Data Capital and Innovation: Evidence from Connected Devices*

Tarikua Erda† Ryan Gilland‡ Ben Lahey§

March 26, 2026

[Latest version]

Abstract

Consumer data is a valuable form of intangible capital and an input in the innovation process but goes unmeasured in standard firm-level datasets. Applying natural language processing to the user manual text of the universe of wireless devices sold in the U.S., we identify and study data capability from connected "smart" devices that collect and transmit user data via apps and cloud servers. Prior commercialization of these data-capable devices is associated with prolific subsequent product introductions and, among startups, a higher likelihood of successful exits. Data-capable public firms are large, highly R&D-intensive, and have dominated patenting in machine learning and AI—domains where data is a core input. Using a triple-difference design, consumer-facing data-capable firms exposed to restrictions on personal data transmission under the European General Data Protection Regulation saw substantial patenting declines particularly in those data-intensive areas, highlighting that first-party data streams from connected devices are a key input into the innovation pipeline. Our approach offers a new, scalable measure of data capital and underscores its implications for firm performance and innovation in the AI-era.

Keywords: Data Economy; Intangible Capital; Innovation; R&D; Internet of Things; AI; Corporate Finance
JEL Codes: E22; G30; O34; L86; K20

*We thank Anton Babrov, Olalekan Bello, Michael Ewens, Xavier Giroud, Thomas Philippon, Robert Seamans, and James Traina for helpful discussions. Support from the Alfred P Sloan Foundation is gratefully acknowledged. All authors have no financial or other ties related to the research project. All errors are ours.

†Erda: New York University, tarikua.erda@nyu.edu.
‡Gilland: New York University, rg4795@nyu.edu.
§Lahey: New York University, bpl9631@nyu.edu.

1 Introduction

The ability to collect, transmit, and analyze user data has become a central form of investment for modern firms. User data shapes product improvement, differentiation, and pricing, and underpins the development of AI and prediction technologies, with venture capital firms pointing to the quality and uniqueness of startups' proprietary data as their biggest competitive edge (Szkutak, 2025). Data has been recognized in the academic literature as a core intangible asset and modeled as a driver of firm dynamics (Farboodi et al., 2019), merger activity (de Cornière & Taylor, 2024), the macroeconomy (Farboodi & Veldkamp, 2025), and allocative efficiency under different property rights regimes (Jones & Tonetti, 2020). Yet, we have relatively little empirical evidence on how exactly data factors into firm performance and innovation, which is the subject of this paper.

Quantifying the value of data capability as an intangible asset has been difficult, primarily due to measurement challenges: standard proxies for intangible investment—accounting intangibles, patents, job postings, or web-based measures—do not directly observe firms' data-collection capability. Existing "gold standard" firm surveys from the US Census Bureau can tell us whether firms report using data, cloud computing, or analytics, but they remain coarse, self-reported, and infrequent.[1]

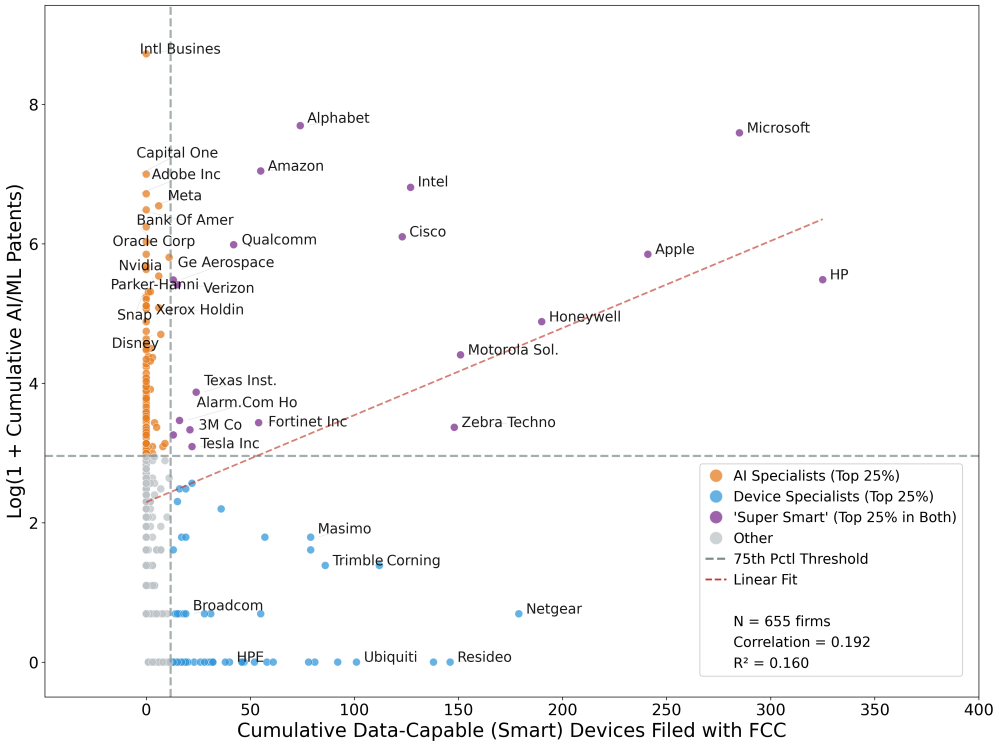
In this paper, we leverage an untapped administrative dataset on wireless products and use connected devices—"smart" products that can connect to apps or the cloud and continuously generate, process, and transmit user data—to develop a novel measure of firms' data capability. We use this measure to document new facts about the role of data in firm performance and innovation. We also exploit a data collection cost shock created by the European Union's General Data Privacy Regulation ("GDPR"), which restricted key forms of personal data transmission and processing for firms with customers in the European Economic Area. Data-capable firms selling consumer-facing smart devices experience significant declines in AI and machine learning patenting relative to those selling industrial smart products, with suggestive evidence of revenue declines as well. That innovation falls in the domains most dependent on first-party data suggests that data functions as an accumulated, depreciable intangible capital—one that serves as a productive input into the innovation pipeline.

Connected devices make a useful setting to study data capital for two reasons. First, the booming growth in these devices defines the modern, Internet of Things ("IoT") era of consumer electronics. The enterprise IoT market stood at $269 billion in 2023, with recent

[1]For example, the Annual Survey of Manufactures (ASM) and the Management and Organizational Practices Survey (MOPS) include valuable questions on IT and management practices, but they do not measure product-level data capability, are only available for sampled firms/years.

estimates citing more than 20 billion connected devices globally and more than 120 zettabytes of global data creation/replication that is projected to grow further in the rest of this decade.[2] Second, the economics of device-generated data differ from the web-tracking setting that dominates much of the existing privacy literature.[3] Such measures of data intensity built on third-party tracking are intertwined with browsers, consent banners, and platform rules. In contrast, as part of ordinary product use, connected devices generate first-party data—data that is proprietary, tied to an installed base, and able to be reused to improve products, refine predictive analytics, and develop complementary products. Measuring the value of data capability embodied in connected devices is therefore a natural and timely step toward understanding data as intangible capital in the AI era. Figure 1 previews this link using our data on smart devices and patent records: there is a positive association between the number of data-capable devices a firm offers and its patenting in AI and machine learning, with a few 'super smart' firms highly active in both domains clustered in the upper right.

Figure 1: Smart Devices and Patenting in AI/Machine Learning Domains



For intuition, consider a robot vacuum. The vacuum generates a steady stream of usage and performance data—cleaning patterns, navigation maps, error logs, battery wear—often routed through a companion app and cloud service. The firm can use these data to improve

[2]See: BusinessWire.com (2021); IOT-Analytics.com (2025)
[3]For instance, see Aridor et al. (2023); Liu (2024).

navigation and reliability, inform product innovation, support add-on services such as software features, replacement parts, and more. The installed, tangible hardware generating this intangible data stream thus enables value creation beyond the initial sale.

We observe connected devices sold in US markets using the mandatory device authorization filings from the Federal Communications Commission ("FCC"), which are required for virtually any wireless device sold in the United States. These filings provide a census of all radio frequency-emitting devices, along with rich product documentation including, notably, user manuals that describe device connectivity, companion applications, and cloud interactions. They are uniquely useful for our purposes because they provide revealed, engineering-focused, product-level capabilities at the point of commercialization. They are thus free from selection concerns that may arise from voluntary self-reporting contexts where firm strategic behavior may matter, and they offer a more reliable paper-trail to track product innovation than patents, whose contents may not make it to market.

To identify devices with data-collection capability, we apply natural language processing ("NLP") methods to user manual texts, focusing on features such as companion-app connectivity and cloud data transmission. Aggregating across devices, we construct a firm-year measure of data-capability that captures whether a firm owns an FCC certification for a data-capable device (and how many).[4] We primarily focus on devices filed between 2007, when smartphones and apps were introduced, and 2019, which is the last year of data where we have detailed and comprehensive data on the FCC website.

We document that data-capability is associated with more prolific new product filings in subsequent years. FCC grantees that have previously commercialized data-capable devices are more likely to file new products in a given year and introduce greater numbers of new devices when they do, consistent with data-enabled products anchoring follow-on innovation.

We then use Compustat and Pitchbook to characterize public and private firms selling wireless devices, finding that data capability has increasingly been concentrated among larger, R&D-intensive and innovative firms over the course of the IoT boom. Public firms that sell data-capable products are generally large in assets and higher valued. They also spend much more on R&D relative to non-data-capable public firms, including in sector-adjusted terms. They have come to dominate patenting activity in AI and machine learning technologies, where data is a critical input. Among early-stage private firms, bringing a wifi-enabled device or a device with a companion app to market was associated with a higher likelihoods for successful exits, particularly IPOs. However, this advantage waned starting

[4]We validate this measure by showing that firms we classify as data-capable are substantially more likely to file deal narrative texts with the SEC that mention collection and usage of data.

in the mid-2010s, concurrent with the broader diffusion of IoT technologies. These patterns, which hold after sector adjustment, indicate that data capability is closely tied to innovative output, R&D strategy, and market position within industries, not merely across them.

Finally, we use a triple-difference design to study the impact of the GDPR's privacy regulation shock that increased firms' compliance costs for using first-party data generated from devices. In particular, we examine how the outcomes of data-collecting firms with exposure to European customers change after GDPR is adopted in 2016 relative to non-data-capable firms with EU exposure. This approach helps to isolate the GDPR's impact from confounders like broader differences in firm technology, managerial quality, and growth opportunities.

We show that EU-exposed data-capable consumer firms experience a 36.5% decline in AI/ML patent filings and an 89% decline in citation-weighted AI/ML patents following GDPR adoption. Firms making smart devices for industrial use—whose data collection falls largely outside GDPR's personal-data scope—show no such decline and if anything increase their AI/ML patenting, consistent with GDPR selectively disrupting innovation that depends on personal consumer data. Financial outcomes are directionally negative for consumer firms but imprecisely estimated; we are expanding sample coverage in ongoing work. Together with the descriptive evidence, these results suggest that first-party data serves as a productive input into innovation: when regulation restricts its collection and reuse, patenting contracts in the domains most reliant on data.

This paper makes several contributions. First, we provide novel evidence on the value of data capital for firm outcomes in the context of connected devices, which constitute a substantial share of data generation in the current digital age. Previous work had documented sales declines of high data-intensity firms after GDPR (see Quan (2024); Liu (2024)) in the context of website data. We extend this to first-party data collection via connected devices. Our finding that AI/ML patenting declines sharply among EU-exposed data-capable consumer firms adds to a small but growing body of evidence that privacy regulation affects the direction and pace of innovation (see Blind et al. (2024); Giorgio et al. (2024)). Our contribution is to identify this effect at the firm level using product-based measures of data capability, and to show that the decline is concentrated among consumer-facing firms whose data collection falls squarely within GDPR's personal-data scope. The consumer/non-consumer contrast—where industrial device makers are unaffected or even benefit—strengthens the case that the mechanism runs through personal data specifically, not through broader regulatory burden. In this way, our paper also provides empirical evidence consistent with data being a non-rival, growth-enhancing asset as modeled in previous theoretical work (see Jones

and Tonetti (2020); Farboodi and Veldkamp (2025)).

Second, our work contributes to a growing literature on intangible capital (Crouzet & Eberly, 2023; Crouzet et al., 2022) by providing a direct, product-level measure of one of its least-observed components—data. Existing empirical proxies for firm data intensity typically rely on disclosures like 10-K filings (Boroomand et al., 2022), online tracking measures (Goldberg et al., 2024; Aridor et al., 2023), or IT and data-processing activity from cloud usage (Demirer et al., 2024), or hiring and data-related patents (Quan, 2024). While informative, this prior work leaves unobserved firms' first-party data generation outside of websites. Our use of FCC authorization filings allows us to contribute a new and comprehensive measure of data capability that is embodied in physical hardware products. Because FCC authorization occurs at or near commercialization and is produced for engineering compliance rather than investor communication, the resulting measure captures deployable data-collection infrastructure with a measurement error that is plausibly orthogonal to firms' strategic disclosure incentives.

In this respect, methodologically, we also add to the literature using NLP to convert unstructured text into structured economic statistics to understand the economy, especially critical areas like artificial intelligence (Babina, Fedyk, He, & Hodson, 2024) and fundamental growth-drivers like innovation (Bergeaud, Jaffe, & Papanikolaou, 2025). Beyond allowing a new measure of data capability in connected devices, our use of user manual text opens new avenues for the study of product innovation and technology in the AI era.

2 Data and Measurement

2.1 Data

FCC Equipment Authorization Filings

Any wireless device marketed or imported into the United States must obtain authorization from the Federal Communications Commission (FCC) before entering commerce. This requirement applies broadly to equipment that emits radio-frequency (RF) energy, including consumer electronics, industrial sensors, medical devices, and Internet-of-Things (IoT) products. Device authorization requires submission of an application—Form 731—identifying the legally responsible party (the "grantee"), the device's technical characteristics, applicable rule parts, and the authorization date. The grantee need not be the physical manufacturer; it may be a brand owner, importer, or distributor, provided it assumes legal responsibility

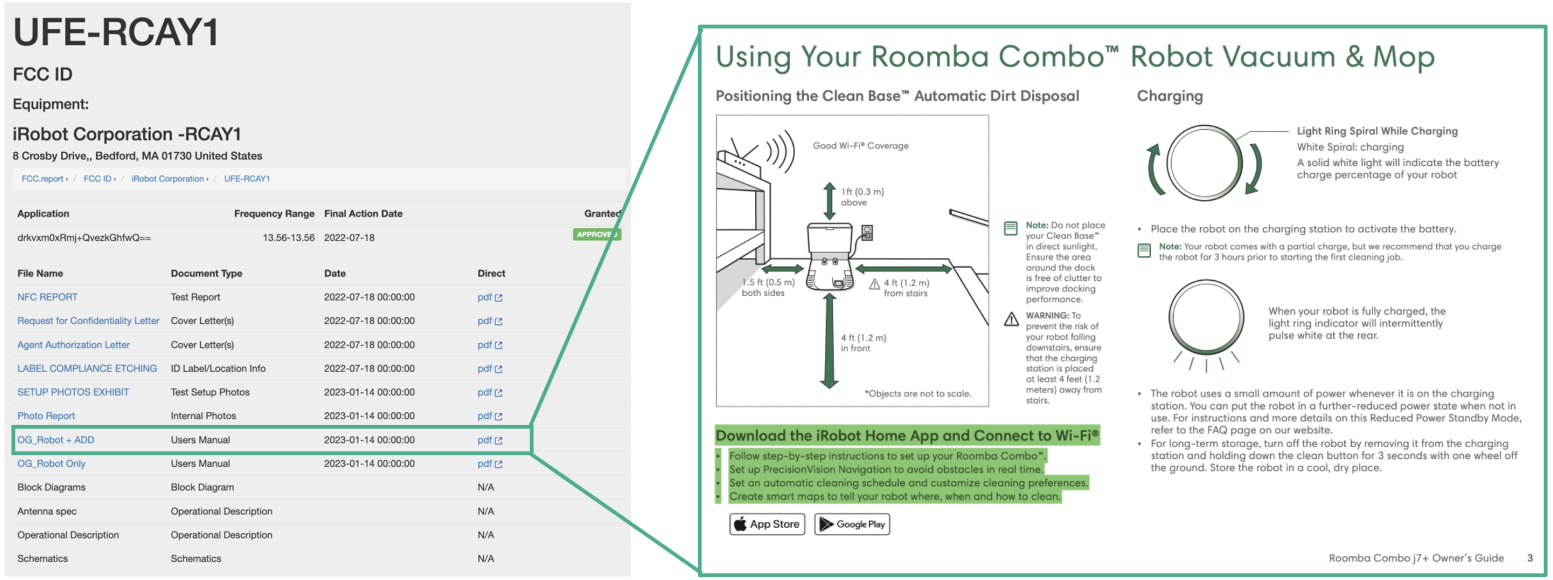
for compliance. For foreign manufacturers, FCC rules require a U.S.-based agent for service of process, ensuring each authorized device is linked to a U.S.-jurisdictional entity. Because authorization precedes market access, these filings are mandatory and independent of firms' financial reporting considerations, performance, or strategic disclosure choices.

We build a device-level FCC dataset by scraping the metadata and application documentation on user manuals associated with device filings.[5] These filings allow us to observe the technical exhibits submitted during the approval process. They also include information on rule parts, frequency bands, and transmission technologies. These materials allow us to infer economically meaningful device features—such as whether a product supports persistent connectivity, data transmission, or companion mobile applications—that distinguish simple RF devices (e.g., garage door openers) from data-capable connected products (e.g., smart wearables and home appliances).

As an illustrative example, Figure 2 shows the FCC equipment authorization records for the Roomba—an autonomous robotic vacuum device—from fcc.report. The top left panel shows an iRobot filing in the FCC Equipment Authorization System, including the exhibit list that typically contains test reports and publicly posted user manuals. The top right panel zooms in on a page from the Roomba's user guide that instructs users to download a companion smartphone app and connect the device to Wi-Fi. We use NLP to detect app-companion and connectivity language, systematically identifying connected devices that likely transmit user or sensor data.

[5]These were accessed from https://fccid.io/ and https://fcc.report/.

Figure 2: FCC record example for the Roomba (FCC ID UFE-RCAY1)



UFE-RCAY1

FCC ID

Equipment:

iRobot Corporation -RCAY1

8 Crosby Drive,, Bedford, MA 01730 United States

FCC.report › / FCC ID › / iRobot Corporation › / UFE-RCAY1

Application	Frequency Range	Final Action Date	Grantee
drkvxm0xRmj+QvezkGhfwQ==	13.56-13.56	2022-07-18	APPROVE

File Name	Document Type	Date	Direct
NFC REPORT	Test Report	2022-07-18 00:00:00	pdf
Request for Confidentiality Letter	Cover Letter(s)	2022-07-18 00:00:00	pdf
Agent Authorization Letter	Cover Letter(s)	2022-07-18 00:00:00	pdf
LABEL COMPLIANCE ETCHING	ID Label/Location Info	2022-07-18 00:00:00	pdf
SETUP PHOTOS EXHIBIT	Test Setup Photos	2023-01-14 00:00:00	pdf
Photo Report	Internal Photos	2023-01-14 00:00:00	pdf
OG_Robot + ADD	Users Manual	2023-01-14 00:00:00	pdf
OG_Robot Only	Users Manual	2023-01-14 00:00:00	pdf
Block Diagrams	Block Diagram		N/A
Antenna spec	Operational Description		N/A
Operational Description	Operational Description		N/A
Schematics	Schematics		N/A

Using Your Roomba Combo™ Robot Vacuum & Mop

Positioning the Clean Base™ Automatic Dirt Disposal

Good Wi-Fi® Coverage

1 ft (0.3 m) above

1.5 ft (0.5 m) both sides

4 ft (1.2 m) from stairs

4 ft (1.2 m) in front

*Objects are not to scale.

Note: Do not place your Clean Base™ in direct sunlight. Ensure the area around the dock is free of clutter to improve docking performance.

WARNING: To prevent the risk of your robot falling downstairs, ensure that the charging station is placed at least 4 feet (1.2 meters) away from stairs.

Download the iRobot Home App and Connect to Wi-Fi®

- Follow step-by-step instructions to set up your Roomba Combo™.
- Set up PrecisionVision Navigation to avoid obstacles in real time.
- Set an automatic cleaning schedule and customize cleaning preferences.
- Create smart maps to tell your robot where, when and how to clean.

App Store Google Play

Charging

Light Ring Spiral While Charging

White Spiral: charging

A solid white light will indicate the battery charge percentage of your robot

- Place the robot on the charging station to activate the battery.

Note: Your robot comes with a partial charge, but we recommend that you charge the robot for 3 hours prior to starting the first cleaning job.

When your robot is fully charged, the light ring indicator will intermittently pulse white at the rear.

- The robot uses a small amount of power whenever it is on the charging station. You can put the robot in a further-reduced power state when not in use. For instructions and more details on this Reduced Power Standby Mode, refer to the FAQ page on our website.
- For long-term storage, turn off the robot by removing it from the charging station and holding down the clean button for 3 seconds with one wheel off the ground. Store the robot in a cool, dry place.

Roomba Combo j7+ Owner's Guide 3

Application Details:

Equipment	Automated robot vacuum
FRN	0015256720
Grantee Code	UFE
Product Code	-RCAY1
Applicant Business	iRobot Corporation
Business Address	8 Crosby Drive, , Bedford,Massachusetts United States 01730
TCB Scope	A1: Low Power Transmitters below 1 GHz (except Spread Spectrum), Unintentional Radiators, EAS (Part 11) & Consumer ISM devices
TCB Email	LST.FRE.CertNotice@ul.com
Applicant	Stephen Pallotta – Principal Compliance Engineer
Applicant Phone	
Applicant Fax	781-430-3001
Applicant Email	spallotta@irobot.com
Applicant MailStop	M/S 12-1
Test Firm	UL LLC
Test Firm Contact	Steven Thompson
Test Firm Phone	
Test Firm Fax	
Test Firm Email	steven.d.thompson@ul.com
Certified By	Stephen Pallotta –

Firm Data

For our analysis of the effects of connected devices on firm cash flows, we obtain regulatory financial filings from Compustat. As a majority of FCC filings are submitted by firms located outside of the United States, we target quarterly financial data for the entirety of firms in the Compustat universe, harmonizing quantities reported in historical local currencies to 2024 USD. We focus on a handful of core measures broadly related to firm performance and innovative output: revenue, R&D expenditures, R&D intensity (defined as R&D normalized by revenue), and market capitalization. We additionally compute Tobin's q for each firm, using the simplified computation discussed in (Chung & Pruitt, 1994). Given we are analyzing a sample of firms that must first be fuzzy-matched across multiple data sources, we prefer to build this measure in a manner that best preserves our linked sample.

We incorporate firm customer segment reporting to build a measure of exposure to markets subject to European Union regulation. SEC reporting requirements require firms to disclose revenues from individual customers that contribute more than 10 percent of consolidated revenue (Board, 2026). Disclosure requirements across other broad categories are more opaque. For instance, the Financial Accounting Standards Board's (FASB) Accounting Standards Codification (ASC) 280-10-55-51 states that public entities must report "revenues from external customers attributed to the public entity's country of domicile and attributed to all foreign countries in total from which the public entity derives revenues. If revenues from external customers attributed to an individual foreign country are material, those revenues shall be disclosed separately." Here "material" is not explicitly defined and thus geographic disclosures are somewhat at the discretion of management. Major accounting firms such as Deloitte and PricewaterhouseCooper take the perspective that, in-line with individual customers, firms should report countries which exceed a contribution of 10 percent to consolidated revenues or long-lived assets [6].

In examining the effects of connected device introduction on longer-run firm outcomes, we leverage deal data from PitchBook. We consider a variety of financing outcomes, in particular initial public offerings, strategic acquisitions, and private equity-led buyouts. While the PitchBook universe extends across public and private firms, complementary financial data for private firms is extremely limited.

[6] The perspectives of Deloitte and PricewaterhouseCooper can be viewed through these links.

Patent Filings

We augment our Compustat sample with patent grant data from the United States Patent and Trademark Office (USPTO). We use the December 7, 2025 revision of the Center for Research in Security Prices (CRSP) to USPTO patent grants crosswalk introduced in Kogan, Papanikolaou, Seru, and Stoffman (2017) as our baseline sample of patents. In this sample, we identify patents as being AI/ML-adjacent using granular USPTO Cooperative Patent Classification (CPC) codes assigned to patents.

In our analyses we emphasize the date a patent was first filed with the USPTO, rather than the date it was granted, to account for the lag in the time between filing and approval. Each patent's citation weight is computed as the ratio of its forward citations to the average forward citations received by all patents in the same CPC3 technology class and grant year, following the approach in Hall et al. (2005). We aggregate these weights to the firm-year level by summing across all patents filed by a given firm in a given year. This weighting adjusts for well-documented differences in citation propensities across technology fields and over time, ensuring that a patent in a sparsely cited class is not mechanically undervalued relative to one in a heavily cited class.

Narratives from SEC Acquisition Filings

For the purpose of validating our FCC-based measure of data capability, we use SEC filings–typically 10-K annual reports or 8-K current reports—which are a required element of acquisitions completed by publicly traded firms. These filings contain both structured purchase price allocations that value identifiable intangibles (e.g., technology, customer relationships, goodwill) and narrative text describing the deal's strategic rationale. We use the 627 acquisitions (2007–2017) compiled by Ewens et al. (2025), which extracts intangible-asset valuations from these filings. The narrative sections often characterize targets' assets in ways not captured by line items, including references to data capabilities, user engagement, monetization potential, and technology infrastructure.

2.2 Developing and Validating our Measure of Data-Capability

Approach for Identifying Data-Capable Devices: We first extract the text from the user manual PDF files with each device authorized by the FCC since 2007, and create OCRs for image-based user manuals. We then apply natural language processing to that text, using an algorithm that searches for textual patterns indicating companion-app connectivity and

cloud data transmission. We use rule-based pattern matching rather than machine learning classifiers because the target language is explicit and literal, and the resulting classification rules, which were refined using manual checks, are transparent and reproducible. Devices are classified as "data-capable" if their user manuals contain either (i) any explicit reference to companion apps, app stores, or brand-specific app names (e.g., "mobile app," "download from Google Play," "Ring app"), or (ii) at least two references to secondary indicators such as remote access functionality, push notifications, account creation, cloud storage, or QR-code-based setup. We additionally flag devices mentioning data synchronization, usage tracking, or server-based storage. Documents that appear to be test reports rather than user manuals are filtered out prior to classification. A more detailed summary on the NLP approach is provided in Appendix section A.1.

Appendix Figure A.1 displays the number of approved FCC filings by grantee country and state for filings concerning data-collecting devices over the 2007-2025 period. Of the 60,151 filings we are able to map to grantee country, 20,652 (34.3 percent) were granted to firms headquartered in the United States. The remaining top five most prevalent jurisdictions with data-collecting device approvals were China and Hong Kong (30.9 and 4.8 percent), Taiwan (7.2 percent), South Korea (4.0 percent), and Japan (2.5 percent). Just outside is Germany (2.1 percent), an EU member. Decomposing the filings granted to firms headquartered in the United States, the most prevalent states are California (29.8 percent of US filings), New Jersey (19.2 percent), Illinois (6.3 percent), Florida (5.8 percent), and New York (3.9 percent). Thus device filings are highly concentrated among a handful of countries, primarily the US and China, and within the US, California.

Validation: To validate our measure of data capability, we leverage a separate, external resource—the texts of purchase price allocations in Securities Exchange Commission (SEC) filings. In particular, we examine whether acquisitions involving companies with data-collecting devices exhibit different linguistic patterns in their SEC filings compared to acquisitions without such devices.

We proceed in three steps. First, we start with the Ewens et al. (2025) Purchase Price Allocation (PPA) dataset. We find a successful match to our FCC grantee database for 175 deals, where either the target or acquiring firm sold FCC-licensed devices. Of these, 31 deals specifically linked to target firms.[7] Next, for each acquisition, we extract text excerpts from the corresponding SEC filings that mention "data" or related terms in the context of business combinations, purchase price allocations, and acquisition rationale. We then use

[7]It is not surprising that we obtained a relatively much smaller set of matches to target firms—the SEC data primarily capture larger public firms while acquisition targets are likely to be smaller, private companies.

NLP to identify filings mentioning keywords related to data strategy and use, including in the areas of user engagement and retention metrics, IoT and connected devices (e.g., wearables like Fitbits). Lastly, we test for statistical significance using Fisher's exact test for a binary indicator of *any* keyword presence.

Table 1: FCC grantees classified as likely data-collectors are substantially more likely to be associated with data use/data strategy related keywords in SEC acquisition filings

	Target-Only (N=31)			**All Deals (N=175)**		
	Has Keyword	Total	%	Has Keyword	Total	%
Data-device deals	7	13	53.8%	45	118	38.1%
No-data-device deals	2	18	11.1%	11	57	19.3%
Difference			+42.7pp			+18.8pp
Odds Ratio		9.33			2.58	
p-value (Fisher's exact)		0.0166**			0.0152**	

Notes: Binary indicator equals 1 if any data-strategy keyword appears in SEC filing excerpts. Sample restricted to high-quality FCC-SEC matches (score ≥ 0.9).
*** p<0.01, ** p<0.05, * p<0.10

Our findings in Table 1 reveal that acquisitions involving data-device companies exhibit significantly higher mentions of data-related strategy keywords across multiple categories. Strikingly, a binary indicator for *any* keyword presence shows that 53.8% of data-capable targets have at least one data-strategy keyword mentioned, compared to only 11.1% of non-data capable targets (odds ratio = 9.33, p=0.0166). Looking at all deals whether either the target or acquirer was a data-capable company similarly shows a large increase in the likelihood of data-use or data-strategy mentions in the deal narrative—mentions that, according to the detailed breakdown of keyword types in Appendix Table A.1—appear to be primarily associated with engagement/retention terms, IoT terminology, and data collection language. When restricting to the smaller set of deals where the *target* (rather than acquirer) is the data-device company, the IoT category still remains significant. These findings suggest that FCC data-device classifications capture meaningful variation in how acquirers discuss data assets and strategy in their SEC disclosures, and thus lend credence that our measurement of the data capability of firms is reasonable.

2.3 Sample Construction

Our data on FCC device filings contains a unique identifier associated with each grantee, but this identifier does not map to any relevant firm data. We thus use a matching algorithm to map firms recorded in FCC device filings to those in the CapitalIQ/Compustat and

PitchBook databases. Details of the matching algorithm are described in the appendix. Overall we match 5,054 (16.9 percent) of the uniquely identified firms in the FCC device filings to those in Compustat, and 10,219 (34.1 percent) to those in PitchBook. The disparity in match rates between Compustat and PitchBook is unsurprising, as any firm can file with the FCC in principle. The firms captured in our Compustat match are largely current or former publicly-listed entities and their subsidiaries, while PitchBook's scope is far broader and aims to capture the universe of "investable" firms.

We then use the latest CRSP to Compustat firm crosswalk supplied by CRSP to map patents to Compustat firms over time. We identify the patent assignee firm in our FCC-Compustat link for 1,139,053 (33.3 percent) of the patents in our initial sample, of which 9,820 are classified as related to Artificial Intelligence ("AI") or machine learning ("ML"). We then build measures of patenting activity both broadly and within AI/ML, as well as weighted measures using the forward citations provided by (Kogan et al., 2017).

Lastly, we use reported customer segments to identify firms with exposure to the European Union. While there is expected to be some reporting bias in firm geographic segments given the lack of a hard reporting requirement, the loose "material" requirement implies the firms that do report splits do so because these geographies represent material concerns. We therefore define firms as being exposed to European Union markets (and thus European Union regulations) based on whether that firm reports revenue or asset exposures to a region or country situated in the European Union. For the purposes of our later analyses, we use sales reported as of the first quarter of 2016 (pre-GDPR adoption) to generate our indicator. Ultimately we identify 66.1 percent of the FCC-Compustat linked firms as being potentially exposed to GDPR.

3 Characterizing Data-Capable Devices and Firms

3.1 Data-Capable Devices & Product Innovation

Table 2 examines whether firms that commercialize data-capable devices subsequently exhibit different product-introduction patterns. We identify "switcher" firms—grantees that had filed non-data products before their first data-device authorization—and require a minimum gap of three years between each firm's first non-data FCC filing and its first data-capable device filing, ensuring sufficient pre-adoption history to assess parallel trends in the event-study design. To recover firms that adopted data-capable devices in the earliest cohorts (2007–2009), we extend the FCC filing history back to 2004; because connected de-

vices with companion apps and cloud connectivity were essentially nonexistent before the smartphone era, all pre-2007 filings are classified as non-data devices. These criteria yield 1,943 switcher firms. We construct a balanced panel for these firms from 2007 to 2018, producing 21,452 firm-year observations. Using a staggered adoption design, we compare each firm's post-switch filing activity (i.e., after it commercialized its first data-capable device) to its own pre-switch baseline, excluding the year of the switch itself to avoid mechanical effects. We control for the number of cumulative prior products and grantee age, both measured in the year prior to the switch. We report both standard Two-Way Fixed Effects and interaction-weighted estimates following Sun and Abraham (2021) to address potential bias from heterogeneous treatment timing.

Overall, the patterns in Table 2 point to an increase in both the probability and intensity of new product filing episodes. Panel A shows that data-device adoption raises the likelihood of filing any new product in a given year by 8.0 percentage points (interaction-weighted estimate), against a baseline probability of 38.0%. Panel B shows that adopters file approximately 0.56 more new products per firm-year, a 40% increase relative to the baseline mean of 1.40. Panel C suggests that continued data-device filings account for a meaningful share of this increase.[8] Given that the decision to develop a data-capable device is itself endogenous to a firm's product strategy, these estimates should be interpreted as descriptive associations rather than causal effects. Still, the findings are consistent with platform-style development: an initial data-enabled product anchors a stream of follow-on filings, raising both the probability and the volume of innovation within a firm's active development periods.

3.2 Data-Capability among Public Firms

Figure 3 shows various summary statistics comparing public firms with at least one data-capable device against 'non-capable' firms with none.

Panel (a) shows that the number of data-capable public firms has grown steadily, overtaking non-capable firms around 2015 and reaching nearly twice as many by 2023 (roughly 270 vs. 145). This is consistent with a broad shift toward connected products since the proliferation of smartphones and app-enabled devices. Data-capable firms are substantially larger and more highly valued: their median assets and market value exceed those of non-capable firms, with the gap widening markedly from the early 2010s onward (Panels b and c). By 2023, median assets for data-capable firms are roughly $4.5 billion, compared with $1.5 billion for non-capable firms; median market values diverge similarly ($6 billion vs. $2.5

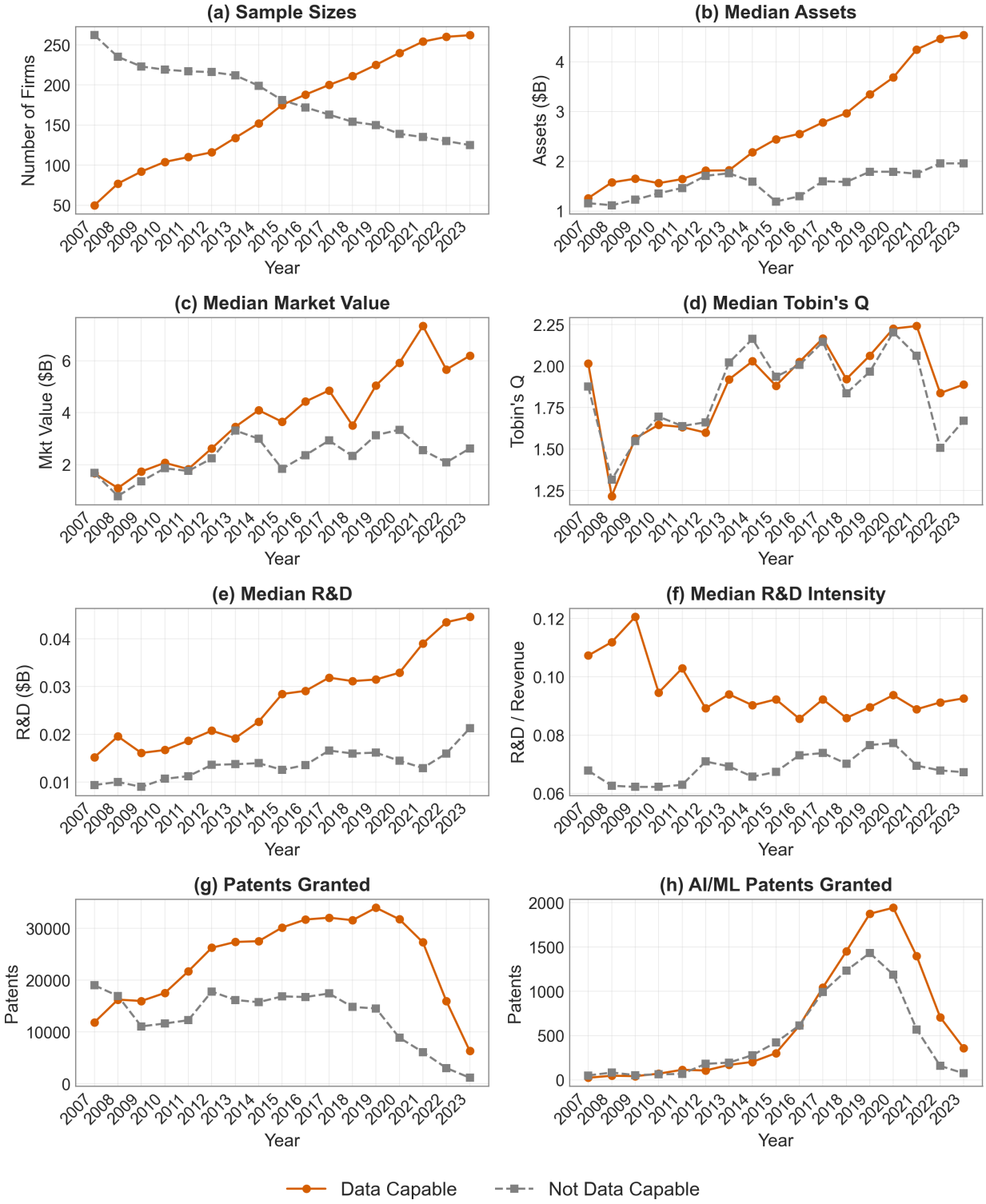
[8]The Sun and Abraham (2021) interaction-weighted estimates are roughly half the TWFE coefficients, consistent with upward bias in standard TWFE under staggered adoption.

Table 2: Data-Capable Device Launches and Subsequent Product Filings

	(1)	(2)	(3)	(4)	(5)
Panel A: Any New Product Filing					
Post-switch (TWFE)	0.274***	0.094***	0.110***	0.108***	0.104***
	(0.011)	(0.013)	(0.016)	(0.016)	(0.016)
Post-switch (S&A)	—	0.065***	0.085***	0.080***	0.080***
		(0.018)	(0.022)	(0.023)	(0.023)
Baseline Mean	0.380	0.380	0.380	0.380	0.380
Baseline SD	0.485	0.485	0.485	0.485	0.485
Panel B: N New Product Filings					
Post-switch (TWFE)	1.836***	0.723***	0.860***	0.818***	0.796***
	(0.100)	(0.082)	(0.109)	(0.107)	(0.106)
Post-switch (S&A)	—	0.503***	0.523***	0.558***	0.561***
		(0.107)	(0.119)	(0.129)	(0.133)
Baseline Mean	1.404	1.404	1.404	1.404	1.404
Baseline SD	3.408	3.408	3.408	3.408	3.408
Panel C: N New Data Device Filings					
Post-switch (TWFE)	0.826***	0.583***	0.796***	0.779***	0.766***
	(0.043)	(0.035)	(0.054)	(0.054)	(0.053)
Post-switch (S&A)	—	0.257***	0.274***	0.269***	0.261***
		(0.003)	(0.059)	(0.060)	(0.060)
Baseline Mean	0.436	0.436	0.436	0.436	0.436
Baseline SD	1.719	1.719	1.719	1.719	1.719
Observations	21,452	21,452	21,452	21,452	21,452
Grantee FE	No	Yes	Yes	Yes	Yes
Country × Year FE	No	No	Yes	Yes	Yes
Equipment × Year FE	No	No	No	Yes	Yes
Product Fn × Year FE	No	No	No	No	Yes

Notes: Sample restricted to switcher firms with at least three years of non-data FCC filings before their first data-device filing. Filing history extends to 2004 using archived FCC device records; all pre-2007 filings are treated as non-data devices, as connected products with companion apps were essentially nonexistent before the smartphone era. The adoption year (t=0) is excluded to avoid mechanical effects. $= 1$ for years ≥ 1 after first data-device filing. S&A IW reports the average interaction-weighted post-treatment estimate (t=1 to t=4) following Sun and Abraham (2021), using the last-treated cohort as control. Controls: firm age at t= -1 and its square, IHS(prior products at t= -1). Standard errors clustered at the grantee level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure 3: Comparing data capable versus non-capable public firms (2007-2023)



billion). Median Tobin's Q tracks relatively closely across the two groups for much of the sample period, though data-capable firms carry a modest premium in the early and late years of the panel (Panel d). These patterns generally hold in sector-adjusted comparisons as well (see Appendix Figure A.2).

Notably, data-capable firms also appear more innovative. Panels (e) and (f) show that data-capable firms spend more on R&D both in absolute terms and as a share of revenue, consistent with data from connected devices serving as an input into product improvement and development. The gap in R&D spending widens over time, while R&D intensity remains persistently higher for data-capable firms throughout the sample. Strikingly, data-capable firms have come to dominate patenting activity. Panel (g) shows that total patents granted to data-capable firms surpass those of non-capable firms by the mid-2010s and pull away sharply. The pattern is even more pronounced for AI and machine-learning patents (Panel h), where data-capable firms account for the vast majority of activity by the late 2010s. This is consistent with data capability and AI innovation being complements: firms that generate first-party data through their connected devices are disproportionately the ones investing in the prediction technologies that use it.

As shown in Figure 1 earlier in the paper, "smart device" innovation is highly concentrated in a small set of large, technologically sophisticated firms: companies that introduce many data-capable FCC devices also tend to be the ones with the deepest AI/ML patent portfolios. In general, we document a positive correlation (0.19) between firms' data-capable device filings and their AI/ML patent output, suggesting that hardware-enabled data collection and artificial intelligence innovation are complementary strategies. Using 75th-percentile thresholds to classify firms, we identify four distinct profiles: "AI Specialists" like Meta and Nvidia that lead in AI and ML patents but with few direct device filings; "Device Specialists" like Ubiquity and Netgear that focus primarily on hardware and smart electronics; "Super Smart" firms like Microsoft and Honeywell that are in the top quartile of both smart device innovation and AI and machine learning patenting; and a long tail of firms below median on both dimensions.

Taken together, these findings indicate that data capability has become increasingly characteristic of larger, more R&D-intensive, and more innovative public firms. Because these differences persist after sector adjustment, they reflect real distinctions between data-capable and non-capable firms within similar industries, rather than merely across-sector variation in size and performance.

3.3 Data-Capability among Early-Stage Firms

We estimate a discrete hazard model from Equation 1 for several outcomes for US Pitchbook firms with respect to device type filings. Let $Y_{it} \in \{0, 1\}$ where Y represents an outcome from {bankruptcy, buyout, merger/acquisition, IPO} and $Y_{it} = 1 \implies Y_{i,t+j}$ are censored for all $j > 0$. N_{it}^X is the number of devices of type X the firm i has filed by t and N_{it} is the firm's count of devices of all types, $X \in$ {data-collecting, wifi-enabled, has companion app}.

$$\Pr(Y_{it} = 1 \mid active) = \beta_1 \mathbb{1}\{N_{it}^X > 0\} + \beta_2 \log(1 + N_{it}) + C\gamma + \lambda_{s(i)\times t} + \lambda_{g(i)} + \epsilon_{it} \quad (1)$$

The control term C contains firm age bins, indicators for the type of the firm's most recent deal (private equity, venture, corporate debt, etc.), and indicators for the most recent firm status observed as of t: profitability, revenue-generating, or trials/development. The λ terms are fixed effects for sector-time $s(i) \times t$ and state $g(i)$. The sample is all firms who ever file a device of type X in the time-frame with years after each firm's first outcome Y event excluded.

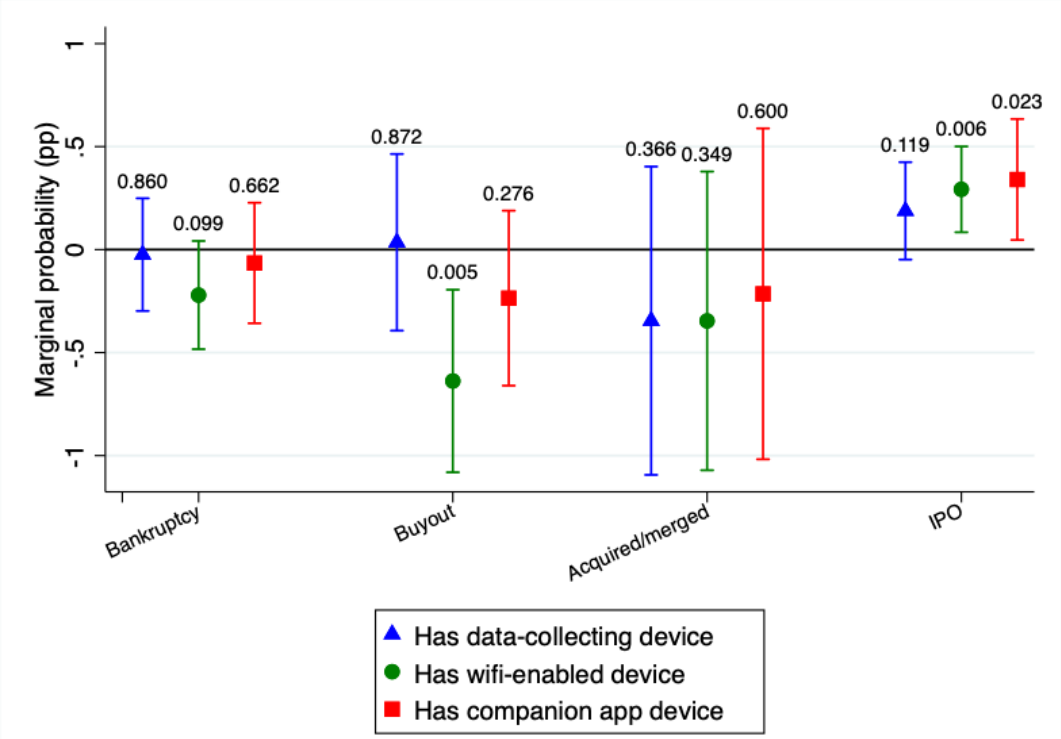
Figure 4 plots β_1 for each outcome from Equation 1 from 2007-2013 in Panel A and from 2014-2019 in Panel B. We split the sample in 2014 to consider the trajectories of data-collecting firms in two periods that appear distinct; as Figure 3 demonstrates, 2014 is the first year in which the median data-collector had more assets, a higher valuation, and spent more on R&D than the median non-data-collector.

Figure 4 shows that—conditional on age, deal histories, profitability, and fixed effects—both firms with wifi-enabled devices and companion app devices were 0.2-0.3 percentage point more likely to IPO than firms without, and firms with wifi-enabled devices were more than a half a percentage point less likely to be bought out.

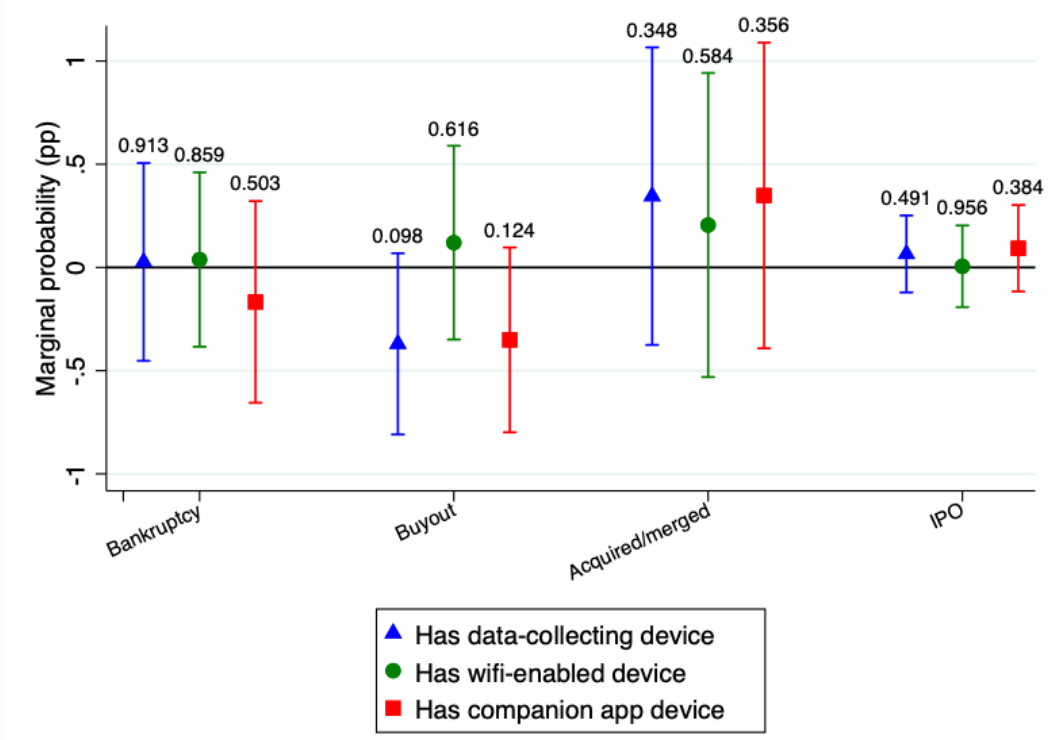
By contrast, there is no difference between early-stage data-collectors and non-data collectors after 2014. In this late period, where public data-collectors dominate non-data collectors (see Figure 3), early-stage data-collectors have no advantages over their non-data-collecting peers.

Figure 4: Differential Exit Trajectories for Early-Stage Data-Collectors

(a) 2007-2013



(b) 2014-2019



4 Data Capability and Firm Outcomes

4.1 Background on GDPR

The EU's General Data Protection Regulation (GDPR)—Regulation (EU) 2016/679—replaced the 1995 Data Protection Directive with a single, directly applicable EU-wide framework governing the processing of personal data. The regulation entered into force on May 24, 2016 and became applicable on May 25, 2018, following a two-year transition period. GDPR has explicit extraterritorial reach: it can apply to entities outside the EU when they offer goods or services to individuals in the EU or monitor their behavior there. GDPR is organized around core processing principles—lawfulness, purpose limitation, data minimization, storage limitation, integrity/confidentiality, and accountability—and grants individuals enforceable rights (e.g., access, erasure, portability, objection).

GDPR is enforced by national supervisory authorities, coordinated through EU cooperation and consistency mechanisms (including the "one-stop-shop" for cross-border processing), with the European Data Protection Board (EDPB) playing a central role in promoting consistent application. The regulation raised compliance and enforcement stakes through (i) procedural obligations such as personal-data breach notification within 72 hours where feasible, and (ii) a two-tier administrative fine structure, including fines up to 20 Euros million or 4% of worldwide annual turnover for the most serious infringements.

GDPR is technology-neutral and therefore applies whenever personal data are processed through connected devices, companion apps, or cloud services. IoT products routinely generate data within GDPR's scope—including persistent device identifiers (IP addresses, RFID tags), location data, and telemetry—when linkable to an identifiable person.[9]

4.2 Estimation and Results

We implement a triple differences-in-differences design from Equation 2 on a balanced quarterly panel of public firms matched across Compustat and the FCC data to estimate the effect of the GDPR on EU-exposed firms with data-collecting devices. We focus on the quarters between 2011q1 and 2019q4 for this analysis. The event date is denoted by t^*, which here we assign as 2016q2, when the EU Parliament adopted the GDPR. The regulation was not implemented until 2018, but we look to include potential firm adjustments in anticipation of the GDPR during the phase-in period in the overall ATT of the GDPR. We restrict the sample to firms active in all quarters from 2015q1-2019q4.

[9] See: EU (2016); EuropeanCommission (2026); EDPB (2026)

$$Y_{it} = \beta_1[\mathbb{1}(t \geq t^*) \times \mathbb{1}_i(\text{EU-exposed}) \times D_i] + \beta_2[\mathbb{1}(t \geq t^*) \times \mathbb{1}_i(\text{EU-exposed})] + \beta_3[\mathbb{1}(t \geq t^*) \times D_i] + \beta_4[\mathbb{1}_i(\text{EU-exposed}) \times D_i] + \lambda_i + \eta_{qtr \times s(i)} + \epsilon_{it} \quad (2)$$

We consider as outcomes Y_{it} the firms' quarterly log market value, log revenue, log R&D spending, Tobin's Q, and inverse-hyperbolic sine (IHS) transformations of patents, citations, and citation-weighted patents for all patent filings on all topics and on AI/ML topics. D_i is a measure of the firm's investment in data-capable devices. We consider several functional forms for D_i – a dummy for any data-capable devices, raw count of data-capable devices, and log data-capable devices – and indicate which is in use for a given set of results. Equation 2 includes firm fixed effects λ_i and quarter $\times$ 2-digit SIC industry $\eta_{qtr \times s(i)}$ fixed effects. We do not include fixed effects for the geography of firms' EU exposures because the granularity of EU exposure reporting is not consistent and often highly aggregated.

The main coefficient of interest here is β_1, which we interpret as the data-mediated impact of the GDPR on exposed firms to the extent that our core identification assumptions hold. These are that (A1) differences in trends between EU-exposed data-collectors and non-data collectors would have evolved similarly to the differences in trends between non-EU data-collectors and non-data collectors in a counterfactual without GDPR, (A2) the compositions of each of the four groups of firms (data-collects/not $\times$ EU-exposed/not) do not change differentially in the post period, and (A3) the GDPR does not affect outcomes for firms we classify as untreated (non-data-collecting or non-EU exposed).

While (A1) is not directly verifiable we estimate generally flat pre-trends in our event study plots, suggesting that groups did not necessarily have sustained differential evolutions leading into GDPR. The primary threat to (A2) is that the GDPR drove differential compositional changes across groups, but Table A.3 compares the characteristics of firms that exit after GDPR to those who do not and no difference in data capability or EU exposure by exit status. Further, Tables A.6-A.6 shows that the main results are robust to balancing the firm panel by including only firms alive in all throughout 2015q1 - 2019q4 (the year before GDPR to the end of the sample). (A3) is threatened if either our EU-exposure or data-collection measures are flawed or if non-EU or non-data-collecting firms are affected by GDPR in ways not controlled for by quarter $\times$ sector fixed effects. The EU-exposure measure is based on customer segment information described in 2.3 and we validate our data-collection measure in 2.2.

We partition the sample by 4-digit SIC industries into a set of consumer device producers (consumer electronics, household appliances, toys and sporting goods,...) and non-consumer device producers (industrial machinery, medical/surgical device production, aerospace and defense, and telecommunications equipment) as a falsification test of our consumer data mechanism: non-consumer device producers should not be affected

do not expect restrictions on customer data transmission in the EU to affect non-consumer device manufacturers, so we partition the sample on industry into a set of consumer device producers and non-consumer device producers by excluding from the consumer set all device filers in industries like industrial machinery, medical/surgical device production, aerospace and defense, and telecommunications. Appendix Table A.3 presents the full list of SIC industry inclusions and exclusions used to create the consumer and non-consumer samples and demonstrates that our results are robust to alternate industry selections.

We demonstrate that this distinction is meaningful by estimating an alteration of Equation 2 among data-capable firms. We plot the coefficient on $consumer \times EU\text{-}exposed \times post\text{-}GDPR$ among all data-capable firms in Figure 5 for revenue. Consumer and non-consumer firms are selected so we do interpret this as a purely descriptive result demonstrating that consumer device firms may have been affected by consumer data restrictions in a way non-consumer firms were not.

Table 4.2 presents β_1 from Equation 2 estimated for financial outcomes, where data-capability is a binary indicator for whether the firm had data-capable device filings before the GDPR. Consumer firms are in Panel A and non-consumer firms are in Panel B. The table suggests that neither consumer nor non-consumer firms exposed to the GDPR were affected financially. Table A.5 repeats this analysis using the count of data devices instead of a data-capability binary for the treatment and shows near-zero, precise effects for non-consumer firms and small, precise, and negative effects on market value and Tobin's Q for the consumer firms. Together these estimates suggest that the largest consumer data device firms may have felt small, negative financial impacts from GDPR, but that financial effects were in general non-existent or limited for both consumer and non-consumer firms.

Table 4.2 presents β_1s on patent filing outcomes using the binary data-capability indicator for consumer firms in Panel A and non-consumer firms in Panel B. While patent filings of all types are unaffected, Columns 2 and 4 of Panel A demonstrate large, negative effects on AI/ML-related patent filing (down 36%) and citation-weighted AI/ML patent filing (down 89%) in the following GDPR announcement. The difference in these effects suggests that the patent filings lost to this shock would have been high impact. In Panel B, non-consumer firms massively increase AI/ML patent filing (up 994%), citation-weighted patent filing (up

Figure 5: GDPR Effect on IHS Revenue, Consumer Data Firms v. Non-Consumer Data Firms

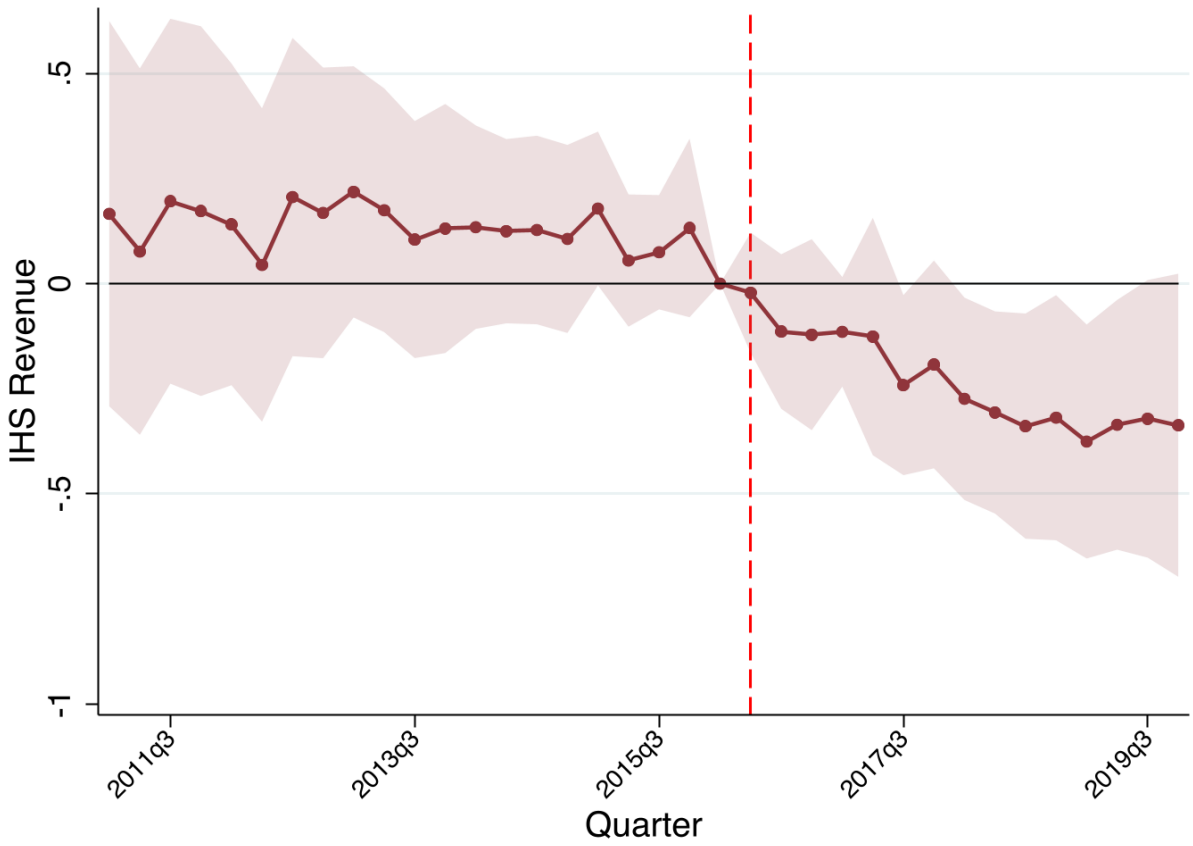


Table 3: Financial Outcomes — Binary Treatment, Unbalanced Panel

Panel A: Consumer				
	(1)	(2)	(3)	(4)
	IHS MktVal	IHS Rev	IHS R&D	IHS Tobin's Q
EU × Post × Data	-0.089	-0.081	-0.183	-0.115
	(0.241)	(0.171)	(0.203)	(0.210)
Dep. var. mean	8.43	6.27	3.51	1.35
Dep. var. SD	2.67	2.76	2.45	0.89
Observations	5,102	5,259	3,570	5,096
Firms	161	161	145	161
Adj. R-sq	0.963	0.981	0.977	0.777
Panel B: Non-Consumer				
	(1)	(2)	(3)	(4)
	IHS MktVal	IHS Rev	IHS R&D	IHS Tobin's Q
EU × Post × Data	0.196	0.239	0.053	-0.144
	(0.202)	(0.146)	(0.133)	(0.145)
Dep. var. mean	8.28	6.38	3.71	1.18
Dep. var. SD	2.47	2.36	2.07	0.59
Observations	5,910	6,081	4,922	5,910
Firms	185	185	173	185
Adj. R-sq	0.969	0.981	0.970	0.742

Regressions include firm and year-by-SIC2 fixed effects. Standard errors are clustered by firm.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

82%), and citation-weighted AI/ML patent filing (up 1,470%). Table A.5 repeats this analysis using the count of data devices instead of a data-capability binary and shows near-zero or insignificant effects of additional data devices on all patenting outcomes.

Table 4: Patent Outcomes (Poisson) — Binary Treatment, Unbalanced Panel

Panel A: Consumer					
	(1) Patents	(2) AI/ML	(3) CW Pat	(4) CW AI/ML	(5) Val Pat ($)
EU × Post × Data	0.034	-0.454***	-0.409	-2.225***	-0.126
	(0.304)	(0.127)	(0.470)	(0.229)	(0.311)
Dep. var. mean	223	27	420	1639	4345
Dep. var. SD	845	113	2481	4130	14518
Observations	945	343	920	275	945
Panel B: Non-Consumer					
	(1) Patents	(2) AI/ML	(3) CW Pat	(4) CW AI/ML	(5) Val Pat ($)
EU × Post × Data	0.408	2.393***	0.597*	2.755**	-0.298
	(0.295)	(0.891)	(0.360)	(1.087)	(0.369)
Dep. var. mean	144	5	99	267	2148
Dep. var. SD	408	17	293	790	10246
Observations	1,252	465	1,209	366	1,252

Regressions include firm and year-by-SIC2 fixed effects. Standard errors are clustered by firm.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

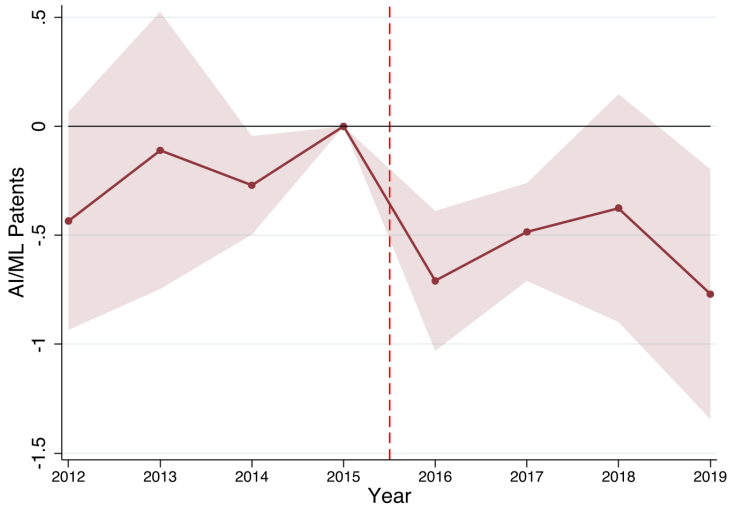
The estimated effects for all firms, but especially non-consumer firms, are massive. We re-estimate these effects in Table A.6 excluding the top 5 consumer patenters and the top 5 non-consumer patenters, respectively, and find that the effects disappear. Thus these patent filing results are driven entirely by the most prolific AI/ML patent filers. These top 5 patent filers among consumer device producers are IBM, Alphabet, Microsoft, Amazon, and Ford and the top 5 among non-consumer device producers are Intel, Qualcomm, Apple, GE Aerospace, and HP.

Figure 6a shows an event study for patent filing for all consumer firms using the binary treatment. It demonstrates a clear trend break after the GDPR was announced in 2016. We do see one significant pre-period coefficient in 2014. Event studies estimated on IHS-transformed outcomes do not exhibit pre-trends, suggesting that the pre-period Poisson coefficients are potentially driven by the sensitivity of count models to large outliers in thin cells rather than by systematic and multi-period differential trends. Our ongoing work to expand the sample aims to improve precision for these analyses. Still, the declining post-

2016 pattern in AI and machine learning patent applications, as well as the ATT results presented in the tables above, are consistent with the interpretation that EU exposed data capable firms saw meaningful innovation slowdowns after GDPR's restrictions to data use were announced.

Figure 6: Patent Filing by Consumer Device Firms Around the GDPR

(a) AI/ML Patents



5 Discussion and Conclusion

This paper develops a novel, product-level measure of data capital by applying natural language processing to the user-manual texts of wireless devices authorized by the FCC. We document that data-capable connected devices—"smart" products that collect and transmit user data via companion apps and cloud services—are increasingly characteristic of large, R&D-intensive, and innovative public firms, and that prior data-device commercialization is associated with more prolific subsequent product introductions and, among startups, higher likelihoods of successful exits.

Our central empirical contribution exploits the GDPR as a quasi-natural experiment restricting firms' ability to collect and process personal data from European customers. Using a triple-difference design that compares EU-exposed data-capable consumer firms to (i) EU-exposed non-data firms, (ii) non-EU data firms, and (iii) non-consumer data firms, we find that GDPR adoption is associated with a 36.5% decline in AI/ML patent filings and an 89.2% decline in citation-weighted AI/ML patents among treated consumer firms. These effects are concentrated in precisely the technology domain where first-party data is a core input. Non-consumer firms—whose connected devices serve industrial rather than personal-data applications—show no such decline, and if anything exhibit increased patenting, consistent with GDPR selectively disrupting innovation that depends on personal consumer data.

These findings have implications for how we think about data as a form of intangible capital. The innovation results suggest that first-party data functions as a productive, depreciable asset: it is accumulated through deliberate investment in data-capable hardware, it is durable and reusable across production cycles, and it complements R&D in producing innovative output. When regulation restricts the collection and reuse of this capital stock—through GDPR's purpose-limitation, data-minimization, and erasure requirements—the returns to complementary R&D investment fall, reducing both the incentive and the capacity to innovate in data-intensive domains like AI.

We highlight two areas where we see room for improvement in ongoing work. The first is statistical power. Our triple-difference design identifies off of roughly 44 firms that are simultaneously EU-exposed, data-capable, and consumer-facing—a small group, even if the patent results prove robust to bootstrap inference, leave-one-out exercises, and pooled estimation on the full 338-firm sample. The null financial results likely reflect this: with 44 treated firms, we simply may not have enough variation to detect effects on noisy quarterly outcomes like market value or revenue. We are working to expand the panel in several ways—extending the sample window past 2019 to gain more post-enforcement years, tightening the

FCC-to-Compustat matching to bring in additional firms, and incorporating international firm data that would let us exploit richer cross-country variation in GDPR exposure.

The second is that several of our variables are more blunt than would be ideal to capture real-world phenomena. EU exposure comes from geographic segment disclosures, which firms report under a vague materiality standard—so some firms we classify as unexposed may in fact sell to European customers, and some we classify as exposed may have minimal EU revenue. The fact that results survive when we swap in continuous EU revenue shares or broader international-exposure indicators is encouraging, but we are also building more granular measures from subsidiary-level CapitalIQ data. The consumer/non-consumer split is similarly rough: it relies on SIC codes, which do not always map cleanly onto whether a firm's connected devices end up in households or on factory floors. Our ongoing work also aims to build device- and firm-level customer-composition measures to do better here. Our data-capability indicator itself is a simple flag for whether a firm sells any connected device—it does not distinguish a company with a single Bluetooth speaker from one with an ecosystem of cloud-connected products and a broad user-base collecting terabytes of data. Thus, our ongoing work aims to build richer measures that can help capture the intensity of data collection more directly.

Notwithstanding these limitations, the paper provides new evidence that data from connected devices acts as a productive intangible asset with material implications for innovation. As the IoT continues to grow and privacy regulation spreads beyond Europe, the tension between protecting consumer data and preserving the innovation incentives that data creates will only become more pressing.

References

Aridor, G., Che, Y.-K., & Salz, T. (2023). The effect of privacy regulation on the data industry: empirical evidence from GDPR. *RAND Journal of Economics*, *54*(4), 695–730. Retrieved 2026-01-28, from https://ideas.repec.org//a/bla/randje/v54y2023i4p695-730.html (Publisher: RAND Corporation)

Babina, T., Fedyk, A., He, A., & Hodson, J. (2024). Artificial intelligence, firm growth, and product innovation. *Journal of Financial Economics*, *151*(C). Retrieved 2026-01-29, from https://ideas.repec.org//a/eee/jfinec/v151y2024ics0304405x2300185x.html (Publisher: Elsevier)

Bergeaud, A., Jaffe, A. B., & Papanikolaou, D. (2025, May). *Natural Language Processing and Innovation Research* [Working Paper]. National Bureau of Economic Research. Retrieved 2026-01-27, from https://www.nber.org/papers/w33821 doi: 10.3386/w33821

Blind, K., Niebel, C., & Rammer, C. (2024). The impact of the EU general data protection regulation on product innovation. *Industry and Innovation*, *31*(3), 337–357.

Board, F. A. S. (2026). *Accounting standards codification.* Retrieved from https://asc.fasb.org/

Boroomand, F., Leiponen, A., & Vasudeva, G. (2022). Does the Market Value Attention to Data Privacy? Evidence from U.S.-listed Firms under the GDPR.

BusinessWire.com. (2021). *Data Creation and Replication Will Grow at a Faster Rate Than Installed Storage Capacity, According to the IDC Global DataSphere and StorageSphere Forecasts.* Retrieved 2026-02-01, from https://www.businesswire.com/news/home/20210324005175/en/Data-Creation-and-Replication-Will-Grow-at-a-Faster-Rate-Than-Installed-Storage-Capacity-According-to-the-IDC-Global-DataSphere-and-StorageSphere-Forecasts

Chung, K. H., & Pruitt, S. W. (1994). A simple approximation of tobin's q. *Financial Management*, *23*(3), 70–74. Retrieved 2026-02-01, from http://www.jstor.org/stable/3665623

Crouzet, N., & Eberly, J. (2023). Rents and Intangible Capital: A Q+ Framework. *The Journal of Finance*, *78*(4), 1873–1916. Retrieved 2026-02-02, from https://onlinelibrary.wiley.com/doi/abs/10.1111/jofi.13231 (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/jofi.13231) doi: 10.1111/jofi.13231

Crouzet, N., Eberly, J. C., Eisfeldt, A. L., & Papanikolaou, D. (2022, August). The Economics of Intangible Capital. *Journal of Economic Perspectives*, *36*(3), 29–52. Retrieved 2026-02-02, from https://www.aeaweb.org/articles?id=10.1257%2Fjep.36

.3.29 doi: 10.1257/jep.36.3.29

de Cornière, A., & Taylor, G. (2024, September). Data-Driven Mergers. *Management Science*, *70*(9), 6473–6482. Retrieved 2026-01-27, from https://pubsonline.informs.org/doi/10.1287/mnsc.2023.04104 (Publisher: INFORMS) doi: 10.1287/mnsc.2023.04104

Demirer, M., Jiménez Hernández, D. J., Li, D., & Peng, S. (2024, February). *Data, Privacy Laws and Firm Production: Evidence from the GDPR* [Working Paper]. National Bureau of Economic Research. Retrieved 2026-01-28, from https://www.nber.org/papers/w32146 doi: 10.3386/w32146

EDPB. (2026). *GDPR Cooperation and Enforcement | European Data Protection Board.* Retrieved 2026-01-29, from https://www.edpb.europa.eu/our-work-tools/support-cooperation-and-enforcement/gdpr-cooperation-and-enforcement_en?utm_source=chatgpt.com

EU, R. (2016, April). *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance).* Retrieved 2026-01-29, from http://data.europa.eu/eli/reg/2016/679/oj (Legislative Body: EP, CONSIL)

EuropeanCommission. (2026). *Data protection explained - European Commission.* Retrieved 2026-01-29, from https://commission.europa.eu/law/law-topic/data-protection/data-protection-explained_en

Ewens, M., Peters, R. H., & Wang, S. (2025, January). Measuring Intangible Capital with Market Prices. *Management Science*, *71*(1), 407–427. Retrieved 2026-01-28, from https://pubsonline.informs.org/doi/10.1287/mnsc.2021.02058 (Publisher: INFORMS) doi: 10.1287/mnsc.2021.02058

Farboodi, M., Mihet, R., Philippon, T., & Veldkamp, L. (2019, May). Big Data and Firm Dynamics. *AEA Papers and Proceedings*, *109*, 38–42. Retrieved 2026-01-27, from https://www.aeaweb.org/articles?id=10.1257/pandp.20191001 doi: 10.1257/pandp.20191001

Farboodi, M., & Veldkamp, L. (2025). *FV_datagrowth.pdf.* Retrieved 2026-01-28, from https://drive.google.com/file/u/1/d/1dXcOXTwdKP40KPk2agxBUUxSdSPu-vYk/view?usp=sharing&usp=embed_facebook

Giorgio, M., Benedikt Frey, C., & Andres, P. (2024). *Redirecting AI: Privacy regulation and the future of artificial intelligence* (VoxEU). Oxford Martin School. (Available at https://cepr.org/voxeu/columns/redirecting-ai-privacy-regulation

-and-future-artificial-intelligence)

Goldberg, S. G., Johnson, G. A., & Shriver, S. K. (2024, February). Regulating Privacy Online: An Economic Evaluation of the GDPR. *American Economic Journal: Economic Policy*, *16*(1), 325–358. Retrieved 2026-01-28, from https://www.aeaweb.org/articles?id=10.1257%2Fpol.20210309&utm_source=chatgpt.com doi: 10.1257/pol.20210309

Hall, B. H., Jaffe, A. B., & Trajtenberg, M. (2005). Market value and patent citations. *The RAND Journal of Economics*, *36*(1), 16–38.

IOT-Analytics.com. (2025). *Number of connected IoT devices growing 14% to 21.1 billion.* Retrieved 2026-02-01, from https://iot-analytics.com/number-connected-iot-devices/

Jones, C. I., & Tonetti, C. (2020, September). Nonrivalry and the Economics of Data. *American Economic Review*, *110*(9), 2819–2858. Retrieved 2026-01-27, from https://www.aeaweb.org/articles?id=10.1257/aer.20191330 doi: 10.1257/aer.20191330

Kogan, L., Papanikolaou, D., Seru, A., & Stoffman, N. (2017, May). Technological Innovation, Resource Allocation, and Growth*. *The Quarterly Journal of Economics*, *132*(2), 665–712. Retrieved 2026-01-27, from https://doi.org/10.1093/qje/qjw040 doi: 10.1093/qje/qjw040

Liu, P.-Y. (2024). Value of Data to U.S. Public Firms: Evidence from GDPR.

Quan, J. (2024). *Junjun_quan_value_of_data.pdf.* Working Paper. Retrieved 2026-01-29, from https://www.dropbox.com/scl/fi/54epblhmp377v12h00itc/Junjun_Quan_Job_Market_Paper.pdf?dl=0&rlkey=zwzilgvfdhsrg665hrpgkt55t

Sun, L., & Abraham, S. (2021, December). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, *225*(2), 175–199. doi: 10.1016/j.jeconom.2020.09.006

Szkutak, R. (2025, January). *VCs say AI companies need proprietary data to stand out from the pack.* Retrieved 2026-01-28, from https://techcrunch.com/2025/01/10/vcs-say-ai-companies-need-proprietary-data-to-stand-out-from-the-pack/

A Appendix

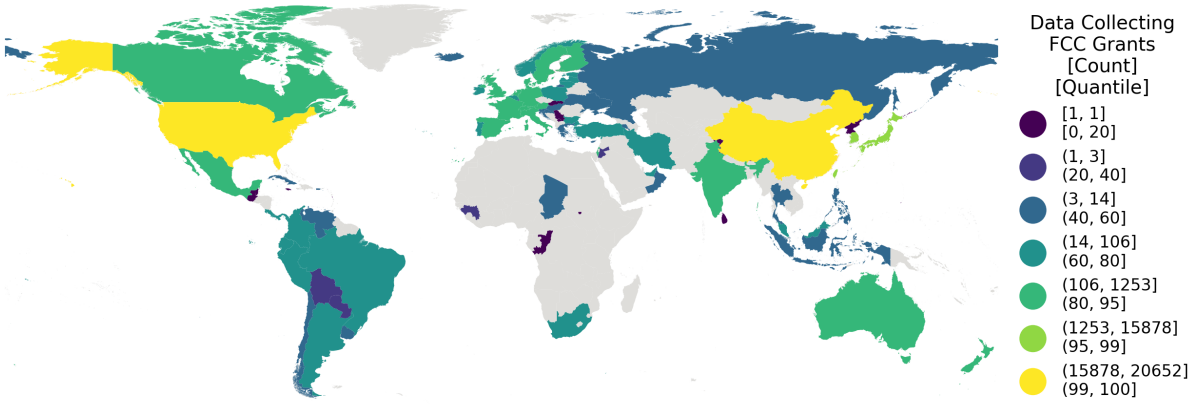
Table A.1: Mean keyword frequency by category: data-device vs. non-data-device deals in SEC acquisition filings

	Target-Only (N=31)				**All Deals (N=175)**			
	Data (n=13)	No Data (n=18)	Diff	p-value	Data (n=118)	No Data (n=57)	Diff	p-value
Engagement / Retention	1.46	0.00	+1.46	0.174	0.42	0.04	+0.38	0.006***
IoT / Connected Devices	2.92	0.06	+2.87	0.080*	0.78	0.25	+0.53	0.046**
Data Collection	0.77	0.00	+0.77	0.232	0.25	0.00	+0.25	0.024**
Monetization	0.46	0.00	+0.46	0.337	0.07	0.00	+0.07	0.207
Advertising	1.92	0.06	+1.87	0.220	0.53	0.44	+0.09	0.824
Unit Economics	6.00	0.00	+6.00	0.337	0.77	0.09	+0.68	0.308
Personalization	0.00	0.06	−0.06	0.331	0.07	0.12	−0.06	0.494

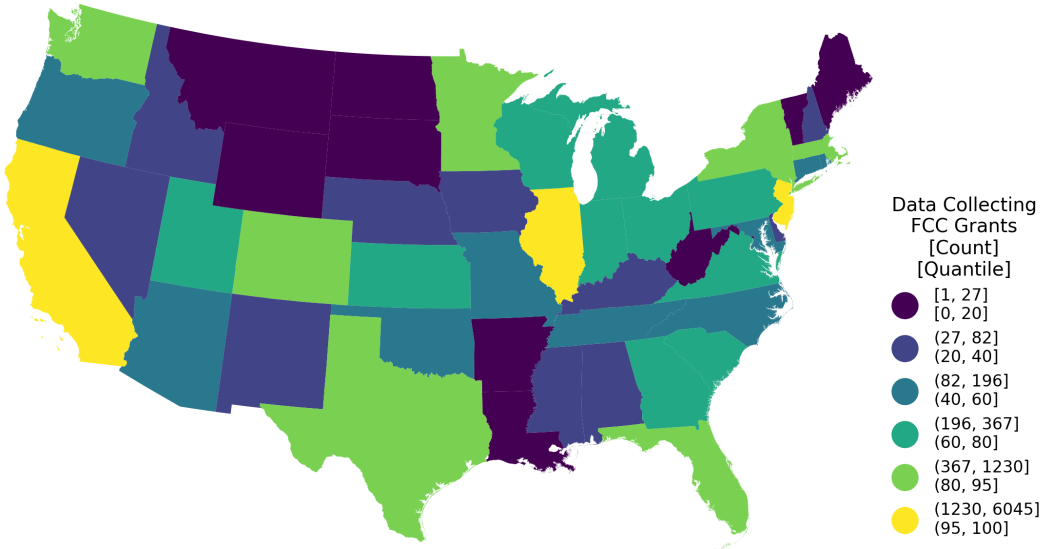
Notes: Mean keyword mentions per deal. Welch's t-test used for difference in means. "Data" refers to deals where the matched firm is classified as a data-device company; "No Data" refers to deals where the matched firm is not. Sample restricted to high-quality FCC-SEC matches (score ≥ 0.9).
*** p<0.01, ** p<0.05, * p<0.10

Figure A.1: Approved FCC filings by grantee country and state headquarters, 2007-2025

(a) Approved FCC filings by country



(b) Approved FCC filings by state



Note: Quantiles and corresponding counts are computed based on entities with at least one approved FCC filing for a data-collecting device.

Figure A.2: Sector-adjusted comparisons across data capable versus non-capable public firms (2007-2023)

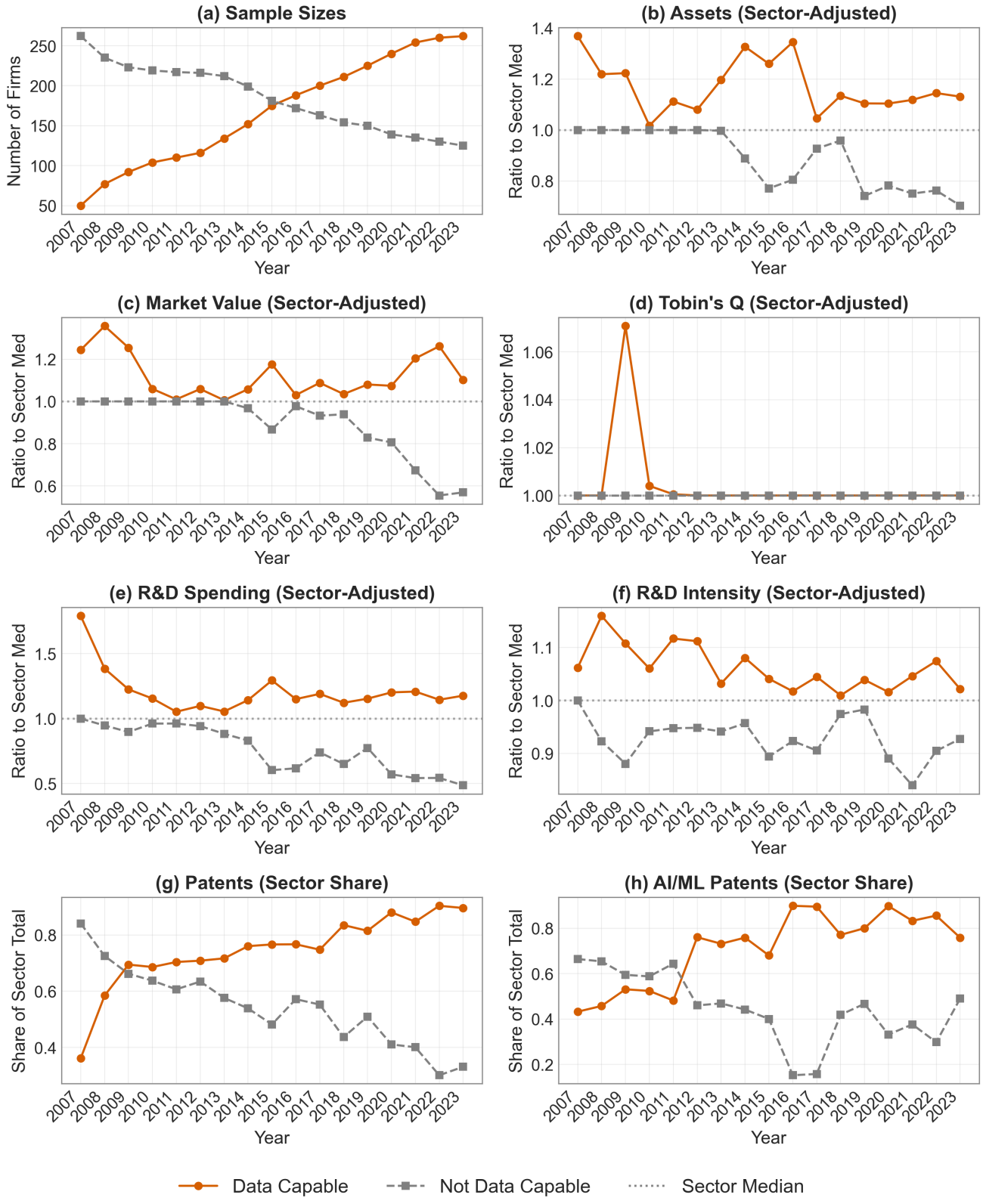
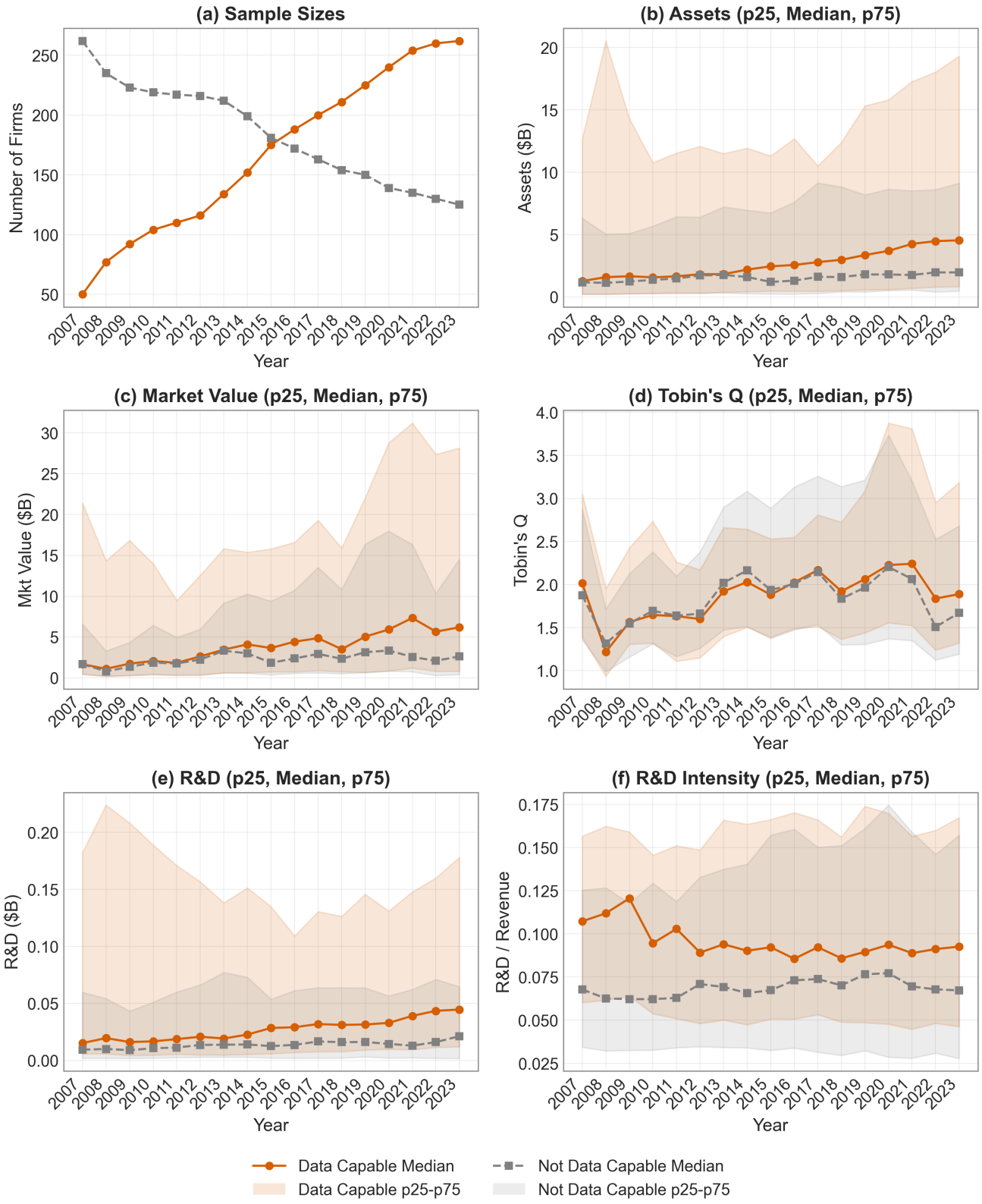


Figure A.3: Comparing data capable versus non-capable public firms on the basis of medians, and 25th and 75th percentiles, (2007-2023)



A.1 Details of NLP Algorithm to Identify Data-Capable Devices

1. Document Filtering: In some instances, some FCC application documents like test reports or cover letters that are not user manuals are saved under that label. So our first step was to exclude such documents by searching for terms characteristic of test reports and compliance documentation (e.g., "test report," "EMC test," "SAR test," "RF exposure evaluation," "label location"). Documents shorter than 500 characters are also excluded.

2. Primary Keyword Detection: A device is classified as having a companion app if its user manual contains any single match from a set of high-confidence patterns. These include: (i) explicit app-type mentions such as "mobile app," "smartphone app," "companion app," "iOS app," or "Android app"; (ii) app installation references such as "download the app" or "install the app"; (iii) app store references including "App Store," "Google Play," or "Play Store"; (iv) brand-specific app names for major IoT platforms and device manufacturers (e.g., Ring, Wyze, Arlo, Nest, Eufy, Tuya, SmartLife, Kasa, Alexa, Google Home); and (v) direct app references such as "our app," "use the app to," or "in-app settings."

3. Secondary Keyword Detection: If no primary keywords are found, a device is classified as data-capable if two or more secondary indicators are present. Secondary keywords capture: (i) smart home platform references (SmartLife, Tuya, eWelink, SmartThings); (ii) generic app functionality phrases ("through the app," "in the app," "app control"); (iii) WiFi/Bluetooth pairing and network setup language; (iv) remote access patterns ("remote view," "view anywhere," "monitor remotely," "live stream"); (v) push notification and alert patterns; (vi) account and cloud patterns ("create account," "sign up," "cloud storage," "cloud recording"); and (vii) QR code setup references.

4. Other Data Collection Indicators: Independent of app detection, we flag devices whose manuals reference explicit data transmission or storage activities, including "data sync," "upload data," "usage data," "collect data," "cloud storage," "store to server," or "transmit data."

A device is classified as "data-capable" if it satisfies either the companion-app criteria (items #2–3) or the other-data-collection criteria (#4). Using this approach, more than a third of devices in our sample are classified as data-capable, with the rate rising over time consistent with the growth of the Internet of Things.

A.2 Details of the Matching Algorithm

1. Standardize firm details (name, address, contact information) across FCC grantees, CapitalIQ, and PitchBook.

2. Exact match FCC grantees to CapitalIQ or PitchBook using firm legal name.

3. Exact match FCC grantees to CapitalIQ or PitchBook using contact information and selecting for the match with the highest name similarity.

4. Exact match on the soundex encoding of each firm's name, the first word of each firm's name, the reported state and city of each firm, or the first three letters of the firm's name, keeping matches where firm names without legal identifiers display over 50 percent similarity.

5. Across the resulting batch of candidates, keep those with an exact match on name and contact information, then keep candidate matches based on the following prioritized sequence of criteria:

   (a) over 95 percent similarity in firm legal names.

   (b) over 75 percent similarity in firm legal names and street addresses.

   (c) over 95 percent similarity of one candidate firm's name against a substring of the other candidate firm's name.

6. For firms mapped to CapitalIQ, map the corresponding CapitalIQ firm to Compustat, or the parent firm's Compustat record if an operating subsidiary, using GVKeys.

A.3 Characteristics of Post-GDPR Exiters

Table A.2: Pre-GDPR (2015Q4) Characteristics: Stayers vs Exiters

Variable	Stayers (mean)	Exiters (mean)	p-value
Market Value ($M)	32438.1	996.1	0.000
Revenue ($M)	3976.1	216.4	0.000
R&D ($M)	322.5	10.1	0.000
Assets ($M)	23113.0	345.5	0.000
Device Count	4.59	0.89	0.058
Data Share	0.35	0.29	0.746
Any Data Device	0.44	0.33	0.533
EU Exposure	0.50	0.33	0.356

Consumer firms, unbalanced panel. t-tests with unequal variances.

A.4 Defining Consumer vs Non-Consumer-oriented Firms

Table A.3: Industries excluded from our sample to identify consumer-oriented firms

SIC Code(s)	Category	Description
Industrial & Computing Equipment		
SIC2 = 35	Machinery	Industrial and commercial machinery and computer equipment
3576	Networking	Computer communications equipment
3578	Point-of-sale	Calculating and accounting machines
3577	Peripherals	Computer peripheral equipment, n.e.c.[a]
Electronic Components & Semiconductors		
3674	Semiconductors	Semiconductors and related devices
3672	PCBs	Printed circuit boards
3679	Components	Electronic components, n.e.c.
Instruments & Medical Devices		
3812, 3823–3829	Instruments	Laboratory, scientific, and measuring instruments
3841–3844	Medical	Surgical and medical instruments and devices
Aerospace & Defense		
3721, 3724, 3728	Aircraft	Aircraft, engines, and parts
3760, 3769	Defense	Guided missiles, space vehicles, and parts
Other Manufacturing		
3714	Auto parts	Motor vehicle parts and accessories
3621	Motors	Motors and generators
3640	Lighting	Electric lighting and wiring equipment
Telecommunications & Carriers		
4812, 4813	Telecom carriers	Telephone and telegraph apparatus
3661, 3663, 3669	Telecom equipment	Communications equipment[b]
Business Services		
7371	Software	Computer programming and data processing
7380, 7389	Services	Miscellaneous business services, n.e.c.
Primary Industries, Transport, & Wholesale		
SIC2 = 10, 12–14	Mining	Metal, coal, oil, and nonmetallic mineral mining
SIC2 = 40–42, 44–47	Transport	Railroad, trucking, water, air, and pipeline transport
SIC2 = 49	Utilities	Electric, gas, and sanitary services
SIC2 = 50	Wholesale	Wholesale trade—durable goods

[a] Turtle Beach Corp (SIC 3577) is retained in the consumer sample. [b] Telecom. equipment manufacturers are excluded from both the consumer and non-consumer samples.

A.5 Additional Results

Table A.4: Financial Outcomes — Count Treatment, Unbalanced Panel

Panel A: Consumer				
	(1)	(2)	(3)	(4)
	IHS MktVal	IHS Rev	IHS R&D	IHS Tobin's Q
EU × Post × Count	-0.027**	0.000	-0.010	-0.023***
	(0.011)	(0.007)	(0.008)	(0.007)
Dep. var. mean	8.43	6.27	3.51	1.35
Dep. var. SD	2.67	2.76	2.45	0.89
Observations	5,102	5,259	3,570	5,096
Firms	161	161	145	161
Adj. R-sq	0.963	0.981	0.976	0.779
Panel B: Non-Consumer				
	(1)	(2)	(3)	(4)
	IHS MktVal	IHS Rev	IHS R&D	IHS Tobin's Q
EU × Post × Count	0.008**	0.005**	0.006**	-0.002
	(0.003)	(0.002)	(0.002)	(0.002)
Dep. var. mean	8.28	6.38	3.71	1.18
Dep. var. SD	2.47	2.36	2.07	0.59
Observations	5,910	6,081	4,922	5,910
Firms	185	185	173	185
Adj. R-sq	0.969	0.981	0.970	0.741

Unbalanced panel (enter by 2015q1, allow exit). Firm + Quarter-by-SIC2 FE. SEs clustered by firm.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.5: Patent Outcomes (Poisson) — Count Treatment, Unbalanced Panel

Panel A: Consumer					
	(1) Patents	(2) AI/ML	(3) CW Pat	(4) CW AI/ML	(5) Val Pat ($)
EU × Post × Count	0.024***	-0.005***	0.005	0.003*	-0.005
	(0.005)	(0.001)	(0.017)	(0.002)	(0.014)
Dep. var. mean	223	27	420	1639	4345
Dep. var. SD	845	113	2481	4130	14518
Observations	945	343	920	275	945
Firms	109	46	106	37	109
Panel B: Non-Consumer					
	(1) Patents	(2) AI/ML	(3) CW Pat	(4) CW AI/ML	(5) Val Pat ($)
EU × Post × Count	0.003	0.008	0.006*	-0.021	0.006
	(0.003)	(0.012)	(0.003)	(0.018)	(0.004)
Dep. var. mean	144	5	99	267	2148
Dep. var. SD	408	17	293	790	10246
Observations	1,252	465	1,209	366	1,252
Firms	145	56	139	43	145

Poisson PML (ppmlhdfe). Unbalanced panel. Firm + Year-by-SIC2 FE. SEs clustered by firm.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

A.6 Robustness to Balanced Panel Restriction

Table A.6: Financial Outcomes — Binary Treatment, Balanced Panel

Panel A: Consumer				
	(1)	(2)	(3)	(4)
	IHS MktVal	IHS Rev	IHS R&D	IHS Tobin's Q
EU × Post × Data	-0.030	-0.066	-0.163	-0.100
	(0.249)	(0.178)	(0.213)	(0.217)
Dep. var. mean	8.54	6.39	3.59	1.33
Dep. var. SD	2.61	2.71	2.46	0.83
Observations	4,898	5,036	3,407	4,898
Firms	146	146	133	146
Adj. R-sq	0.962	0.980	0.976	0.766
Panel B: Non-Consumer				
	(1)	(2)	(3)	(4)
	IHS MktVal	IHS Rev	IHS R&D	IHS Tobin's Q
EU × Post × Data	0.165	0.285*	0.078	-0.198
	(0.211)	(0.151)	(0.138)	(0.150)
Dep. var. mean	8.41	6.49	3.79	1.18
Dep. var. SD	2.38	2.26	2.03	0.58
Observations	5,618	5,777	4,670	5,618
Firms	167	167	157	167
Adj. R-sq	0.968	0.980	0.968	0.750

Balanced panel (2015q1–2019q4). Firm + Quarter-by-SIC2 FE. SEs clustered by firm.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.7: Financial Outcomes — Count Treatment, Balanced Panel

Panel A: Consumer				
	(1)	(2)	(3)	(4)
	IHS MktVal	IHS Rev	IHS R&D	IHS Tobin's Q
EU × Post × Count	-0.026**	0.000	-0.010	-0.023***
	(0.011)	(0.007)	(0.008)	(0.007)
Dep. var. mean	8.54	6.39	3.59	1.33
Dep. var. SD	2.61	2.71	2.46	0.83
Observations	4,898	5,036	3,407	4,898
Firms	146	146	133	146
Adj. R-sq	0.963	0.980	0.976	0.769
Panel B: Non-Consumer				
	(1)	(2)	(3)	(4)
	IHS MktVal	IHS Rev	IHS R&D	IHS Tobin's Q
EU × Post × Count	0.007***	0.005**	0.006**	-0.003
	(0.003)	(0.002)	(0.002)	(0.002)
Dep. var. mean	8.41	6.49	3.79	1.18
Dep. var. SD	2.38	2.26	2.03	0.58
Observations	5,618	5,777	4,670	5,618
Firms	167	167	157	167
Adj. R-sq	0.968	0.980	0.968	0.749

Balanced panel (2015q1–2019q4). Firm + Quarter-by-SIC2 FE. SEs clustered by firm.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.8: Patent Outcomes (Poisson) — Binary Treatment, Balanced Panel

Panel A: Consumer					
	(1) Patents	(2) AI/ML	(3) CW Pat	(4) CW AI/ML	(5) Val Pat ($)
EU × Post × Data	0.034	-0.454***	-0.409	-2.225***	-0.126
	(0.304)	(0.127)	(0.470)	(0.229)	(0.311)
Dep. var. mean	229	27	432	1639	4463
Dep. var. SD	856	113	2514	4130	14696
Observations	920	343	895	275	920
Firms	105	46	102	37	105
Panel B: Non-Consumer					
	(1) Patents	(2) AI/ML	(3) CW Pat	(4) CW AI/ML	(5) Val Pat ($)
EU × Post × Data	0.408	2.393***	0.597*	2.755**	-0.298
	(0.295)	(0.891)	(0.360)	(1.087)	(0.369)
Dep. var. mean	148	5	101	270	2211
Dep. var. SD	414	17	296	794	10395
Observations	1,215	461	1,180	362	1,215
Firms	139	55	134	42	139

Poisson PML (ppmlhdfe). Balanced panel (2015q1–2019q4). Firm + Year-by-SIC2 FE. SEs clustered by firm.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.9: Patent Outcomes (Poisson) — Count Treatment, Balanced Panel

Panel A: Consumer					
	(1)	(2)	(3)	(4)	(5)
	Patents	AI/ML	CW Pat	CW AI/ML	Val Pat ($)
EU × Post × Count	0.024***	-0.005***	0.005	0.003*	-0.005
	(0.005)	(0.001)	(0.017)	(0.002)	(0.014)
Dep. var. mean	229	27	432	1639	4463
Dep. var. SD	856	113	2514	4130	14696
Observations	920	343	895	275	920
Firms	105	46	102	37	105
Panel B: Non-Consumer					
	(1)	(2)	(3)	(4)	(5)
	Patents	AI/ML	CW Pat	CW AI/ML	Val Pat ($)
EU × Post × Count	0.003	0.008	0.006*	-0.021	0.006
	(0.003)	(0.012)	(0.003)	(0.018)	(0.004)
Dep. var. mean	148	5	101	270	2211
Dep. var. SD	414	17	296	794	10395
Observations	1,215	461	1,180	362	1,215
Firms	139	55	134	42	139

Poisson PML (ppmlhdfe). Balanced panel (2015q1–2019q4). Firm + Year-by-SIC2 FE. SEs clustered by firm.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.10: Patent Outcomes (Poisson) — Binary, Unbalanced, Excl. Top 5 Patenters

Panel A: Consumer					
	(1)	(2)	(3)	(4)	(5)
	Patents	AI/ML	CW Pat	CW AI/ML	Val Pat ($)
EU × Post × Data	-0.216	0.200	-0.368	-2.659	-0.121
	(0.331)	(1.936)	(0.338)	(1.971)	(0.424)
Dep. var. mean	80	5	318	321	2484
Dep. var. SD	164	16	2490	1082	6986
Observations	891	284	866	216	891
Firms	103	39	100	30	103
Panel B: Non-Consumer					
	(1)	(2)	(3)	(4)	(5)
	Patents	AI/ML	CW Pat	CW AI/ML	Val Pat ($)
EU × Post × Data	0.473	-0.454	0.383	-2.610	-0.548*
	(0.451)	(1.530)	(0.421)	(13.203)	(0.324)
Dep. var. mean	76	2	55	155	1115
Dep. var. SD	184	9	134	607	3838
Observations	1,207	420	1,164	293	1,207
Firms	140	51	134	38	140

Poisson PML. Unbalanced panel. Top 5 patenters excluded separately per panel.

Firm + Year-by-SIC2 FE. SEs clustered by firm.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$