to developers with only one app. Empirically, we first compare apps from single-app developers with those from developers with two or more apps (Table 6, Columns 1–3). We then interact the treatment effect with the logarithm of portfolio size to assess whether the innovation response weakens as portfolio size increases (Table 6, Column 4).

Table 6. Innovation Response by Developer Portfolio

		DV: Feature Update (Count)			
		(1) Single App Dev	(2) Multi App Dev	(3) All Binary Dev	(4) All Cont. Dev
Affected	× After ATT	.086*** (.014)	.023*** (.005)	.086*** (.014)	.030*** (.005)
Multi-App Developer	× After ATT			.036*** (.010)	
Affected × Multi-App Developer	× After ATT			-.063*** (.015)	
Portfolio Size (log)	× After ATT				.014*** (.002)
Affected × Portfolio Size (log)	× After ATT				-.018*** (.003)
Constant		.147*** (.004)	.130*** (.001)	.115*** (.005)	.134*** (.001)
App Fixed Effects		Yes	Yes	Yes	Yes
Month Fixed Effects		Yes	Yes	Yes	Yes
Apps		1,677	11,153	12,830	12,830
Clusters (developers)		1,677	7,229	8,906	8,906
Observations		44,134	296,731	340,865	340,865
Adj. R ²		.279	.317	.311	.312

Notes: Portfolio Size is the number of the developer's apps in May 2020 and is assigned to all apps of the developer. Multi-App Developer indicates a Portfolio Size of two or more. Portfolio Size (log) is mean-centered, so that the coefficient on Affected × After ATT in Column (4) gives the effect at the average portfolio size. The sample excludes 572 apps with missing developer-level information. Observations are app-months. The announcement month June 2020 is excluded from the estimation. Clustered standard errors at the developer level in parentheses. +, *, **, *** denote significance at the 10%, 5%, 1%, and .1% levels, respectively.

Table 6 reports the results. The innovation response is stronger among apps from single-app developers, with an estimated increase of .086 feature updates per app-month (Column 1), compared with

.023 for apps from multi-app developers (Column 2). Column (3) confirms that this difference is statistically significant. This pattern extends beyond the distinction between single-app and multi-app developers. Using portfolio size as a continuous measure, Column (4) shows that doubling the number of apps is associated with a reduction of about .013 feature updates per app-month in the estimated effect. Online Appendix I reports the corresponding estimates by portfolio size group.

These findings are consistent with a weaker innovation response among developers with greater potential access to first-party data. We cannot directly observe whether developers use these data across their apps, and portfolio size may also capture other differences such as their resources. The portfolio results therefore do not establish the first-party data channel directly, but complement the findings on category opt-in rates and advertising dependence in showing a stronger innovation response among apps that we expect to be more exposed to ATT.

7. Mechanism: Adaptation to Reduced Cross-App Tracking

Our main argument is that the increase in innovation reflects app developers' efforts to adapt to the changes brought about by the ATT policy. Two paths of adaptation follow from the role of cross-app tracking in advertising-based monetization and user acquisition. First, tracking enabled developers to earn more from targeted ads within their apps. As ATT reduces these returns, developers may change how they capture value from their users. Developers trade off direct payments against access to user data (Kummer and Schulte 2019). With lower returns from advertising, developers thus have an incentive to shift toward in-app purchases or subscriptions (Kesler 2026). Providing new features or content would then increase app utility that users may be willing to pay for (Appel et al. 2020). Second, advertising-related SDKs could transmit information about users' in-app behavior to advertising networks, helping them direct campaigns toward users likely to take desired actions (Wernerfelt et al. 2025). ATT restricts the cross-app tracking underlying this process when users decline consent, weakening developers' ability to optimize user acquisition. Developers may therefore place greater emphasis on features that retain existing users and attract new ones. Features that encourage continued use can also generate first-party data, while features that facilitate sharing can attract users through existing users' social connections (Aral and Walker 2011).

These changes may partly compensate for the loss of third-party tracking without necessarily changing the app's monetization model.

To examine the proposed shift toward direct payments, we analyze whether affected apps become more likely to introduce in-app purchases. Hazard-model estimates provide no evidence of such an increase after the ATT announcement, either before or after implementation.⁸ If anything, affected apps that already offer in-app purchases are less likely to abandon them in the announcement phase (Online Appendix J1). Developers may instead expand paid offerings in apps that already offer in-app purchases or adjust paid offerings without introducing in-app purchases. We therefore turn to the content of feature updates. Using the release notes of updates already classified as feature updates, we identify the specific features they introduce or extend. For each type of feature, we estimate how the monthly number of such updates changes among affected apps relative to unaffected apps after the announcement. We express each estimate as a percentage of the pre-announcement mean monthly count of the corresponding updates among affected apps. Online Appendix J2-J3 describes the analysis and reports the full results.

Although such updates are relatively uncommon, affected apps that already offer in-app purchases release more feature updates that extend their paid offerings, such as subscriptions, premium features, or purchasable content, than unaffected apps with in-app purchases (45%; monetization, Table J4). Affected apps without in-app purchases also release more updates of this kind than their unaffected counterparts (76%). The pattern is consistent with developers extending existing paid offerings⁹ or introducing new paid options that do not require in-app purchases.

The more detailed analysis of feature updates also allows us to examine whether the increase in innovation reflects efforts to attract and retain users and support the use of first-party data. Updates that make it easier for users to share content or invite others to the app increase by approximately 56% (sharing

⁸ This finding is broadly consistent with Kesler (2026), who compares more than 580,000 apps on the Apple App Store and Google Play Store in a difference-in-differences design and finds only very small increases in the prevalence of paid apps and in-app purchases after the implementation of ATT. The effects are larger among the apps most exposed to ATT but remain negligible in magnitude.

⁹ We cannot directly measure changes in the number of in-app purchase offerings because the list of in-app purchases on an app's App Store page displays at most ten offerings. Once this limit is reached, we cannot observe further increases in the total number of offerings.

and referral, Table J3). Updates related to user accounts and personalization, such as user profiles and recommendations tailored to individual preferences, increase by 23% overall and 39% after implementation (first-party data, Table J3). For features intended to encourage continued use (engagement, Table J3) we find no overall increase. Among apps without in-app purchases these updates increase by 27% (engagement, Table J4). These responses accompany broader improvements to apps' main functions and design, for which feature updates increase by 30% and 41% (core functionality and design and usability, Table J3).

Taken together, these patterns are consistent with developers adapting to reduced access to cross-app tracking through improvements to core app functions and features that expand paid offerings, help attract and retain users, and support the collection and use of first-party data.

8. Robustness Checks

8.1 Extension to all English-Language iOS Apps

Our main analyses focus on apps that appeared in the Top-100 rankings. To examine whether the innovation response extends beyond this group, we re-estimate the same specifications using all English-language iOS apps listed on January 1, 2020. This broader sample comprises 1,294,110 apps (62% of all listed iOS apps), of which 567,062 are classified as affected and 434,294 as unaffected using the same criteria as above. Free apps without SDK information before the announcement are not assigned to either group.

Innovation activity is lower in this broader sample. Before the announcement, affected apps release about one-quarter as many feature updates as those in our main sample (.044 per month), while unaffected apps release .033 per month. To account for differences between the two groups, we match apps before estimation. Because popularity metrics such as reviews, ratings, and rankings are observed only for top-ranked apps, we estimate propensity scores using pre-announcement update activity, in-app purchases, and app age, while otherwise following the original matching procedure (Online Appendix K1). This results in 434,294 matched pairs. Applying this procedure to the apps of our main sample reproduces our main results (Online Appendix K1), so differences between the two samples reflect the population of apps and not the matching.

Table 7 reports the results. Innovation activity declines after the announcement for affected and unaffected apps alike (Columns 1–2). The coefficient on *Affected* $\times$ *After ATT* is positive and statistically significant for the number of feature updates (Column 3) and for the likelihood of a feature update (Column 5). The effect corresponds to an increase of about 8% relative to the pre-announcement mean of affected apps and to about 10% in a pseudo-Poisson model (Column 4). It is thus considerably smaller than among top-ranked apps.

Table 7. Effect on *Feature Updates* in the Full Population of (English) Apps

	DV: Feature Update (Count)				Feature Update (Binary)
	(1) Affected OLS	(2) Unaffected OLS	(3) DiD OLS	(4) DiD PPML	(5) DiD LPM
After ATT	-.012*** (.000)	-.014*** (.000)			
Affected $\times$ After ATT			.002*** (.000)	.095*** (.009)	.002*** (.000)
Constant	.032*** (.000)	.033*** (.000)	.025*** (.000)	-1.599*** (.002)	.021*** (.000)
App Fixed Effects	Yes	Yes	Yes	Yes	Yes
Month Fixed Effects	No	No	Yes	Yes	Yes
Mean (Before ATT)	.032	.033	.032		.027
Clusters	434,294	434,294	868,588	188,119	868,588
Observations	11,609,939	11,599,134	23,209,073	4,987,216	23,209,073
Adj. R ² / Pseudo R ²	.211	.192	.202	.172	.178

Notes: The sample comprises 434,294 affected apps matched one-to-one to unaffected apps on the propensity score (Online Appendix K1). Column (4) reports pseudo-Poisson estimates for apps with at least one feature update during the observation period. Observations are app-months. The announcement month June 2020 is excluded from the estimation. Clustered standard errors at the app level in parentheses. ⁺, *, **, *** denote significance at the 10%, 5%, 1%, and .1% levels, respectively.

Although the “long tail” of niche products in digital markets such as app markets can be relevant for users (Brynjolfsson et al. 2010, 2011), commercially meaningful activity is often concentrated in a much smaller subset of apps, specifically in those that appear and persist in top charts (Kapoor and Agarwal 2017). Many apps outside this subset likely attract limited attention and receive little post-launch investment or updating. In such a setting, a platform policy shock that operates through advertising and user acquisition may translate into clearer behavioral responses among high-attention, actively maintained apps than among the broader pool of long-tail apps. The smaller effect in the full population is consistent with this view. At the same time, the estimate for the full population averages over a heterogeneous set of apps. Economically

relevant apps outside the Top-100 may have responded in similar ways to those in our focal sample, while the long tail of low-activity apps can dilute the effect in aggregate. Online Appendix K2–K3 reports further specifications, the comparison without matching, and an event study for the full population.

8.2 Measurement Sensitivity

Our analysis relies on a novel measure of innovation in app updates that follows the framework of Nowak et al. (2025). We discuss how this approach builds on earlier work to address the challenges of identifying innovation from app platform data. Two main sources of platform information can be used to detect changes in software applications.

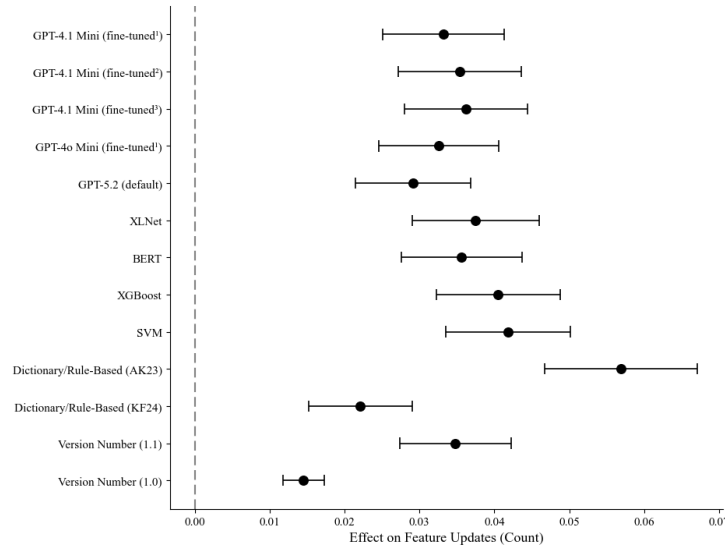
The first source is an app update’s version number (e.g., Boudreau 2012, Wen and Zhu 2019). In principle, versioning conventions could provide signals about the scale or nature of an update. However, in practice, there is substantial heterogeneity in how version numbers are assigned in mobile app development. Interviews with professional app developers indicate that versioning logic typically varies across teams, firms, and products, often reflecting internal development cycles and product strategies instead of versioning conventions. Respondents also noted strategic reasons to deviate from formal conventions, since Apple’s review process may subject major version changes (e.g., version 1.5 to version 2.0) to closer scrutiny than minor updates (e.g., version 1.5 to version 1.6).

The second source is release notes (e.g., Foerderer et al. 2018, Agarwal et al. 2023, Kircher and Foerderer 2024). These are informative because, during our observation period, developers were already required to report what actually had changed in an update. Release notes therefore often contain more direct information about the actual content of updates and, in principle, should better reflect innovative activity. At the same time, they are often terse, ambiguous, or strategically worded, and reporting practices vary across developers (Yang et al. 2022). Earlier natural language processing approaches such as keyword-based approaches thus may reach their methodological limit. Against this background, the LLM-based measure offers an important advantage. It is better suited to interpreting heterogeneous and imperfect textual descriptions and therefore promises to reduce measurement error relative to earlier approaches. While the

measure remains imperfect, it offers a promising way to infer innovative activity from the limited information observable on app platforms.

This measurement challenge makes it important to examine whether our main results are robust to alternative ways of capturing update activity and researcher design decisions. We therefore re-estimate our main specifications using a range of proxies derived from earlier work and design variations. Specifically, we compare our baseline LLM-based measure to version number heuristics (Wen and Zhu 2019), prior dictionary- and rule-based approaches (Agarwal and Kapoor 2023, Kircher and Foerderer 2024), supervised machine-learning classifiers, and variations in training data and model choice of LLMs. Online Appendix L1 provides details on each measure.

Figure 5. Sensitivity Analysis of LLM-based Innovation Variable



Notes: The figure reports coefficient estimates from DiD specifications (Affected $\times$ After ATT) estimated on the panel of matched apps ($N = 356,107$; clusters = 13,402). The announcement month June 2020 is excluded. Horizontal whiskers denote 95% confidence intervals. Fine-tuned models were trained on a human-labeled training set: ¹ $n = 2,000$ with representative class distribution (main variable), ² $n = 2,000$ with balanced class distribution, and ³ $n = 1,000$ with representative class distribution. For all LLM-based classifications, temperature was set to 0 and the seed to 94032. The dictionary/rule-based classifiers replicate the methods proposed by (AK23) Agarwal and Kapoor (2023) and (KF24) Kircher and Foerderer (2024). Further details are provided in Online Appendix L1.

As shown in Figure 5, the estimated effect on complementor innovation activity is positive and statistically significant across all measurement approaches. The LLM-based variants have similar estimates across training data and model choice. The supervised machine-learning classifiers have estimates in the same range or slightly above. The estimates of the earlier approaches are more dispersed. This is expected,

as classifications based on predefined terms or version numbers are more prone to measurement error and differ in how many updates they classify as innovative (Online Appendix L1). Version-number measures are particularly restrictive. Major version changes identify only a small subset of updates and therefore likely capture relatively clear-cut cases while missing much innovative activity. The fact that the estimated effect remains positive across all alternatives suggests that our main finding is not driven by the specific operationalization of innovation. Online Appendix L2 further shows that classification errors do not differ before and after the announcement and would, if anything, reduce the estimated effect.

9. Discussion

In this study, we empirically examined how a platform governance change that constrains complementors' value capture affects their innovation behavior in the context of Apple's App Tracking Transparency policy. We use Apple's announcement of the policy as a quasi-exogenous shock and show that apps exposed to the policy increased their innovation activity relative to unaffected apps. After the announcement, affected apps sustained their innovation activity, with renewed increases around the implementation of the policy, while unaffected apps continued to decline. This corresponds to a relative increase of about 20% in the number of feature updates for affected apps. Further analyses suggest that this increase in innovation is best understood as a form of adaptive innovation. ATT reduced the returns to cross-app tracking and thereby weakened an important basis of value-capture for many apps. The innovation response is stronger among more exposed apps: in app categories with lower tracking opt-in rates, among apps whose in-app monetization relies solely on advertising, and among developers with small app portfolios. This adaptation appears to have primarily taken place within developers' existing monetization models. Their additional feature updates include improvements to core app functions and design, extensions of paid offerings, and features that help attract and retain users and support the collection and use of first-party data.

A supplementary analysis of all English-language iOS apps shows that the adaptive response is not confined to apps that appear in Top-100 rankings. The relative post-announcement effect is positive in the full population as well, at about 8%, compared with about 20% among apps that appear in the Top-100 rankings. Part of this gap may reflect the long tail of low-activity and hobbyist apps, where innovation is

limited to begin with and responses are difficult to detect in the aggregate. Part may also reflect a weaker response outside the top ranks. This pattern suggests that privacy governance stimulates innovation most among complementors with sufficient demand, visibility, and organizational capacity to respond. In this sense, privacy-oriented platform governance may not simply reduce or preserve ecosystem dynamism; it may also redistribute opportunities toward already stronger complementors.

Our study makes several contributions to the literature on digital platform ecosystems and the economics of privacy. First, we extend research that examines how platform operators use platform governance to shape complementors' incentives and innovation activities (Tiwana et al. 2010, Wareham et al. 2014, Foerderer et al. 2018, Rietveld et al. 2019, 2021, Agarwal et al. 2023). We show that implementing policies on stricter data privacy can act as one such instrument and, in doing so, can help sustain or even enhance the value created on the platform. In this sense, privacy governance can redirect distinct groups of complementors toward new directions of innovation without losing them as contributors to the platform.

Our findings also contribute to the literature on the economic consequences of privacy governance. Existing work has often emphasized the negative effects of privacy restrictions (Goldfarb and Tucker 2012, Johnson et al. 2020, Janßen et al. 2022, Aridor et al. 2025). In related work on mobile platforms, Kircher and Foerderer (2024) find that Google's ban on targeted advertising in children's apps reduced complementor innovation. Our results provide nuance to this picture. A key difference is that the ATT policy does not constitute a categorical ban but introduced a user consent-based restriction that impaired targeted advertising but did not eliminate it completely. This distinction appears consequential. When a governance change fully forecloses a value-capture logic, complementors scale back innovation because expected returns diminish sharply and the economic basis for continued development erodes (Kircher and Foerderer 2024). In contrast, when monetization is weakened but still partially recoverable, complementors may respond by adapting their products within their existing monetization strategies in an effort to restore value capture potential. Our results therefore suggest that the consequences of privacy governance for complementor innovation depend on the form the restriction takes. Consent-based privacy governance may induce adaptive innovation even when it imposes substantial economic costs on complementors.

The study further contributes to the growing empirical literature on ATT and the timing of adaptive reactions. Existing studies have documented declines in advertising effectiveness and publisher revenue, shifts in monetization, and adjustment in app components following the implementation of ATT (Kraft et al. 2024, Cheyre et al. 2026, Kesler 2026). We add to this literature in two ways. First, our results show that developers' responses began already at the policy announcement stage. This suggests that ATT operated not only through its later technical implementation, but also as an information shock that triggered anticipatory adjustment. In platform ecosystems, announcements of future governance changes can therefore matter in their own right because complementors start adapting before enforcement takes effect (Wen and Zhu 2019). Second, our study demonstrates the importance of identifying heterogeneous exposure within the iOS ecosystem (Mayya and Viswanathan 2025). ATT is often discussed as if it affected iOS apps uniformly. Our results challenge this view. By distinguishing between apps with and without ad-related SDKs, we show that the innovation response to ATT differs substantially across complementors depending on their underlying value-capture model.

Our findings also speak to managerial practice. For platform operators, they suggest that privacy-enhancing governance need not uniformly suppress complementor innovation. Compared with outright bans, moderated interventions such as consent-based restrictions may preserve room for adaptation and thereby mitigate some of the more severe innovation losses that more categorical restrictions might induce. Our findings also suggest that the timing of governance changes matters. Because complementors already adjusted after Apple's announcement, platform operators may be able to reduce disruption by communicating major governance changes early and credibly, particularly when these changes affect complementors' value-capture opportunities and when immediate implementation would otherwise cut off revenue streams before complementors have time to adapt.

This study is subject to limitations that provide opportunities for future research. First, our measure of innovation captures observable feature and content updates reliably, but it cannot directly assess the degree of novelty, quality, or downstream user value of these efforts. Second, although our analyses provide evidence consistent with adaptation and exposure heterogeneity, we cannot directly observe all internal

strategic responses by developers, such as investments in first-party data capabilities or cross-platform reallocations (Mayya and Zhang 2025). Finally, our research design is aimed at capturing immediate and mid-term effects after the announcement with coverage of the technical implementation. However, it does not allow us to assess long-run effects.

References

- Acquisti A, Taylor C, Wagman L (2016) The Economics of Privacy. *Journal of Economic Literature* 54(2):442–492.
- Adjerid I, Acquisti A, Telang R, Padman R, Adler-Milstein J (2016) The Impact of Privacy Regulation and Technology Incentives: The Case of Health Information Exchanges. *Management Science* 62(4):1042–1063.
- Adjust (2025) ATT opt-in rates: The latest benchmarks by app category and country. Retrieved (October 14, 2026), <https://www.adjust.com/blog/att-opt-in-rates-2025/>.
- Agarwal S, Kapoor R (2023) Value Creation Tradeoff in Business Ecosystems: Leveraging Complementarities While Managing Interdependencies. *Organization Science* 34(3):1216–1242.
- Agarwal S, Miller CD, Ganco M (2023) Growing platforms within platforms: How platforms manage the adoption of complementor products in the presence of network effects? *Strategic Management Journal* 44(8):1879–1910.
- Airship (2021) Push Notifications & Mobile Engagement: 2021 Benchmarks. Retrieved (October 14, 2026) <https://www.airship.com/resources/benchmark-report/push-notifications-mobile-engagement-2021-benchmarks/>.
- Anderson EG, Parker GG, Tan B (2014) Platform Performance Investment in the Presence of Network Externalities. *Information Systems Research* 25(1):152–172.
- Appel G, Libai B, Muller E, Shachar R (2020) On the monetization of mobile apps. *International Journal of Research in Marketing* 37(1):93–107.
- Apple (2020) Details for app privacy questions now available. Retrieved (October 14, 2026), <https://developer.apple.com/news/?id=hx9s63c5>.
- Aral S, Walker D (2011) Creating Social Contagion Through Viral Product Design: A Randomized Trial of Peer Influence in Networks. *Management Science* 57(9):1623–1639.
- Aridor G, Che YK, Hollenbeck B, Kaiser M, McCarthy D (2025) Evaluating the Impact of Privacy Regulation on E-Commerce Firms: Evidence from Apple’s App Tracking Transparency. *Management Science*:mns.2024.06600.
- Bhargava HK (2022) The Creator Economy: Managing Ecosystem Supply, Revenue Sharing, and Platform Design. *Management Science* 68(7):5233–5251.
- Boudreau KJ (2010) Open Platform Strategies and Innovation: Granting Access vs. Devolving Control. *Management Science* 56(10):1849–1872.
- Boudreau KJ (2012) Let a Thousand Flowers Bloom? An Early Look at Large Numbers of Software App Developers and Patterns of Innovation. *Organization Science* 23(5):1409–1427.
- Brynjolfsson E, Hu Y (Jeffrey), Simester D (2011) Goodbye Pareto Principle, Hello Long Tail: The Effect of Search Costs on the Concentration of Product Sales. *Management Science* 57(8):1373–1386.
- Brynjolfsson E, Hu Y (Jeffrey), Smith MD (2010) Research Commentary—Long Tails vs. Superstars: The Effect of Information Technology on Product Variety and Sales Concentration Patterns. *Information Systems Research* 21(4):736–747.
- Burtch G, Ghose A, Wattal S (2015) The Hidden Cost of Accommodating Crowdfunder Privacy Preferences: A Randomized Field Experiment. *Management Science* 61(5):949–962.
- Campbell J, Goldfarb A, Tucker C (2015) Privacy Regulation and Market Structure. *Journal of Economics & Management Strategy* 24(1):47–73.

- Casadesus-Masanell R, Hałaburda H (2014) When Does a Platform Create Value by Limiting Choice? *Journal of Economics & Management Strategy* 23(2):259–293.
- Casadesus-Masanell R, Hervas-Drane A (2015) Competing with Privacy. *Management Science* 61(1):229–246.
- Ceccagnoli M, Forman C, Huang P, Wu DJ (2012) Cocreation of Value in a Platform Ecosystem: The Case of Enterprise Software. *MIS Quarterly* 36(1):263–290.
- Cheyre C, Leyden BT, Baviskar S, Acquisti A (2026) The Impact of Apple’s App Tracking Transparency Framework on the App Ecosystem. https://papers.ssrn.com/abstract_id=4453463.
- Claussen J, Kretschmer T, Mayrhofer P (2013) The Effects of Rewarding User Engagement: The Case of Facebook Apps. *Information Systems Research* 24(1):186–200.
- Comino S, Manenti FM, Mariuzzo F (2019) Updates management in mobile applications: iTunes versus Google Play. *Journal of Economics & Management Strategy* 28(3):392–419.
- Conti A, Santaló J (2026) The Hidden Costs of Fairness in Platform Markets: The Dynamics of Lowering Developer Royalty Rates. *Management Science*:mns.2024.05861.
- Curchod C, Patriotta G, Cohen L, Neysen N (2020) Working for an Algorithm: Power Asymmetries and Agency in Online Work Settings. *Administrative Science Quarterly* 65(3):644–676.
- Cutolo D, Kenney M (2021) Platform-Dependent Entrepreneurs: Power Asymmetries, Risks, and Strategies in the Platform Economy. *Academy of Management Perspectives* 35(4):584–605.
- Foerderer J (2020) Interfirm Exchange and Innovation in Platform Ecosystems: Evidence from Apple’s Worldwide Developers Conference. *Management Science* 66(10):4772–4787.
- Foerderer J, Kude T, Mithas S, Heinzl A (2018) Does Platform Owner’s Entry Crowd Out Innovation? Evidence from Google Photos. *Information Systems Research* 29(2):444–460.
- Foerderer J, Lueker N, Heinzl A (2021) And the Winner Is ...? The Desirable and Undesirable Effects of Platform Awards. *Information Systems Research* 32(4):1155–1172.
- Gao Y, Gu M, Liu D (2025) Behind Privacy Labels: Data Tracking and Advertising Competition on App Platforms. *MIS Quarterly* 49(4):1595–1626.
- Garg R, Telang R (2013) Inferring App Demand from Publicly Available Data. *MIS Quarterly* 37(4):1253–1264.
- Gawer A, Henderson R (2007) Platform Owner Entry and Innovation in Complementary Markets: Evidence from Intel. *Journal of Economics & Management Strategy* 16(1):1–34.
- Ghazawneh A, Henfridsson O (2013) Balancing platform control and external contribution in third-party development: the boundary resources model. *Information Systems Journal* 23(2):173–192.
- Ghose A, Han SP (2014) Estimating Demand for Mobile Applications in the New Economy. *Management Science* 60(6):1470–1488.
- Godinho de Matos M, Adjerid I (2022) Consumer Consent and Firm Targeting After GDPR: The Case of a Large Telecom Provider. *Management Science* 68(5):3330–3378.
- Goldberg SG, Johnson GA, Shriver SK (2024) Regulating Privacy Online: An Economic Evaluation of the GDPR. *American Economic Journal: Economic Policy* 16(1):325–358.
- Goldfarb A, Que VF (2023) The Economics of Digital Privacy. *Annual Review of Economics* 15:267–286.
- Goldfarb A, Tucker C (2011) Online Display Advertising: Targeting and Obtrusiveness. *Marketing Science* 30(3):389–404.
- Goldfarb A, Tucker C (2012) Shifts in Privacy Concerns. *The American Economic Review* 102(3):349–353.
- Gopal RD, Hidaji H, Patterson RA, Rolland E, Zhdanov D (2018) How Much to Share with Third Parties? User Privacy Concerns and Website Dilemmas. *MIS Quarterly* 42(1):143–164.
- Hagiu A (2006) Pricing and commitment by two-sided platforms. *The RAND Journal of Economics* 37(3):720–737.
- Huang P, Ceccagnoli M, Forman C, Wu DJ (2013) Appropriability Mechanisms and the Platform Partnership Decision: Evidence from Enterprise Software. *Management Science* 59(1):102–121.
- Janßen R, Kesler R, Kummer M, Waldfogel J (2022) GDPR and the Lost Generation of Innovative Apps. *NBER Working Paper No. 30028*, National Bureau of Economic Research, Cambridge, MA.

- Jia J, Jin GZ, Wagman L (2021) The Short-Run Effects of the General Data Protection Regulation on Technology Venture Investment. *Marketing Science* 40(4):661–684.
- Johnson GA, Shriver SK, Du S (2020) Consumer Privacy Choice in Online Advertising: Who Optes Out and at What Cost to Industry? *Marketing Science* 39(1):33–51.
- Johnson GA, Shriver SK, Goldberg SG (2023) Privacy and Market Concentration: Intended and Unintended Consequences of the GDPR. *Management Science* 69(10):5695–5721.
- Kapoor R, Agarwal S (2017) Sustaining Superior Performance in Business Ecosystems: Evidence from Application Software Developers in the iOS and Android Smartphone Ecosystems. *Organization Science* 28(3):531–551.
- Kesler R (2026) The Impact of Apple's App Tracking Transparency on App Monetization. <https://papers.ssrn.com/abstract=4090786>.
- Kircher T, Foerderer J (2024) Ban Targeted Advertising? An Empirical Investigation of the Consequences for App Development. *Management Science* 70(2):1070–1092.
- Kraft L, Skiera B, Koschella T (2024) Economic Impact of Opt-in versus Opt-out Requirements for Personal Data Usage: The Case of Apple's App Tracking Transparency (ATT). https://papers.ssrn.com/abstract_id=4598472.
- Kummer M, Schulte P (2019) When Private Information Settles the Bill: Money and Privacy in Google's Market for Smartphone Applications. *Management Science* 65(8):3470–3494.
- Lee DJ, Ahn JH, Bang Y (2011) Managing Consumer Privacy Concerns in Personalization: A Strategic Analysis of Privacy Protection. *MIS Quarterly* 35(2):423–444.
- Lee G, Raghu TS (2014) Determinants of Mobile Apps' Success: Evidence from the App Store Market. *Journal of Management Information Systems* 31(2):133–170.
- Leyden BT (2025) Platform design and innovation incentives: Evidence from the product rating system on Apple's App Store. *International Journal of Industrial Organization* 99:103133.
- Mayya R, Viswanathan S (2025) Delaying Informed Consent: An Empirical Investigation of Mobile Apps' Upgrade Decisions. *Management Science* 71(8):7113–7135.
- Mayya R, Zhang J (2025) Privacy Spillovers across Competing Platforms. *Working paper*, New York University Stern School of Business, New York.
- Miller AR, Tucker C (2018) Privacy Protection, Personalized Medicine, and Genetic Testing. *Management Science* 64(10):4648–4668.
- Miric M, Jeppesen LB (2020) Does piracy lead to product abandonment or stimulate new product development? Evidence from mobile platform-based developer firms. *Strategic Management Journal* 41(12):2155–2184.
- Miric M, Ozalp H, Yilmaz ED (2023) Trade-offs to using standardized tools: Innovation enablers or creativity constraints? *Strategic Management Journal* 44(4):909–942.
- Moehlmann M, Zalmanson L, Henfridsson O, Gregory RW (2021) Algorithmic Management of Work on Online Labor Platforms: When Matching Meets Control. *MIS Quarterly* 45(4):1999–2022.
- Nowak R, Figge P, Haeussler C (2025) AI-Based Measurement of Innovation: Mapping Expert Insight into Large Language Model Applications. <https://arxiv.org/abs/2508.02430>.
- Parker GG, Van Alstyne MW (2018) Innovation, Openness, and Platform Control. *Management Science* 64(7):3015–3032.
- Parker GG, Van Alstyne MW (2005) Two-Sided Network Effects: A Theory of Information Product Design. *Management Science* 51(10):1494–1504.
- Peukert C, Bechtold S, Batikas M, Kretschmer T (2022) Regulatory Spillovers and Data Governance: Evidence from the GDPR. *Marketing Science* 41(4):746–768.
- Rafieian O, Yoganarasimhan H (2021) Targeting and Privacy in Mobile Advertising. *Marketing Science* 40(2):193–218.
- Rahman HA, Karunakaran A, Cameron LD (2024) Taming Platform Power: Taking Accountability into Account in the Management of Platforms. *Academy of Management Annals* 18(1):251–294.

- Rietveld J, Ploog JN, Nieborg DB (2020) Coevolution of Platform Dominance and Governance Strategies: Effects on Complementor Performance Outcomes. *Academy of Management Discoveries* 6(3):488–513.
- Rietveld J, Schilling MA (2021) Platform Competition: A Systematic and Interdisciplinary Review of the Literature. *Journal of Management* 47(6):1528–1563.
- Rietveld J, Schilling MA, Bellavitis C (2019) Platform Strategy: Managing Ecosystem Value Through Selective Promotion of Complements. *Organization Science* 30(6):1232–1251.
- Rietveld J, Seamans R, Meggiorin K (2021) Market Orchestrators: The Effects of Certification on Platforms and Their Complementors. *Strategy Science* 6(3):244–264.
- Sharma P, Sun Y, Wagman L (2021) The Differential Effects of New Privacy Protections on Publisher and Advertiser Profitability. https://papers.ssrn.com/abstract_id=3503065.
- Shi R, Aaltonen A, Henfridsson O, Gopal R (2023) Comparing Platform Owners' Early and Late Entry into Complementary Markets. *MIS Quarterly* 47(4):1727–1744.
- Sokol DD, Zhu F (2021) Harming Competition and Consumers under the Guise of Protecting Privacy: An Analysis of Apple's iOS 14 Policy Updates. *Cornell Law Review Online* 107(3):101–127.
- Sun T, Yuan Z, Li C, Zhang K, Xu J (2024) The Value of Personal Data in Internet Commerce: A High-Stakes Field Experiment on Data Regulation Policy. *Management Science* 70(4):2645–2660.
- Tilson D, Lyytinen K, Sørensen C (2010) Research Commentary—Digital Infrastructures: The Missing IS Research Agenda. *Information Systems Research* 21(4):748–759.
- Tiwana A (2015) Evolutionary Competition in Platform Ecosystems. *Information Systems Research* 26(2):266–281.
- Tiwana A, Konsynski B, Bush AA (2010) Platform Evolution: Coevolution of Platform Architecture, Governance, and Environmental Dynamics. *Information Systems Research* 21(4):675–687.
- Tsai JY, Egelman S, Cranor L, Acquisti A (2011) The Effect of Online Privacy Information on Purchasing Behavior: An Experimental Study. *Information Systems Research* 22(2):254–268.
- Wareham J, Fox PB, Cano Giner JLC (2014) Technology Ecosystem Governance. *Organization Science* 25(4):1195–1215.
- Wen W, Zhu F (2019) Threat of platform-owner entry and complementor responses: Evidence from the mobile app market. *Strategic Management Journal* 40(9):1336–1367.
- Wernerfelt N, Tuchman A, Shapiro BT, Moakler R (2025) Estimating the Value of Offsite Tracking Data to Advertisers: Evidence from Meta. *Marketing Science* 44(2):268–286.
- Xu H, Teo HH, Tan BCY, Agarwal R (2012) Research Note—Effects of Individual Self-Protection, Industry Self-Regulation, and Government Regulation on Privacy Concerns: A Study of Location-Based Services. *Information Systems Research* 23(4):1342–1363.
- Yang AZH, Hassan S, Zou Y, Hassan AE (2022) An empirical study on release notes patterns of popular apps in the Google Play Store. *Empirical Software Engineering* 27(2):55.
- Yin PL, Davis JP, Muzyrya Y (2014) Entrepreneurial Innovation: Killer Apps in the iPhone Ecosystem. *The American Economic Review* 104(5):255–259.
- Yoo Y, Henfridsson O, Lyytinen K (2010) Research Commentary—The New Organizing Logic of Digital Innovation: An Agenda for Information Systems Research. *Information Systems Research* 21(4):724–735.
- Zhang Y, Li J, Tong TW (2022) Platform governance matters: How platform gatekeeping affects knowledge sharing among complementors. *Strategic Management Journal* 43(3):599–626.
- Zhu F, Liu Q (2018) Competing with complementors: An empirical look at Amazon.com. *Strategic Management Journal* 39(10):2618–2642.

Online Appendix

Opt-in to Innovation? The Effects of Apple's App Tracking Transparency Policy on Complementor Innovation

Appendix A: Industry Reaction and Enforcement of the ATT Policy	1
Appendix B: Identification of Advertising-Related SDKs.....	4
Appendix C: Release Note Classification	6
Appendix D: App Development Seasonality	8
Appendix E: Before-After Estimations	9
Appendix F: Sensitivity of the Main Estimates	11
Appendix G: Timing of the Innovation Response.....	20
Appendix H: App Category Opt-in Rates	21
Appendix I: Developer Portfolio Groups	23
Appendix J: Mechanism Analyses	24
Appendix K: Extension to All English-Language iOS Apps	32
Appendix L: Sensitivity of the LLM-based Innovation Variable.....	36
References	39

Appendix A: Industry Reaction and Enforcement of the ATT Policy

A1 Opposition to the ATT Policy

Figure A1 shows a full-page newspaper advertisement that Facebook (now Meta) placed in major US newspapers in December 2020 as part of a public campaign opposing the ATT policy. In an accompanying statement, Facebook argued that the policy would severely reduce the effectiveness of personalized advertising for app developers and small businesses and that it would “force businesses to turn to subscriptions and other in-app payments” as alternative sources of revenue (Levy 2020). The campaign illustrates that affected complementors understood the scope of the policy well before its implementation and anticipated pressure to adjust their value capture models.

Figure A1. Facebook Newspaper Advertisement Opposing the ATT Policy



Notes: Full-page advertisement placed by Facebook (now Meta) in major US newspapers (December 16, 2020).

A2 ATT Enforcement and Compliance

A key assumption underlying our empirical strategy is that Apple’s enforcement of its ATT policy is robust and effective. Specifically, the credibility of our analysis rests on three conditions: (i) Apple consistently enforces its governance policies; (ii) the ATT technical framework effectively prevents the unauthorized

collection of the Identifier for Advertisers (IDFA) and app developers comply with the requirements of the ATT policy; and (iii) the policy applies to a substantial share of the relevant user base.

First, Apple's iOS is a closed platform-based ecosystem with a long-standing reputation for rigorous review processes of new app releases and updates to already listed apps. Violations of Apple's policies are often met with strict penalties, which can even extend to the removal of apps from the App Store and the termination of developer accounts. A prominent example is the removal of Epic Games' Fortnite app and the termination of their developer account on the App Store, following Epic Games' attempt to circumvent Apple's payment system.¹ The credibility of this enforcement is also reflected in the behavior of affected complementors. Facebook (now Meta), the policy's most vocal opponent, publicly stated that it had "no choice but to show" Apple's consent prompt, as noncompliance would risk retaliation from Apple (Levy 2020).

Second, empirical evidence indicates that the ATT policy is effective in practice. Kollnig et al. (2022) demonstrate that, when correctly implemented at the technical level, the ATT framework effectively prevents the collection of the IDFA without user consent. Although some loopholes may still exist, Kesler et al. (2026) find that 87.4% of the iOS apps they examined did not collect any device IDs for advertising without user authorization.

Third, the consent requirement applies only to users running an iOS version that supports the ATT framework and whose data can be tracked. Complementors could anticipate that the policy would reach most users quickly. The ATT policy was announced as part of the new major iOS version, iOS 14. Major iOS versions have historically been adopted rapidly, and iOS 14 reached about 50% of users within six weeks of its release (Clover 2020). Moreover, most users could actually be tracked before the policy. Among iOS users, 93% are adults and in principle trackable. Under the preceding opt-out setting, 79% of these adult

¹ See: <https://www.theverge.com/2020/8/13/21366438/apple-fortnite-ios-app-store-violations-epic-payments> or <https://www.theverge.com/2020/8/28/21406013/apple-epic-games-fortnite-developer-account-terminated-no-longer-available> (Retrieved October 14, 2026).

users had not restricted tracking and were thus trackable by default (Kraft et al. 2024). Under the ATT policy, these users receive the consent prompt.

Taken together, complementors could expect the ATT policy to be enforced, effective, and binding for a substantial share of their users.

Appendix B: Identification of Advertising-Related SDKs

Table B1 documents the classification of SDKs with advertising functionality that underlies our treatment identification variable *ATT Affected*. SDKs qualify as advertising-related if MixRank’s functional tagging assigns them a keyword related to “advertising / ad”. As such, SDKs whose functionality is not primarily directed at advertising and which likely do not directly rely on the IDFA (e.g., general-purpose analytics SDK) are not classified as advertising-related. The table groups the resulting 1,736 unique SDKs by their functionality category.

Table B1. Advertising-Related SDKs by Functionality Category

Category	SDKs	Typical Tags	Description	Top Examples
Ad networks & exchanges	1,538	advertising; monetization; native-ads; video; targeting	Serve advertisements within apps and broker ad inventory between publishers and advertisers.	AdMob, MoPub, Meta Audience Network, AppLovin
Ad-mediation & bidding platforms	556	advertising; mediation	Manage the allocation of an app’s ad inventory across multiple ad networks; adapters connect a mediation layer to a specific network.	ironSource, Fyber, OpenMediation adapters (AdMob, Unity)
Rewarded & offer-wall networks	34	advertising; offer-wall; incentives; rewards	Serve ad formats in which users receive in-app rewards for completing actions such as viewing videos or installing apps.	Tapjoy, PlayHaven, Fyber Offer Wall
Mobile measurement & attribution	35	advertising; tracking; analytics; deep-linking	Record app installs and in-app conversions and assign them to preceding advertising campaigns.	AppsFlyer, Adjust, Branch Metrics, Google Conversion Tracking
Ad-enabled analytics, audience measurement & retargeting	91	advertising; analytics; dmp; targeting; retargeting	Aggregate user-level data across apps into profiles used for audience segmentation, campaign measurement, and re-addressing previous users.	Flurry, Adobe Experience Cloud, Quantcast
Ad-enabled engagement & marketing platforms	13	advertising; marketing; push-notifications; retargeting	Provide messaging, push notification, and campaign management functions, including advertising and retargeting modules.	Urban Airship, Pushwoosh, Localytics, Swrve
Location-based advertising	28	advertising; location; geofencing; beacon	Serve or target advertisements based on device location data.	Factual, GroundTruth (xAd), Verve

OS-level advertising frameworks	2	advertising; tracking	Apple frameworks that expose the advertising identifier to apps (AdSupport) or provided native ad serving (iAd, discontinued).	AdSupport, iAd Framework (discontinued)
---------------------------------	---	-----------------------	--	---

Notes: Categories reflect SDK functionality and are not mutually exclusive; multi-function SDKs (e.g., MoPub, AppLovin) appear in each applicable category. Counts therefore sum to more than the 1,736 distinct treatment-defining SDKs.

Appendix C: Release Note Classification

Table C1 illustrates our classification of release notes with representative examples. The examples cover the most prevalent classification and error patterns. For each release note, we report the human-coded true label, the classification of a fine-tuned LLM (Nowak et al. 2025), and two alternative approaches used in prior work. The dictionary-based approach classifies an update as innovative if its release note contains at least one of the terms “new”, “added”, “upgrade”, or “major” (Kircher and Foerderer 2024). The versioning-based approach classifies an update as innovative when the first digit of the version number changes relative to the previous release, for example from 1.4 to 2.0 (Wen and Zhu 2019). The examples illustrate the limitations of both alternatives. The dictionary-based approach misclassifies maintenance updates that announce fixes with feature-related vocabulary and misses feature updates described without such vocabulary. The versioning-based approach misclassifies major version jumps that bundle routine maintenance and misses feature updates released as minor versions. The LLM classification assigns updates based on the described content of the release. In the benchmark of Nowak et al. (2025), fine-tuned LLMs recover human-coded innovation judgments more accurately than dictionary-, versioning-, and classical machine-learning-based approaches. Yet, the LLM applied here also occasionally confuses subcategories. In most cases, these confusions occur within the maintenance- or innovation-relevant subcategories and thus leave the binary innovation classification unaffected.

Table C1. Representative Examples of Release Note Classifications

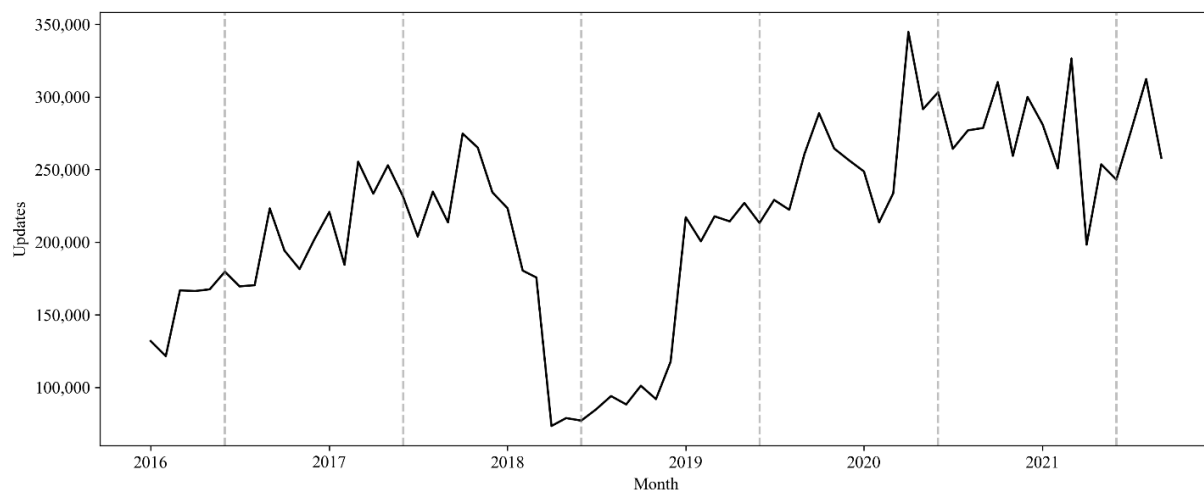
Release Note	Version Change	True Label	Dictionary	Versioning	LLM
Problems with Metal graphic api solved, now the drawings won't be showed with a black background	2.0 → 3.0	4 (0)	0	1	3 (0)
This new version contains patches and performance improvements.	5.3.272 → 5.3.274	6 (0)	1	0	6 (0)
<ul style="list-style-type: none"> • New set of levels - Spain • New game mode - Limited Number of Moves • New app icon • Translation corrected 	1.0.0 → 1.1.0	2 (1)	1	0	2 (1)
If enabled, the app will show the viewer a prompt to take an action at a designated time during a live stream.	5.9 → 5.10.1	1 (1)	0	0	1 (1)

Notes: True labels follow the categorization scheme of the classification prompt (see Figure 2), with human coding as reference. Numbers denote update categories; the binary innovation classification underlying our main dependent variable is reported in parentheses. The dictionary- and versioning-based approaches yield only a binary innovation classification.

Appendix D: App Development Seasonality

Figure D1 displays the monthly count of app updates from January 2016 to September 2021. The series suggests seasonality in app development and release cycles. In particular, there are recurring visible increases around June (dashed lines), which plausibly align with Apple's Worldwide Developers Conference, when developers typically gain access to new platform announcements, SDKs, and beta releases. There also appear to be further spikes around major autumn iOS release windows, when Apple opens submission for apps built against the newest iOS versions (e.g., iOS 13 on September 19, 2019). By contrast, update counts tend to decline around Christmas, which is consistent with Apple's holiday period, when app submission and review activity is typically limited and releases therefore need to be prepared earlier. The pronounced decline in 2018 is harder to interpret. While it could indicate a major change in App Store development conditions, it may also reflect incomplete coverage during that period. As this interval is excluded from the main analysis, we do not examine these patterns further.

Figure D1. Updates Trend (2016-2021)



Notes: Dashed lines denote the month in which Apple conducted its Worldwide Developers Conference each year (June). Updates trend provided for all apps that are listed on the iOS App Store as of January 1, 2020, and display an English app description on their App Store profile ($N = 1,294,110$).

Appendix E: Before-After Estimations

Table E1 reports regressions of feature updates on the post-announcement indicator, estimated separately for affected and unaffected apps. The models take the following form:

$$I_{i,t} = \beta_0 + \beta_1 \text{After } ATT_t + X_{i,t} + \eta_i + \epsilon_{i,t}, \quad (\text{E1})$$

where $I_{i,t}$ denotes the *Feature Updates* of app i in month t . *After* ATT_t equals 1 for months following the announcement. β_1 captures within-group change in innovation activity after the announcement. App fixed effects η_i absorb time-invariant differences between apps. $X_{i,t}$ is a vector of time-varying controls.

Table E1. Before-After Regression on *Feature Updates* (Count)

	DV: Feature Updates (Count)			
	ATT Affected		ATT Unaffected	
	(1) Baseline	(2) Baseline + Controls	(3) Baseline	(4) Baseline + Controls
After ATT	−.025*** (.003)	−.019*** (.003)	−.058*** (.003)	−.053*** (.003)
Toprank		.036*** (.005)		.022*** (.004)
Reviews		.001*** (.000)		.001*** (.000)
Developer Portfolio Size		−.001*** (.000)		.000 (.000)
Developer Portfolio Breadth		.007 (.010)		−.032*** (.009)
In-App Purchases		−.016 (.013)		−.032 (.023)
Constant	.169*** (.002)	.205*** (.039)	.162*** (.001)	.266*** (.031)
App Fixed Effects	Yes	Yes	Yes	Yes
Clusters	6,701	6,456	6,701	6,374
Observations	178,818	172,251	177,289	168,614
Adj. R ²	.330	.337	.272	.277

Notes: Observations are app-months. The controlled specifications exclude 572 apps with missing developer-level information. The announcement month June 2020 is excluded from the estimation. Clustered standard errors at the app level in parentheses. +, *, **, *** denote significance at the 10%, 5%, 1%, and .1% levels, respectively.

The coefficient on *After ATT* is negative and statistically significant in both groups. In the baseline specifications, feature updates fall by .025 per app-month among affected apps (Column 1) and by .058 among unaffected apps (Column 3). Relative to the pre-announcement means of .170 and .163, this corresponds to declines of about 15% and 36%. The estimates change little when time-varying controls are added (Columns 2 and 4). These within-group changes include the decline in early 2020 that both groups share (Figure 3 of the main article).

Appendix F: Sensitivity of the Main Estimates

F1 Alternative Matching Procedures

Our main sample matches each affected app to one unaffected app on the propensity score without replacement. This procedure retains 6,701 of the 10,668 affected apps. Three concerns may arise from this procedure. Matched apps may differ in when they entered the panel, part of the affected apps remains unmatched, and the results may depend on the propensity score and its variables.

We address each concern with an alternative matched sample (Table F1). First, we additionally require an exact match on the month in which an app enters the panel (Column 1). Our panel is unbalanced, as apps continue to enter during the pre-announcement period. Apps tend to release more updates shortly after their launch than once they are established. A recently released app and an app that has been available for several years may thus follow different trajectories even if their pre-announcement averages are similar. Exact matching on the entry month compares apps at the same stage of their lifecycle. Second, we match with replacement (Column 2). In the main sample, 3,967 affected apps remain unmatched. These are mainly apps with very high update and review activity, for which comparable apps are scarce. Matching with replacement retains all affected apps, as an unaffected app can serve as the match for several affected apps. Unaffected apps are weighted by the number of affected apps they are matched to. Third, we use coarsened exact matching (Column 3; Iacus et al. 2012). This procedure does not rely on an estimated propensity score. Apps are grouped into cells defined by coarsened values of the matching variables. Unaffected apps are weighted so that their distribution across cells equals that of affected apps. The procedure also retains all affected apps.

The estimates confirm the main results. The coefficient on *Affected* $\times$ *After ATT* is positive and statistically significant in all samples (Panel A). It ranges from .032 to .056 additional feature updates per app-month, compared with .033 in the main sample. Panel B shows that the effect is larger after implementation than in the announcement phase in every sample. Panel C shows that the likelihood of a feature update is also higher for affected apps in all samples. The samples in Columns (2) and (3) include affected apps with high update activity that remain unmatched in the main sample. As such, their pre-

announcement mean is higher (.378). Relative to this mean, the estimates correspond to increases of about 15% and 9%.

Table F1. Effect on *Feature Updates* (Count) with Alternative Matching Procedures

	DV: Feature Updates (Count)		
	(1) Entry Month OLS	(2) With Replacement OLS	(3) Coarsened Exact Matching OLS
<i>Panel A. Feature Update (Count)</i>			
Affected $\times$ After ATT	.032*** (.004)	.056*** (.014)	.033*** (.010)
<i>Panel B. Feature Update (Count), by phase</i>			
Affected $\times$ Announcement Phase	.023*** (.004)	.037* (.014)	.021* (.010)
Affected $\times$ Post-Implementation	.045*** (.005)	.085*** (.016)	.052*** (.011)
<i>Panel C. Feature Update (Binary)</i>			
Affected $\times$ After ATT	.021*** (.003)	.032*** (.008)	.017** (.005)
Pre-announcement mean (Count)	.169	.378	.378
Pre-announcement mean (Binary)	.125	.247	.247
Affected apps	6,522	10,668	10,668
Unaffected apps	6,522	3,247	15,386
Largest weight	1	302	15.9
Clusters	13,044	13,915	26,054
Observations	347,532	370,081	695,178

Notes: Each coefficient block reports a separate regression with app and month fixed effects. (1) One-to-one nearest-neighbor matching on the propensity score without replacement, with exact matching on the month of entry into the panel. (2) Nearest-neighbor matching on the propensity score with replacement, unaffected apps weighted by the number of matches. (3) Coarsened exact matching on pre-announcement updates, reviews, review rating, and toprank position. Pre-announcement means refer to affected apps. Observations are app-months. The announcement month June 2020 is excluded from the estimation. Clustered standard errors at the app level in parentheses. +, *, **, *** denote significance at the 10%, 5%, 1%, and .1% levels, respectively.

F2 Alternative Model Specification

Our main specifications estimate linear models on the monthly count of feature updates. As the dependent variable is a non-negative count with many zeros, we re-estimate the difference-in-differences models with a pseudo-Poisson maximum likelihood estimator with two-way fixed effects (Correia et al. 2020). Table F2 reports the results.

Table F2. Pseudo-Poisson Estimates of the Effect on *Feature Updates* (Count)

	DV: Feature Updates (Count)			
	(1) Baseline PPML	(2) Baseline + Controls PPML	(3) Split Baseline PPML	(4) Split Baseline + Controls PPML
Affected $\times$ After ATT	.283*** (.030)	.311*** (.030)		
Affected $\times$ Announcement Phase			.196*** (.030)	.216*** (.031)
Affected $\times$ Post-Implementation			.443*** (.038)	.487*** (.039)
Toprank		.220*** (.019)		.222*** (.019)
Reviews		.001*** (.000)		.001*** (.000)
Developer Portfolio Size		-.011*** (.002)		-.011*** (.002)
Developer Portfolio Breadth		-.017 (.050)		-.017 (.050)
In-App Purchases		.098 (.071)		.101 (.070)
Constant	-.930*** (.008)	-.700*** (.166)	-.931*** (.008)	-.693*** (.165)
App Fixed Effects	Yes	Yes	Yes	Yes
Month Fixed Effects	Yes	Yes	Yes	Yes
Clusters	7,869	7,487	7,869	7,487
Observations	209,263	199,103	209,263	199,103
Pseudo R ²	.217	.222	.217	.222

Notes: Apps without any feature update during the observation period are not identified in models with app fixed effects and are excluded (5,533 of 13,402 apps). The controlled specifications exclude a further 382 apps with missing developer-level information. Observations are app-months. The announcement month June 2020 is excluded from the estimation. Clustered standard errors at the app level in parentheses. ⁺, *, **, *** denote significance at the 10%, 5%, 1%, and .1% levels, respectively.

The estimates confirm the main results. The coefficient on *Affected $\times$ After ATT* is positive and statistically significant (Column 1) and similar when controls are included (Column 2).² Columns (3) and (4) separate the two policy phases. The coefficient on *Affected $\times$ Post-Implementation* is about twice the coefficient on *Affected $\times$ Announcement Phase*. This pattern mirrors the results in the main article, with a divergence that sets in after the announcement and becomes most pronounced after implementation.

² Pseudo-Poisson coefficients are log-point changes. The proportional effect is $\exp(\beta) - 1$, relative to the level affected apps would have reached without ATT. For Column (1), this is about 33%.

F3 Comparison Within App Categories

Our matching procedure is based on app-level activity and popularity before the announcement and does not condition on app categories. Differences between categories that are constant over time are absorbed by the app fixed effects. However, categories may follow different trajectories over the observation period, for example because of potential category-specific changes in demand during the COVID-19 pandemic. Such developments are not captured by month fixed effects, which is why we replace them with category-by-month fixed effects. The effect is then identified from affected and unaffected apps within the same category and month.

Table F3 reports the results. The coefficient on *Affected* $\times$ *After ATT* remains positive and statistically significant for both outcomes, with and without controls (Columns 1, 2, 4, and 5). The estimates are slightly larger than in the main specification. Column (3) shows that the effect is again larger after implementation than in the announcement phase.

Table F3. Effect on *Feature Updates* with Category-by-Month Fixed Effects

	DV: Feature Updates (Count)			Feature Updates (Binary)	
	(1) Baseline OLS	(2) Baseline + Controls OLS	(3) Split Baseline OLS	(4) Baseline LPM	(5) Baseline + Control LPM
Affected $\times$ After ATT	.038*** (.004)	.039*** (.004)		.024*** (.003)	.025*** (.003)
Affected $\times$ Announcement Phase			.031*** (.004)		
Affected $\times$ Post-Implementation			.049*** (.005)		
Toprank		.027*** (.003)			.019*** (.002)
Reviews		.001*** (.000)			.000*** (.000)
Developer Portfolio Size		-.001** (.000)			-.000 (.000)
Developer Portfolio Breadth		-.007 (.007)			-.005 (.004)
In-App Purchases		-.019+ (.011)			-.005 (.006)
Constant	.132*** (.001)	.179*** (.027)	.132*** (.001)	.100*** (.001)	.121*** (.015)
App Fixed Effects	Yes	Yes	Yes	Yes	Yes

Category $\times$ Month Fixed Effects	Yes	Yes	Yes	Yes	Yes
Clusters	13,402	12,830	13,402	13,402	12,830
Observations	356,107	340,865	356,107	356,107	340,865
Adj. R ²	.309	.315	.309	.270	.275

Notes: All models include app fixed effects and category-by-month fixed effects for 26 app categories. Two categories contain unaffected apps only (106 apps) and do not contribute to the estimates. The controlled specifications exclude 572 apps with missing developer-level information. Observations are app-months. The announcement month June 2020 is excluded from the estimation. Clustered standard errors at the app level in parentheses. +, *, **, *** denote significance at the 10%, 5%, 1%, and .1% levels, respectively.

F4 Sample Restricted to Free Apps

In the main sample, most unaffected apps are paid apps (5,460 of 6,701). Paid apps may differ from affected apps in ways that also shape how they develop over time. We therefore assess whether our findings hold when the unaffected group is restricted to free apps and exclude all paid apps from the main sample. Table F4 reports the results. The effect on feature updates remains positive and statistically significant (Column 1) and is larger after implementation than in the announcement phase (Column 2). The same holds for the likelihood of a feature update (Column 3).

Table F4. Effect on *Feature Updates* in a Sample Restricted to Free Apps

	DV: Feature Updates (Count)		Feature Updates (Binary)
	(1) Baseline OLS	(2) Split Baseline OLS	(3) Baseline LPM
Affected $\times$ After ATT	.034*** (.009)		.021*** (.006)
Affected $\times$ Announcement Phase		.022* (.010)	
Affected $\times$ Post-Implementation		.052*** (.010)	
Constant	.152*** (.004)	.152*** (.004)	.115*** (.003)
App Fixed Effects	Yes	Yes	Yes
Month Fixed Effects	Yes	Yes	Yes
Clusters	7,942	7,942	7,942
Observations	211,504	211,504	211,504
Adj. R ²	.332	.332	.286

Notes: The sample is the main sample without paid apps (6,701 affected and 1,241 unaffected apps). Observations are app-months. The announcement month June 2020 is excluded from the estimation. Clustered standard errors at the app level in parentheses. +, *, **, *** denote significance at the 10%, 5%, 1%, and .1% levels, respectively.

F5 Potential Interference Between Affected and Unaffected Apps

A core assumption of our empirical strategy is that the feature updates of an unaffected app do not depend on whether other apps are affected by ATT. Two features of the setting could violate this assumption. First, developers with apps in both groups can shift development effort between them. If effort moves toward the affected app, its unaffected sibling app could receive less updates, and the estimate would overstate the effect. Second, additional feature updates of affected apps may prompt developers of unaffected apps in the same category to adjust their own update activity. The estimate would be attenuated if these developers follow and inflated if they fall back. The category-by-month fixed effects (Online Appendix F3) absorb common category trends but not interactions between the two groups within a category.

To examine the first channel, we re-estimate the main specification without developers that have apps in both groups and test whether apps with a sibling app in the other group receive different updates after the announcement. Table F5 reports the results. Developers with apps in both groups account for 1,419 of the 13,402 matched apps. All have multiple apps by construction, and their median portfolio contains 30 apps, compared with 8 for other developers. Because the innovation response declines with portfolio size (Section 6.3), Column (2) restricts the analysis to multi-app developers. After excluding developers with apps in both groups, the estimates remain in line with the main findings (see Table 3 and Table 6 of the main article): .036 feature updates per app-month overall (Column 1) and .026 among multi-app developers (Column 2). Apps with a sibling app in the other group receive the same updates after the announcement as other apps of their group (Columns 3–4; .012 and .008, neither statistically significant). Together with the stable estimates in Columns (1)–(2), this gives no indication that these developers shift effort between affected and unaffected apps in a way that would affect our main conclusions.

Table F5. Developers with Apps in Both Groups

	DV: Feature Updates (Count)			
	(1) Full sample, w/o mixed	(2) Multi-app w/o mixed	(3) Unaffected, multi-app	(4) Affected, multi-app
Affected $\times$ After ATT	.036*** (.004)	.026*** (.005)		
After ATT $\times$ Dev. Has Affected App			.012 (.008)	
After ATT $\times$ Dev. Has Unaffected App				.008 (.011)
After ATT $\times$ Portfolio Size (log)			.014*** (.002)	-.002 (.003)
Constant	.133*** (.001)	.130*** (.001)	.119*** (.001)	.152*** (.001)
App Fixed Effects	Yes	Yes	Yes	Yes
Month Fixed Effects	Yes	Yes	Yes	Yes
Apps	11,983	9,734	5,646	5,507
Clusters	11,983	9,734	4,019	3,605
Observations	318,132	258,756	149,566	147,165
Adj. R ²	.304	.316	.275	.349

Notes: Mixed developers have affected and unaffected apps in the matched sample (1,419 apps). Multi-app developers have two or more apps in May 2020. Standard errors clustered at the app level in Columns (1)–(2) and at the developer level in Columns (3)–(4) in parentheses. Observations are app-months. The announcement month June 2020 is excluded from the estimation. ⁺, *, **, *** denote significance at the 10%, 5%, 1%, and .1% levels, respectively.

To examine the second channel, we use variation in the share of affected apps across categories. If competitive responses influence our estimates, we would expect the estimated effect to vary with this share. The share ranges from about a quarter in Navigation, Productivity, and Reference to over three quarters in News and Magazines. Table F6 provides no evidence of such variation; the interaction estimate is .005 and not statistically significant and the ATT interaction estimate remains at a similar magnitude to the main estimate reported in the article.

Table F6. Exposure to Affected Apps in the Category

	DV: Feature Update (Count)
Affected $\times$ After ATT	.038*** (.004)
Affected $\times$ After ATT $\times$ Share of Affected Apps	.005 (.040)
Constant	.132*** (.001)
App Fixed Effects	Yes

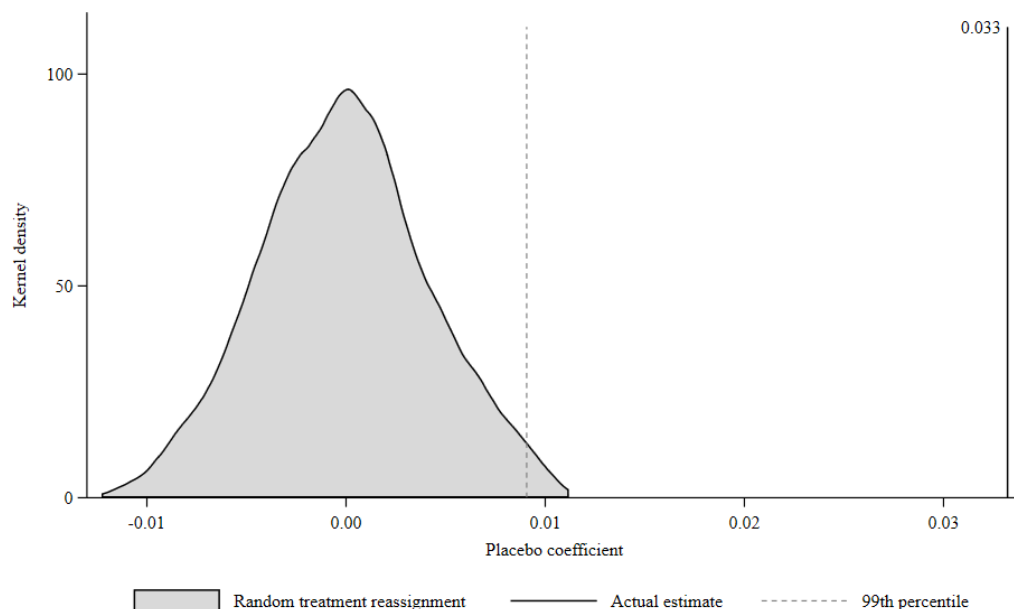
Category $\times$ Month Fixed Effects	Yes
Apps	13,402
Clusters	13,402
Observations	356,107
Adj. R ²	.309

Notes: *Share of Affected Apps* is the share of affected apps among the apps of the category in May 2020, centered at its mean across unaffected apps (.384). Standard errors clustered at the app level in parentheses. Observations are app-months. The announcement month June 2020 is excluded from the estimation. +, *, **, *** denote significance at the 10%, 5%, 1%, and .1% levels, respectively.

F6 Placebo Treatment Assignment and Timing

We conduct two placebo tests to assess whether an estimate of this size could arise without the policy. First, we randomly reassign treatment status across apps, holding the number of affected apps constant, and re-estimate Equation (1) of the main article. Figure F1 shows the distribution of the resulting 500 placebo estimates. The distribution is tightly centered at zero with a standard deviation of .004. The actual estimate of .033 is eight standard deviations above its mean and is more than three times the largest placebo estimate (.010).

Figure F1. Distribution of Placebo Estimates with Randomly Reassigned Treatment Status



Notes: The figure shows the kernel density of 500 placebo estimates obtained by randomly reassigning treatment status across apps and re-estimating Equation 1. The dependent variable is the number of feature updates.

Second, we assign a placebo announcement date in the midpoint of the pre-announcement period (December 2019) and re-estimate the model on the pre-announcement months. As in the main specification, we exclude the month of the placebo announcement. Table F7 reports the results. The placebo estimates are

close to zero and negative, $-.009$ ($p = .09$) for the number of feature updates and $-.005$ ($p = .16$) for the likelihood of a feature update.

Table F7. Test with a Placebo Announcement Date

	DV: Feature Updates (Count)	Feature Updates (Any)
	(1) Baseline OLS	(2) Baseline LPM
Affected $\times$ After Placebo Date	$-.009^+$ (.005)	$-.005$ (.003)
Constant	.167*** (.001)	.127*** (.001)
App Fixed Effects	Yes	Yes
Month Fixed Effects	Yes	Yes
Clusters	13,402	13,402
Observations	141,755	141,755
Adj. R ²	.297	.258

Notes: The models are estimated on the pre-announcement period (June 2019 to May 2020) with a false announcement date in December 2019. As in the main specification, the month of the false announcement is excluded. Observations are app-months. Clustered standard errors at the app level in parentheses. ⁺, *, **, *** denote significance at the 10%, 5%, 1%, and .1% levels, respectively.

Appendix G: Timing of the Innovation Response

Table G1. Effect on *Feature Updates* by Policy Phase

	DV: Feature Update (Count)		Feature Update (Binary)	
	(1) Baseline OLS	(2) Baseline + Controls OLS	(3) Baseline LPM	(4) Baseline + Controls LPM
Affected × Announcement Phase	.025*** (.004)	.026*** (.004)	.014*** (.003)	.015*** (.003)
Affected × Post-Implementation	.046*** (.005)	.049*** (.005)	.033*** (.003)	.034*** (.003)
Toprank		.028*** (.003)		.019*** (.002)
Reviews		.001*** (.000)		.000*** (.000)
Developer Portfolio Size		−.001** (.000)		−.000 (.000)
Developer Portfolio Breadth		−.006 (.007)		−.005 (.004)
In-App Purchases		−.018 (.011)		−.005 (.006)
Constant	.133*** (.001)	.179*** (.027)	.101*** (.001)	.121*** (.015)
Difference Between Phases	.022*** (.004)	.023*** (.004)	.018*** (.003)	.019*** (.003)
App Fixed Effects	Yes	Yes	Yes	Yes
Month Fixed Effects	Yes	Yes	Yes	Yes
Clusters	13,402	12,830	13,402	12,830
Observations	356,107	340,865	356,107	340,865
Adj. R ²	.307	.313	.268	.272

Notes: The announcement phase covers July 2020 to March 2021, the post-implementation phase April to September 2021. Both coefficients are relative to the pre-announcement period. The difference between phases is the post-implementation coefficient minus the announcement-phase coefficient. Observations are app-months. The controlled specifications exclude 572 apps with missing developer-level information. The announcement month June 2020 is excluded from the estimation. Clustered standard errors at the app level in parentheses. +, *, **, *** denote significance at the 10%, 5%, 1%, and .1% levels, respectively.

Appendix H: App Category Opt-in Rates

H1 Opt-in Rates over Time

Table H1. App Category Opt-in Overview

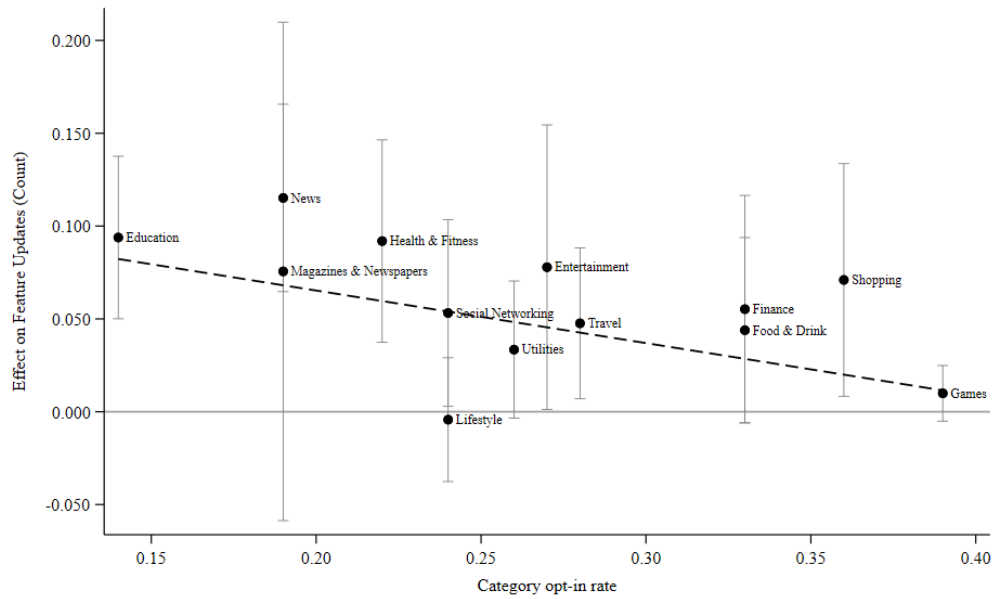
App Category	Opt-in Rate	High/Low Split	Opt-in Rate Rank		
			(2025)	(2024)	(2023)
Education	14 %	Low	13	13	13
Entertainment	27 %	Low	6	5	5
Finance	33 %	High	3	3	4
Food & Drink	33 %	High	3	3	3
Games	39 %	High	1	1	1
Health & Fitness	22 %	Low	10	10	11
Lifestyle	24 %	Low	8	9	8
Magazines & Newspapers	19 %	Low	11	11	9
News	19 %	Low	11	11	9
Shopping	36 %	High	2	2	2
Social Networking	24 %	Low	8	7	7
Travel	28 %	Low	5	6	6
Utilities	26 %	Low	7	7	11

Notes: Opt-in rates reported for 2025. Data available for listed app categories only. The median opt-in rate across apps is 33% (N = 8,220 apps and 218,057 app-months). Categories at or above median are classified as High. Categories below are classified as Low. Rankings shown for 2023-2025.

H2 Category-level Analysis

Opt-in rates are available for 13 of the 26 app categories, which cover 8,220 of the 13,402 matched apps (68% of affected and 54% of unaffected apps). As the variation in opt-in rates is at the category level, we complement the app-level estimates of Table 4 of the main article with one estimate per category. Figure H1 plots these estimates against the corresponding opt-in rates. The estimated effect is largest in categories with low opt-in rates, such as news, education, and health & fitness, and close to zero in games, the category with the highest opt-in rate. A regression of these estimates on category opt-in rates, weighted by the precision of the estimates, results in a slope of $-.28$ ($p = .01$). The Spearman rank correlation is $-.57$ ($p = .04$). These results are consistent with the app-level finding of a stronger innovation response where opt-in rates are lower, though inference rests on a small number of categories.

Figure H1. Effect on *Feature Updates* (Count) by Category Opt-in Rate

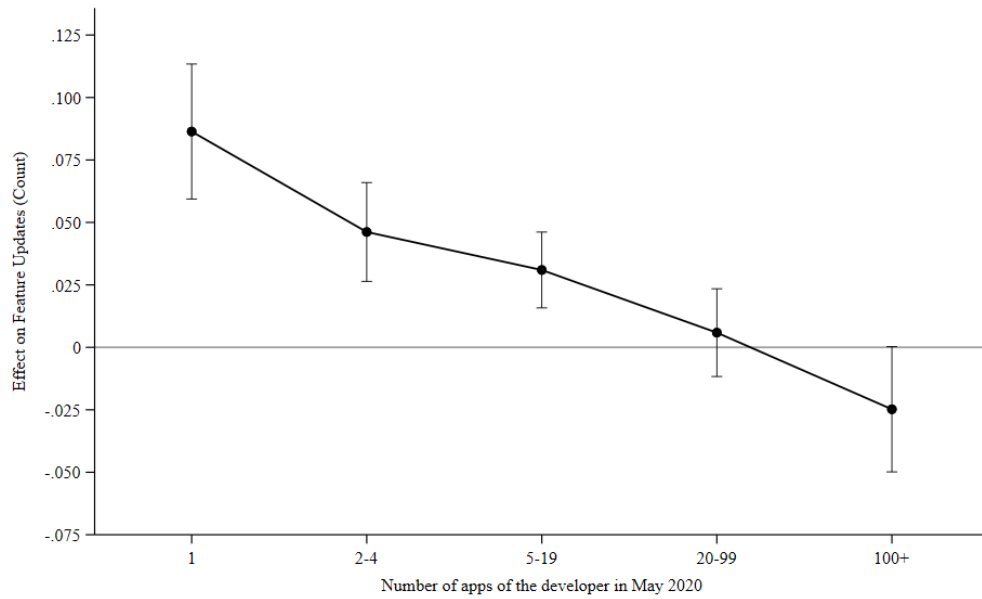


Notes: Each point is the coefficient on $\text{Affected} \times \text{After}$ ATT from a separate difference-in-differences regression for apps in the category, with app and month fixed effects and standard errors clustered at the app level. The dependent variable is the number of feature updates. Vertical bars indicate 95% confidence intervals. The dashed line is the linear fit weighted by the inverse of the squared standard errors of the estimates. Opt-in rates are category-level rates for 2025. The announcement month June 2020 is excluded from the estimation. Categories comprise between 77 (magazines & newspapers) and 3,523 (games) matched apps.

Appendix I: Developer Portfolio Groups

Figure I1 shows the innovation response separately for five groups of developer portfolio size. The effect declines with every step in portfolio size. It is largest for apps from single-app developers (.086) and remains statistically significant for developers with two to four apps (.046) and five to nineteen apps (.031). The estimate is close to zero for 20 to 99 apps (.006) and negative and marginally significant for 100 or more apps ($-.025, p = .054$).

Figure I1. Effect on *Feature Updates* (Count) by Developer Portfolio Size



Notes: The figure shows estimates of *Affected* $\times$ *After ATT* from separate regressions for five groups of apps, defined by the average number of apps of the app's developer in the pre-announcement period. The dependent variable is the number of feature updates. All models include app and month fixed effects. Vertical bars indicate 95% confidence intervals based on standard errors clustered at the developer level. The groups comprise 1,677, 2,524, 4,231, 3,244, and 1,154 apps from 1,677, 2,324, 3,056, 1,580, and 269 developers. Portfolio size refers to all apps of a developer in the App Store, of which only those in the matched sample are considered in the estimation. The announcement month June 2020 is excluded from the estimation.

Appendix J: Mechanism Analyses

J1 Potential Adoption and Abandonment of In-App Purchases

Table J1 reports discrete-time hazard estimates for entry into and exit from in-app purchases on the main sample. For an app i at risk in month t , the hazard of the event follows

$$h_{it} = 1 - \exp[-\exp(\alpha_t + \gamma \text{Affected}_i + \beta \text{Affected}_i \times \text{AfterATT}_t)], \quad (\text{J1})$$

where α_t are month fixed effects that trace the baseline hazard, γ captures the difference in baseline hazards between affected and unaffected apps, and β is the proportional shift in the hazard of affected apps after the announcement relative to unaffected apps, with $\exp(\beta) - 1$ reported as the implied hazard shift. The split phases specification replaces $\beta \text{Affected}_i \times \text{AfterATT}_t$ with separate terms for the announcement phase and the post-implementation phase. Apps leave the risk set after their first event.

Table J1. Hazard of Entering and Exiting In-App Purchases

	DV: IAP Entry		IAP Exit	
	(1) DiD	(2) Split Phases	(3) DiD	(4) Split Phases
Affected $\times$ After ATT	-.239 ⁺ (.143)		-.409* (.199)	
Affected $\times$ Announcement Phase		-.296 ⁺ (.163)		-.561** (.215)
Affected $\times$ Post-Implementation		-.110 (.231)		-.182 (.241)
Affected	1.665*** (.080)	1.665*** (.080)	.693*** (.169)	.693*** (.169)
Month Fixed Effects	Yes	Yes	Yes	Yes
Implied hazard shift	-21%	-26% / -10%	-34%	-43% / -17%
Pre-ATT hazard, affected apps	.023	.023	.005	.005
Events	1,085	1,085	820	820
At-risk app-months	189,596	189,596	148,833	148,833
Apps	8,603	8,603	6,355	6,355

Notes: Complementary log-log discrete-time hazard models with month fixed effects. The risk set for entry consists of app-months in which an app had no in-app purchases at the start of the month, the risk set for exit of app-months in which it had; apps leave the risk set after their first event, and repeat events of the same app are not used. Implied hazard shifts are $\exp(\beta) - 1$; for the phase models, announcement phase / post-implementation. Pre-ATT hazard is the monthly event rate of affected apps at risk before the announcement. The announcement month June 2020 is excluded. Standard errors clustered at the app level in parentheses. ⁺, *, **, *** denote significance at the 10%, 5%, 1%, and .1% levels, respectively.

Affected apps have a higher baseline hazard of entering in-app purchases, but their hazard does not rise after the announcement. The estimates are negative in both phases, significant at the 10% level in the announcement phase and insignificant after implementation. The hazard of exiting in-app purchases

declines for affected apps by about 43% in the announcement phase and is unchanged after implementation. Affected apps therefore do not adopt in-app purchases at a higher rate, and those that have them are less likely to give them up while the policy is pending.

J2 In-Depth Classification of Feature Content

The analysis of feature content builds on the classification of release notes described in the main article. The fine-tuned model identifies updates that introduce new features or new content. In a second step we now classify what these extensions add to an app. The sample comprises all 52,878 feature updates of the apps in the main sample between June 2019 and September 2021.

We use a general-purpose model (GPT-5.2) with a few-shot prompt because the content categories differ from the feature–maintenance distinction on which the fine-tuned classifier was trained. The prompt defines 13 categories and instructs the model to select all categories that describe what an update introduces or substantially extends (Figure J1). Accompanying bug fixes, performance improvements, minor polish, and generic statements receive no label, even if they would relate to one of the specified categories. An update can receive multiple labels.

Figure J1. Prompt for Feature Content Classification

ONLY provide category numbers in response, separated by commas (for example: 2,6). Provide 0 if none applies.

The following app update text has been classified as containing a new feature or new content.

Determine ALL categories that describe what is NEWLY ADDED or SUBSTANTIALLY EXTENDED in this update.

Do NOT assign a category for:

- bug fixes, crash fixes, performance or stability improvements, even if they concern one of the categories (for example, “fixed a login issue” is not category 1, “fixed an issue with buying subscriptions” is not category 6)
- minor tweaks or polish of existing functions
- greetings, requests to rate or review the app, links to social media, general app descriptions, and generic phrases such as “we made the app better for you”

- (1) Accounts & user data – sign-up or login options, user profiles, syncing a user's data across devices, onboarding questionnaires or in-app surveys about the user
- (2) Personalization – recommendations, personalized feeds or home screens, favorites, watchlists or wishlists, content tailored to the user's past behavior or preferences

- (3) Notifications & re-engagement – push notifications, reminders, alerts, home-screen widgets, daily rewards or streaks
- (4) Social & community – interaction between users of the app: friends or followers, chat or messaging, comments, leaderboards, multiplayer, user groups or clans
- (5) Sharing & referral – sharing content to other apps or people, invitations, referral rewards
- (6) New paid offer – a subscription, membership, premium tier, free trial or purchasable item is introduced for the first time or as a new option
- (7) Features for paying users – new functions or content that are available only to subscribers, members or premium users of an existing paid offer
- (8) Paid removal of ads – an option to pay in order to remove or reduce advertising
- (9) Advertising – changes to the ads shown in the app (fewer or more ads, rewarded ads, new ad formats), other than category 8
- (10) Privacy & tracking – tracking permission prompts, consent or privacy settings, user controls over data
- (11) New content – levels, episodes, items, articles, catalog additions, seasonal events
- (12) Core functionality – new or extended functions that serve the app's main purpose and fit none of the categories above
- (13) Design & usability – redesigns, navigation changes, dark mode, accessibility, new languages

Examples:

Update text: “Sign in with Apple is here! Also added 20 new levels and fixed a crash on iPad.”

Response: 1,11

Update text: “Introducing Premium: go ad-free and unlock cloud backup. Minor improvements.”

Response: 6,8,7

Update text: “Bug fixes and performance improvements. Thanks for using our app, please leave a review!”

Response: 0

Update text: {*release note*}

Notes: Release notes are whitespace-normalized and truncated at 3,000 characters. Requests use gpt-5.2 with temperature 0, a fixed seed (94032) and a limit of 30 completion tokens. Each distinct text is classified once.

Table J2 lists the categories with a short definition, reports the share of feature updates with each label, and maps the categories to the adaptation paths (theme) discussed in the main article. Core functionality (51%), design and usability (34%), and new content (32%) are the most frequent labels. Categories related to monetization, data collection, or user growth are less common, with shares between 0.3% and 7%. For 11% of feature updates, the release notes announce new features without describing them sufficiently to assign a category.

Table J2. Categories of Feature Content

	Category	Definition	Theme	Share of Feature Updates (%)
1	Accounts and user data	Features that identify users or collect information about them	First-party data	5.7
2	Personalization	Features that tailor what the app shows to the individual user	First-party data	4.5
3	Notifications and re-engagement	Features that bring users back to the app	Engagement	6.7
4	Social and community	Features that enable interaction between users of the app	Engagement	6.0
5	Sharing and referral	Features through which users carry the app or its content to others	Growth	3.2
6	New paid offer	A paid option introduced for the first time	Monetization	4.0
7	Features for paying users	New functions or content restricted to users of an existing paid offer	Monetization	2.4
8	Paid removal of ads	An option to pay in order to remove or reduce advertising	Monetization	.3
9	Advertising	Changes to the ads shown in the app, other than a paid removal	Other	.6
10	Privacy and tracking	Features that concern the handling of user data and consent	Other	1.0
11	New content	Additions within the app’s existing functions	Engagement	31.6
12	Core functionality	New or extended functions that serve the app’s main purpose and fit no other category	Other	50.7
13	Design and usability	Changes to how the app looks and is operated	Other	34.1
0	No category	The note announces new features or content without naming an addition that fits a category		11.3

Notes: Shares refer to the 52,878 feature updates of the apps in the main sample between June 2019 and September 2021. An update can have several labels, so shares do not sum to 100%.

To validate the classification, we manually coded 400 randomly selected release notes using the category definitions in the prompt. Against these codes, the model achieves F1-scores of .808 when micro-averaged across labels, .729 when macro-averaged across categories, and .810 when weighted by category frequency. Precision is .743 and recall is .886. The gap between precision and recall reflects a tendency to assign additional labels. While the model reliably captures most labels assigned by the coder, it also assigns

additional labels, mainly for core functionality and design and usability, to updates that the coder classifies as adjustments to existing functions, compatibility updates, or localization. The lower macro-averaged F1-score reflects weaker agreement for rare categories, which appear in few validation texts. Additional labels can inflate category counts in both groups and periods. To the extent that this inflation is proportionally similar across groups and periods, the relative effects reported in section J3 are less sensitive to these errors because each estimate is scaled by the pre-announcement mean of the corresponding count (see also Online Appendix L).

J3 Effects by Content Category and In-App Purchase Status

Table J3 reports the estimates by content category that correspond to the mechanism analysis of the main article. The increase in feature updates is concentrated in core functionality (30%), design and usability (41%), sharing and referral (56%), and personalization (37%). The pooled estimates show no change in new content, notifications, or monetization categories. Consistent with ATT's consent requirements, updates related to privacy and tracking increase after implementation.

Table J3. Effect on *Feature Updates* (Count) by Content Category

	Dependent Variable	Mean (Pre-ATT)	Affected × After ATT	Relative Size	Affected × Announcement	Affected × Implementation
	Feature updates (all)	.170	.033*** (.004)	20%	.025*** (.004)	.046*** (.005)
	First-party data (1, 2)	.017	.004** (.001)	23%	.002 (.001)	.007*** (.001)
1	Accounts and user data	.010	.001 (.001)	5%	.000 (.001)	.001 (.001)
2	Personalization	.008	.003** (.001)	37%	.001 (.001)	.006*** (.001)
	Engagement (3, 4, 11)	.082	.001 (.003)	1%	-.001 (.003)	.003 (.003)
3	Notifications and re-engagement	.012	-.001 (.001)	-9%	-.001 (.001)	-.001 (.001)
4	Social and community	.013	.001 (.001)	11%	.002 ⁺ (.001)	.001 (.001)
11	New content	.068	-.000 (.002)	0%	-.001 (.002)	.001 (.003)
5	Sharing and referral	.004	.003*** (.001)	56%	.002** (.001)	.003** (.001)

	Monetization (6, 7, 8)	.012	.001 (.001)	8%	.001 (.001)	.001 (.001)
6	New paid offer	.008	.001 (.001)	10%	.001 (.001)	.001 (.001)
7	Features for paying users	.004	.001 (.001)	15%	.001 (.001)	.001 (.001)
8	Paid removal of ads	.001	-.000** (.000)	-44%	-.000+ (.000)	-.001*** (.000)
	Other (9, 10, 12, 13)	.102	.032*** (.003)	31%	.023*** (.003)	.044*** (.004)
9	Advertising	.002	-.000 (.000)	-22%	-.000 (.000)	-.001+ (.000)
10	Privacy and tracking	.001	.001+ (.000)	55%	-.000 (.000)	.002*** (.000)
12	Core functionality	.076	.023*** (.003)	30%	.018*** (.003)	.030*** (.003)
13	Design and usability	.053	.022*** (.002)	41%	.016*** (.002)	.031*** (.003)
	No category	.019	.002+ (.001)	12%	.002 (.001)	.003+ (.001)

Notes: Each row is a separate regression of the number of feature updates per app-month carrying the label on the interaction terms, with app and month fixed effects, as in Table 3, Column 1 (N = 356,107 app-months, 13,402 apps). The first row repeats the main outcome. Themes count an update once if it has any of the theme's categories. Mean (Pre-ATT) refers to affected apps. The relative effect is the coefficient divided by Mean (Pre-ATT), computed from unrounded estimates. An update can have several labels, so category estimates do not add up to the total. The announcement month June 2020 is excluded from the estimation. Clustered standard errors at the app level in parentheses. +, *, **, *** denote significance at the 10%, 5%, 1%, and .1% levels, respectively.

We also differentiate estimates based on apps' in-app purchase status in May 2020. This distinction matters for monetization updates because the groups differ in their existing offerings. Most affected apps combine advertising with in-app purchases (66%). Unaffected apps include paid apps and free apps without advertising, and only 23% offer in-app purchases. Apps without in-app purchases often already charge an upfront price, offer paid options outside the app, or operate without direct user payments. The pooled estimate in Table J4 therefore compares groups with different scope to extend their paid offerings. Within each status group, monetization-related updates increase by 76% among apps without in-app purchases and by 45% among apps with them. The content of the response also differs in other respects. Apps without in-app purchases release more updates related to new content and sharing, and their engagement-related updates increase by 27%. Among apps with in-app purchases, the increase is confined to core functionality, design and usability, and new paid offerings.

Table J4. Effect on *Feature Updates* (Count) by In-App Purchase (IAP) Status

Dependent Variable	w/o IAP			IAP			Difference (IAP – w/o IAP)
	Mean (Pre-ATT)	Affected × After ATT	Relative Size	Mean (Pre-ATT)	Affected × After ATT	Relative Size	
Feature updates (all)	.143	.053*** (.006)	37%	.183	.033*** (.008)	18%	–.020* (.010)
First-party data (1, 2)	.020	.003+ (.002)	16%	.015	.003 (.002)	19%	–.000 (.003)
1 Accounts and user data	.012	.001 (.001)	6%	.009	–.000 (.002)	–3%	–.001 (.002)
2 Personalization	.009	.003* (.001)	29%	.007	.002 (.002)	26%	–.001 (.002)
Engagement (3, 4, 11)	.044	.012*** (.003)	27%	.101	–.000 (.005)	0%	–.012+ (.006)
3 Notifications and re-engagement	.012	–.001 (.001)	–5%	.011	–.003 (.002)	–29%	–.003 (.003)
4 Social and community	.010	.002+ (.001)	25%	.015	.002 (.002)	17%	.000 (.002)
11 New content	.026	.012*** (.003)	44%	.089	–.001 (.005)	–1%	–.012* (.005)
5 Sharing and referral	.003	.004*** (.001)	112%	.005	.001 (.002)	25%	–.003 (.002)
Monetization (6, 7, 8)	.004	.003** (.001)	76%	.016	.007** (.002)	45%	.005+ (.003)
6 New paid offer	.002	.002* (.001)	90%	.011	.006** (.002)	50%	.004+ (.002)
7 Features for paying users	.001	.001+ (.001)	116%	.006	.002 (.002)	40%	.001 (.002)
8 Paid removal of ads	.000	.000 (.000)	79%	.001	–.001* (.000)	–53%	–.001** (.000)
Other (9, 10, 12, 13)	.100	.043*** (.005)	43%	.103	.029*** (.006)	29%	–.014+ (.008)
9 Advertising	.001	.000 (.000)	6%	.002	–.000 (.001)	–3%	–.000 (.001)
10 Privacy and tracking	.002	.002** (.001)	95%	.001	–.000 (.001)	–44%	–.002* (.001)

12	Core functionality	.078	.028*** (.004)	36%	.075	.023*** (.005)	30%	-.005 (.007)
13	Design and usability	.052	.030*** (.004)	59%	.053	.019*** (.005)	36%	-.011+ (.006)
	No category	.020	.003 (.002)	13%	.018	.004 (.002)	21%	.001 (.003)

Notes: In-app purchase status refers to May 2020. Each row reports three regressions of the number of feature updates per app-month carrying the label: among apps without in-app purchases (N = 195,911 app-months, 7,425 apps), among apps with in-app purchases (N = 160,196 app-months, 5,977 apps), and a pooled model that interacts *Affected* $\times$ *After ATT* and *After ATT* with in-app purchase status (N = 356,107 app-months, 13,402 apps), whose interaction coefficient is reported as the difference. All models include app and month fixed effects, as in Table 3, Column 1. The first row repeats the main outcome. Theme rows count an update once if it has any of the theme's categories. Mean (Pre-ATT) refers to affected apps. The relative effect is the coefficient divided by Mean (Pre-ATT), computed from unrounded estimates. An update can have several labels, so category estimates do not add up to the total. The announcement month June 2020 is excluded from the estimation. Clustered standard errors at the app level in parentheses. +, *, **, *** denote significance at the 10%, 5%, 1%, and .1% levels, respectively.

Appendix K: Extension to All English-Language iOS Apps

K1 Matching and Comparison with the Main Sample

Reviews, ratings, and rankings are observed for apps in the top charts only. For the full population, we therefore estimate the propensity score from three variables that are available for all apps: the average number of updates per month before the announcement, an indicator for in-app purchases in May 2020, and app age in months in May 2020. We match one-to-one without replacement with a caliper of .05, as in the main analysis. Unaffected apps are the smaller group in this sample, so each of them is matched to its nearest affected app. All 434,294 unaffected apps find a match, which retains 77% of the 567,062 affected apps. Unmatched affected apps are mainly apps with in-app purchases, for which comparable unaffected apps are scarce. Table K1 reports the balance before and after matching. Standardized differences after matching are below .04 for all three matching variables.

Table K1. Propensity Score Matching and Pre-Announcement Differences of Groups

Matching Variable	Pre-Matching			Post-Matching		
	Mean Affected	Mean Unaffected	SMD	Mean Affected	Mean Unaffected	SMD
Updates per month	.091	.066	.109	.067	.066	.007
In-App Purchases	.224	.067	.457	.067	.067	-.000
App Age (months)	47.0	47.0	.000	45.9	47.0	-.039

Notes: The pre-matching sample comprises 567,062 affected and 434,294 unaffected apps, the post-matching sample 434,294 affected and 434,294 unaffected apps. Matching covariates are averaged over the pre-announcement period. SMD denotes the standardized mean difference.

Because the matching in the full population rests on fewer and different variables than the matching in the main sample, we check whether this difference affects the estimates. To this end, we apply the matching of the full population to the 26,054 apps of the pre-matching sample. This results in 5,270 matched pairs, with standardized differences below .10 for all matching variables. Table K2 compares the estimates. With the matching of the full population, the effect among top-ranked apps is .046 feature updates per app-month, reflecting an increase of about 25% relative to the pre-announcement mean of affected apps and about 36% in the pseudo-Poisson model. These values are close to the main results of .033, 20%, and 33%. The smaller effect in the full population therefore reflects the population of apps and not the matching variables.

Table K2. Effect on *Feature Updates* (Count) in the Main Sample and the Full Population

	DV: Feature Updates (Count)		
	(1) Main Sample, Main Matching	(2) Main Sample, Matching on Full Population	(3) Full Population
<i>Feature Update (Count), OLS</i>			
Affected $\times$ After ATT	.033*** (.004)	.046*** (.005)	.002*** (.000)
Mean (Before ATT)	.170	.186	.032
Relative effect	20%	25%	8%
<i>Feature Update (Count), PPML</i>			
Affected $\times$ After ATT	.283*** (.030)	.309*** (.031)	.095*** (.009)
Implied effect	33%	36%	10%
<i>Feature Update (Binary), LPM</i>			
Affected $\times$ After ATT	.022*** (.003)	.029*** (.003)	.002*** (.000)
Mean (Before ATT)	.127	.140	.027
Relative effect	17%	21%	7%
Matched pairs	6,701	5,270	434,294

Notes: Column (1) repeats the main results (Table 3 and Table F2). Column (2) applies the matching of the full population, on pre-announcement update activity, in-app purchases, and app age, to the apps of the main sample. Column (3) repeats Table 7. All models include app and month fixed effects. Mean (Before ATT) refers to affected apps. Relative effects are the coefficient divided by Mean (Before ATT), computed from unrounded estimates. Implied effects of the pseudo-Poisson models are $\exp(\text{coefficient}) - 1$. Observations are app-months. The announcement month June 2020 is excluded from the estimation. Clustered standard errors at the app level in parentheses. ⁺, *, **, *** denote significance at the 10%, 5%, 1%, and .1% levels, respectively.

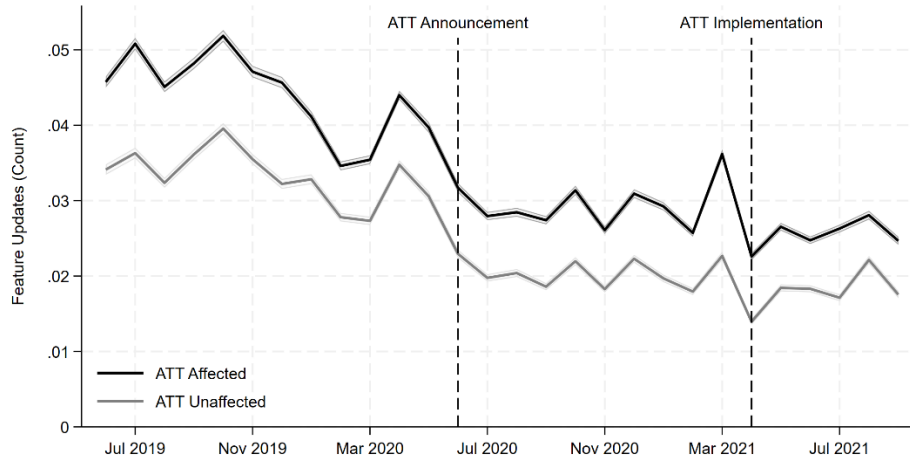
K2 The Full Population without Matching

Table K3 reports the comparison without matching. Both groups lose a large share of their feature updates after the announcement, 37% for affected and 43% for unaffected apps (before-after regressions with app fixed effects). Because affected apps start from a higher level (.044 against .033 feature updates per month), their absolute decline is larger, and the coefficient on *Affected $\times$ After ATT* in the linear models is negative (Columns 1 and 3). In proportional terms, the comparison shows an increase of about 10% (Column 2), close to the matched estimate. Figure K1 shows the monthly means of both groups.

Table K3. Effect on *Feature Updates* without Matching

	DV: Feature Update (Count)		Feature Update (Binary)
	(1)	(2)	(3)
	DiD	DiD	DiD
	OLS	PPML	LPM
Affected $\times$ After ATT	-.002*** (.000)	.094*** (.008)	-.001*** (.000)
Constant	.031*** (.000)	-1.318*** (.002)	.025*** (.000)
App Fixed Effects	Yes	Yes	Yes
Month Fixed Effects	Yes	Yes	Yes
Clusters	1,001,356	220,122	1,001,356
Observations	26,776,016	5,841,176	26,776,016
Adj. R ² / Pseudo R ²	.273	.214	.228

Notes: Column (2) reports pseudo-Poisson estimates for apps with at least one feature update during the observation period. Observations are app-months. The announcement month June 2020 is excluded from the estimation. Clustered standard errors at the app level in parentheses. +, *, **, *** denote significance at the 10%, 5%, 1%, and .1% levels, respectively.

Figure K1. Feature Updates by Group in the Full Population without Matching

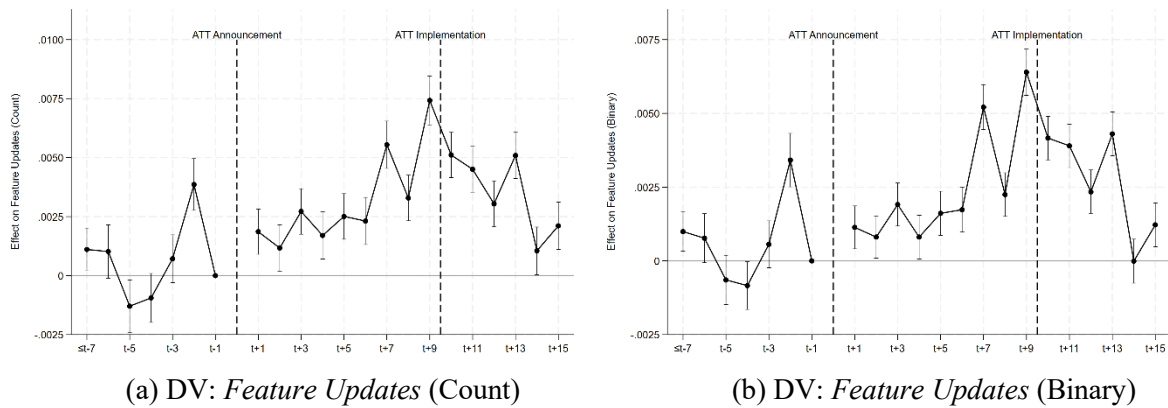
Notes: Monthly means of the number of feature updates for all 567,062 affected and 434,294 unaffected apps. Shaded bands indicate 95% confidence intervals.

K3 Event Study

Figure K2 reports event-study estimates for the matched pairs in the full population. Panel (a) shows the number of feature updates, Panel (b) the likelihood of a feature update. Before the announcement, estimates in both panels fluctuate around zero without systematic drift, although the April 2020 coefficients are similar in magnitude to the estimated post-announcement effect. The monthly estimates are noisier than in the main sample because innovation activity is rare in this population, so individual months are sensitive to small

differences in update activity. In the post-announcement period, both panels show a divergence that sets in after the announcement. The estimates are positive throughout and rise again markedly around the implementation of ATT in April 2021, before they return toward (a) or below (b) the level of the announcement phase. Overall, the dynamic estimates show a response in the full population that follows the timing found in the main sample.

Figure K2. Effect on *Feature Updates* over Time in the Full Population



Notes: The figure plots coefficients on the interaction of the affected indicator with month indicators, estimated on the matched sample with app and month fixed effects and without controls ($N = 23,209,073$; clusters = 868,588). The reference month is May 2020 ($t-1$), the month before the ATT announcement. The announcement month June 2020 (t) is excluded from the estimation. Months before December 2019 are pooled into the first coefficient ($\leq t-7$). Bars indicate 95% confidence intervals based on standard errors clustered at the app level.

Appendix L: Sensitivity of the LLM-based Innovation Variable

L1 Alternative Innovation Measures

To assess whether the results depend on the specific operationalization of innovation, we reconstruct the innovation variable using several alternative approaches from prior literature as well as additional machine-learning-based classifiers and large language models (LLMs).

First, we implement two dictionary- and rule-based approaches based on prior work. Following Kircher and Foerderer (2024), an update is classified as innovative if its release note contains at least one of the terms “new”, “added”, “upgrade”, or “major”. Following Agarwal and Kapoor (2023), an update is classified as innovative if the release note exceeds 200 characters or contains at least one of the keywords “introduce”, “feature”, “support”, “performance”, “improve”, “enable”, “update”, “enhance”, “modify”, “optimize”, “fast”, “adjust”, or “multitask”. To ensure comparability across these text-based measures, we apply a common preprocessing pipeline to release notes including tokenization, lowercasing, removal of punctuation, numbers, symbols, and English stop words, followed by stemming. The same stemming procedure is applied to the dictionary terms.

Second, we construct version-number-based proxies for innovation building on Wen and Zhu (2019). Their approach classifies an update as innovative when the major component of the version number changes relative to the previous release (first digit, e.g., $1.4 \rightarrow 2.0$). Because major-version changes are rare in our setting and may miss feature changes, we also consider a more permissive variant that classifies an update as innovative when either the major or minor component of the version number changes (second digit, e.g., $1.4 \rightarrow 1.5$).

Third, we use several supervised machine-learning classifiers to label release notes, including XGBoost, support vector machines (SVM), XLNet, and BERT. These models are specified on the same training data used for the main article and use tuned hyperparameters from the associated replication materials (Nowak et al. 2025). In addition, we implement several large language model classifiers using the same inference and output-processing pipeline as in the replication package, with deterministic settings (temperature = 0; seed = 94032).

Across these approaches, some disagreement in the resulting innovation labels is expected, as different methods vary in how accurately they recover human-coded innovation judgments (Nowak et al. 2025). The relevant question for our purposes is therefore not whether all methods produce identical classifications, but whether the substantive downstream results are robust across measures with different underlying assumptions and measurement properties (Carlson and Burbano 2026).

Table L1 shows the overlap in positive innovation labels and the Jaccard similarity between the GPT-4.1 Mini fine-tuned classifier from the main article and the alternative approaches to construct the innovation variable. Agreement is highest for the other fine-tuned LLMs and GPT-5.2, followed by the supervised ML classifiers, while the rule-based and version-number proxies diverge more strongly. However, the estimated effect on complementor innovation activity remains positive across all measures (see Figure 5 of the main article).

Table L1. Agreement of Alternative Innovation Measures with the Baseline LLM Classifier

Approach	Share Innovative Updates	Overlap (Same Positive Labels)	Agreement (Jaccard Similarity)
GPT-4.1 Mini (fine-tuned ¹)	.478		
GPT-4.1 Mini (fine-tuned ²)	.491	.970	.919
GPT-4.1 Mini (fine-tuned ³)	.498	.970	.906
GPT-4o Mini (fine-tuned ¹)	.474	.954	.920
GPT-5.2 (default)	.447	.878	.831
XLNet	.520	.938	.817
BERT	.477	.892	.808
XGBoost	.502	.844	.701
SVM	.502	.844	.700
Dictionary/Rule-Based (AK23)	.791	.883	.498
Dictionary/Rule-Based (KF24)	.340	.591	.528
Version Number (1.1)	.441	.483	.336
Version Number (1.0)	.072	.088	.083

Notes: Overlap is the share of release notes labelled as innovative by the baseline that the alternative approach also labels as innovative. Agreement (Jaccard similarity) is defined as the number of updates classified as innovative by both measures divided by the number classified as innovative by at least one of the two measures. Fine-tuned models were trained on a human-labeled training set: ¹ n = 2,000 with representative class distribution (main variable), ² n = 2,000 with balanced class distribution, and ³ n = 1,000 with representative class distribution. For all LLM-based classifications, temperature was set to 0 and the seed to 94032. The dictionary/rule-based classifiers replicate the methods proposed by (AK23) Agarwal and Kapoor (2023) and (KF24) Kircher and Foerderer (2024).

L2 Classification Errors over Time

A concern with the LLM-based measure is that classification errors could change after the announcement, for instance if release notes after the announcement describe changes in ways that the classifier reads less accurately. Errors that are more frequent in one period or group could then affect the difference-in-differences estimates. We assess this with the validation holdout of 1,000 human-coded release notes, of which 523 fall into our observation period. Table L2 reports the accuracy of the classification by policy phase (Panel A). Accuracy is between .92 and .94 in all phases and the error rates before and after the announcement do not differ (17 of 218 and 21 of 305 notes, $p = .82$). The errors have the same structure in both periods. Most are maintenance updates described as general improvements that the model classifies as feature updates. The share of such false positives is slightly lower after the announcement.

For the 252 notes whose apps belong to the affected or unaffected group, Panel B reports the error rates by group and period. The cells are small and the differences between groups are not statistically significant ($p = .11$ for false positives). Still, their direction is informative. After the announcement, false positives decline for affected apps and increase for unaffected apps. If anything, classification errors would therefore reduce the estimated effect, not inflate it.

Table L2. Classification Accuracy by Policy Phase and Group

	Notes	Feat. share, Truth	Feat. share, LLM	Acc.	F1	FN Rate	FP Rate
<i>Panel A. By policy phase</i>							
Before announcement	218	.404	.454	.922	.909	.034	.108
Announcement phase	186	.495	.527	.935	.937	.033	.096
After implementation	119	.353	.378	.924	.897	.071	.078
<i>Panel B. By group and period</i>							
Affected, before announcement	81	.370	.420	.926	.906	.033	.098
Affected, after announcement	77	.364	.364	.974	.964	.036	.020
Unaffected, before announcement	55	.473	.509	.927	.926	.038	.103
Unaffected, after announcement	39	.564	.615	.897	.913	.045	.176
<i>Reference</i>							
Full holdout	1,000	.438	.464	.922	.914	.059	.093

Notes: The validation holdout comprises 1,000 randomly selected release notes with human-coded labels, of which 523 fall into the observation period (June 2019 to September 2021, excluding June 2020). Panel B uses the 252 of these notes whose apps belong to the affected or unaffected group. The false negative rate is the share of feature updates classified as maintenance updates, the false positive rate the share of maintenance updates classified as feature updates. Classification by the fine-tuned GPT-4.1 Mini model (temperature = 0, seed = 94032).

References

- Agarwal S, Kapoor R (2023) Value Creation Tradeoff in Business Ecosystems: Leveraging Complementarities While Managing Interdependencies. *Organization Science* 34(3):1216–1242.
- Carlson NA, Burbano V (2026) The use of LLMs to annotate data in management research: Foundational guidelines and warnings. *Strategic Management Journal* 47(3):699–725.
- Clover J (2020) iOS 14 Poised to Surpass iOS 13 Installations as Adoption Nears 50%. Retrieved (October 14, 2026), <https://www.macrumors.com/2020/10/27/ios-14-adoption-nears-50-percent/>.
- Correia S, Guimarães P, Zylkin T (2020) Fast Poisson estimation with high-dimensional fixed effects. *The Stata Journal* 20(1):95–115.
- Iacus SM, King G, Porro G (2012) Causal Inference without Balance Checking: Coarsened Exact Matching. *Political Analysis* 20(1):1–24.
- Kesler R, Kraft L, Skiera B (2026) Existence, Antecedents and Consequences of Non-Compliance in Mobile App Markets. <https://papers.ssrn.com/abstract=4871225>.
- Kircher T, Foerderer J (2024) Ban Targeted Advertising? An Empirical Investigation of the Consequences for App Development. *Management Science* 70(2):1070–1092.
- Kollnig K, Shuba A, Kleek MV, Binns R, Shadbolt N (2022) Goodbye Tracking? Impact of iOS App Tracking Transparency and Privacy Labels. *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22)*, 508–520.
- Kraft L, Skiera B, Koschella T (2024) Economic Impact of Opt-in versus Opt-out Requirements for Personal Data Usage: The Case of Apple's App Tracking Transparency (ATT). https://papers.ssrn.com/abstract_id=4598472.
- Levy D (2020) Speaking Up for Small Businesses. Retrieved (October 14, 2026), <https://about.fb.com/news/2020/12/speaking-up-for-small-businesses/>.
- Nowak R, Figge P, Haeussler C (2025) AI-Based Measurement of Innovation: Mapping Expert Insight into Large Language Model Applications. <https://arxiv.org/abs/2508.02430>.
- Wen W, Zhu F (2019) Threat of platform-owner entry and complementor responses: Evidence from the mobile app market. *Strategic Management Journal* 40(9):1336–1367.