# Privacy Laws and Value of Personal Data

Mehmet Canayaz[*], Ilja Kantorovitch[†] and Roxana Mihet[‡]

December 15, 2021

**Abstract**

We analyze how the adoption of the California Consumer Protection Act (CCPA), which limits buying or selling consumer data, heterogeneously affects firms with and without previously gathered data on consumers. Exploiting a novel and hand-collected data set of 11,436 conversational-AI firms with rich personal data on identifiable U.S. consumers, we find that the CCPA gives a strong protection and advantage to *firms* with in-house data on consumers. First, products of these firms experience significant appreciations in customer ratings and are able to collect more customer data relative to their competitors after the adoption of the CCPA. Second, publicly traded firms with in-house data exhibit higher valuations, profitability, asset utilization, and they invest more after the adoption of the CCPA. Third, earnings of such firms can be more accurately predicted by analysts. To rationalize these empirical findings, we build a general equilibrium model where firms produce final goods using labor and data in the form of intangible capital, which can be traded with other firms subject to an iceberg transportation cost. When the introduction of the CCPA increases the transportation cost, firms without in-house data suffer the most because they cannot adequately substitute the previously externally purchased data, while firms with in-house data expand their market share.

**Keywords**: Privacy, Voice Data, In-House Data, Big Data, Intangible Capital
**JEL-Codes:** D80, G30, G31, G38, L20, O30

---

[*]Smeal College of Business, Penn State. Contact: mcanayaz@psu.edu
[†]Department of Economics, EPFL. Contact: ilja.kantorovitch@epfl.ch
[‡]Department of Business and Economics, UNIL and SFI. Contact: roxana.mihet@unil.ch.
*Acknowledgements*: We thank seminar participants at the University of Lausanne and ESSEC for their invaluable feedback, as well as Simona Abis, Avi Goldfarb, and Laura Veldkamp for useful discussions and suggestions.

# 1 Introduction

Today's firms gather vast amounts of consumer data to produce better and more innovative products, predict customer demand more accurately, increase operational efficiency, expand into new markets, and design well-targeted and more profitable marketing campaigns.[1] The use of consumer data however also raises important challenges. One first order concern is privacy. In general, too strict data protection laws may hinder firms, whereas full transparency of consumer data disregards consumer privacy with possible discriminatory outcomes (Acquisti et al., 2016). Furthermore, if data protection laws are not timely they may have unintended consequences, as firms that have a lead on collecting and utilizing previously-collected consumer data can be advantaged after collection of data or its purchase from third parties are restricted.

It has proven difficult to test the systematic interplay between consumer privacy regulations and business outcomes for firms, as lack of representative data has plagued studies that aim to provide large scale empirical evidence. Studies on the subject concentrate mainly on publicly traded firms and extreme events such as data leaks. We know little about the entire population of firms that have access to valuable personal data on customers. We also know little about consumer attitudes towards the privacy of their sensitive data, how businesses navigate the markets for personal data or collect personal data on their customers, and how these parties respond to and get affected by changing consumer privacy laws.

We fill this gap in the literature by exploiting a novel and hand-collected data set of 11,436 conversational-AI firms that have access to detailed voice-generated data on U.S. consumers. These firms operate through intelligent personal assistants, computers and mobile devices. They have continuous and almost unlimited access to rich personal data through 19,334 unique products. They continuously follow and listen to U.S. consumers, exploit recorded data about them, and receive instantaneous feedback from them through speech recognition and natural language understanding technologies. Since these firms can profile customers through their voices, they can build precise

[1] According to the International Data Corporation (IDC), global spending on big data and analytics solutions will increase to over 215 billion US Dollars in 2021, a 10% increase from 2020. The IDC forecasts a further compound annual growth rate of 12.8% between 2021 and 2025 (Vesset and George, 2021). This reflects an increase in the volume of data globally generated data, which increased exponentially and reached 33 Zetabytes in 2018 with a projected volume of 175 Zetabytes in 2025 (Rydning, 2018).

customer profiles and infer, even if they may not directly collect, very personal information such as customer gender, personal habits, and even ethnic background, which other businesses may not have access to.

We analyze how the adoption of the California Consumer Protection Act (CCPA) impacted these previously unstudied firms. This act, which was introduced in June 2018, gave Californian consumers new rights regarding the information companies collect about them. It came into effect on January 1, 2020. The CCPA is an ideal setting for studying the value of privacy for the following reasons. First, the staggered adoption of the CCPA allows us to separate out the effects of CCPA from the effects of other contemporaneous shocks at the product, firm, and country levels. It therefore allows us to better identify the impact of privacy rules on firm and product outcomes. Second, our sampling period of 2016 to 2021 has been characterized by incredible growth in conversational-AI adoption, which contains the entire staggered adoption of the CCPA.

Our empirical approach draws inferences based on the comparison of customer ratings and comments about conversational-AI products in a given day after controlling for fixed product and therefore firm characteristics. We exploit a panel of 11,627,772 observations at the firm-product-day level. We compare products of conversational-AI firms with and without in-house data. We define firms with in-house data as firms that have gathered more customer feedback (negative or positive) per voice-AI products than their competitors before the CCPA's introduction.

We find that the staggered adoption of the CCPA gives a strong advantage to firms with in-house data as opposed to firms that rely primarily on buying external data to improve their operations. Voice-enabled products of firms with lots of in-house data prior to the CCPA experience significant appreciations in customer ratings and additional feedback, i.e., data, from customers after the CCPA. In particular, we estimate appreciations in customer satisfaction by up to 14% (t-stat=5.61) and customer feedback by up to 1,045% (t-stat = 2.19). These results are robust to controlling for rich fixed-effects structures and sub-sample tests on publicly-traded and private businesses.

We also examine the financial ramifications of in-house data before and after the adoption of the CCPA. Our tests on publicly-traded firms provide additional evidence consistent with the argument that CCPA benefits firms with in-house data. Such firms exhibit 18% increase in firm value relative to the sample mean. They also experience a 4.7% increase in operating margin, a 0.7% increase in capital expenditures to assets, and a 4.1% increase in asset utilization (i.e. sales to assets). Additional analyses suggest that earnings of firms with in-house data became easier to predict as uncertainty about the effects of CCPA is resolved once the CCPA is signed into law. Collectively, our findings suggest that California's consumer protection act protects and benefits a select group of firms that have access to previously collected personal consumer data.

To rationalize our empirical findings, we build a simple theoretical model in which firms use data in the form of intangible capital as input in their production function. Firms can either produce data in-house or buy it externally from other firms, which is subject to iceberg transportation costs representing regulatory and technical challenges in using data to its full potential. Importantly, firms are different along one dimension: Sophisticated firms find it cheap to produce data in-house, whereas naive firms face higher costs and rely more on acquiring external data.

We show that a tightening of privacy regulation in the form of an increase in iceberg transportation costs may increase the profits of sophisticated firms at the expense of naive firms, as the latter find it more costly to substitute externally acquired data. This matches our empirical results, as we expect sophisticated firms to enter the conversational-AI space earlier. In future work, we plan on expanding our theoretical analysis to more closely match the partial non-rival nature of data and intangible capital and consider other aspects, such as dynamic effects or firm entry and exit.

This paper pushes the knowledge frontier in several ways. We gather a first data set on firms that have access to valuable personal information on U.S. consumers in the form of voice-generated data. We quantify the value of voice-generated data using a novel data set of conversational-AI firms and the CCPA as a shock to consumer data acquisition. Additionally, we identify the channels through which firms' profitability and size change in response to data regulation. In particular, we look at their products, investments, asset utilization, and ability to predict consumer preferences. We find that the amount of accumulated in-house data is crucial for business outcomes post-CCPA. Careful theoretical and empirical analyses of these issues inform us not only of the direct impact of data privacy laws on business dynamics and consumer feedback but also of the unintended consequences that a non-unified regulatory environment can have on business competition.

## 2 Literature Review

Questions regarding the economics of data, privacy, and sharing personal information have received some attention in the literature. Theoretically, data can empower growth and innovation because data has the special economic property of non-rivalry (Jones and Tonetti, 2020; Cong et al., 2021), which means that data can be used by any number of firms simultaneously without losing its value, e.g., for the training of algorithms. Despite the economic benefits, the gathering, processing, and sharing of personal information are subject to data regulations due to the threat of identity theft (Acquisti et al., 2016) and to protect privacy which is valued by consumers (Tang, 2019).

To the best of our knowledge, we are the first to quantify the economic value of voice-generated data using a unique data set of conversational-AI firms and the California Consumer Protection Act as a shock to firms' ability to purchase consumer personal data. We choose to focus on firms in the conversational-AI space because these firms inherently rely on consumer data to improve their products. Moreover, we exploit the introduction of the CCPA as it is a data privacy regulation that accounts for best data practices within the newest technologies. Older data privacy laws such as the Family Education Rights and Privacy Act (FERPA) and the Health Information Portability and Accountability Act (HIPAA), studied by Khansa et al. (2012) among others, are still important and useful, but they do not account for newer information technologies or for the different ways that data is being used to predict consumer preferences or to price-discriminate. While event studies have been conducted to explore the market effect of security and privacy breaches on firms, none has attempted to determine the impact, in terms of resulting market reactions, of the CCPA legislation itself.

We contribute to the literature that studies the impact of the newest data regulations on firm performance. We focus on the CCPA instead of the more studied GDPR as CCPA's staggered implementation, regional nature, and the fact that it only applies to Californian residents' information offer more variation to establish the causal relationship between data regulation and firm outcomes. It has been reported that the GDPR had a negative impact on the performance of firms in the airline industry (Aridor et al., 2020), and on European firms' ability to attract investment (Jia et al., 2021). Moreover, the GDPR seems to have negatively impacted innovation, since AI startups are re-allocating their limited resources to deal with the implications of the GDPR (Bessen et al., 2020). However, some studies find a neutral or even positive effect of the GDPR. Godinho de Matos and Adjerid (2021) show that consumers' opt-in decisions for a large European telecommunications provider have increased after the GDPR, leading to an increase in sales due to more effective targeted advertising. We find a similar effect of higher profitability and higher customer ratings in the aftermath of CCPA, but only for firms with more in-house data.

Our analysis contributes to the debate on whether the use of new information technologies such as AI and big data distorts competition dynamics and ends up creating winner-takes-all effects. The theoretical literature has warned about the emergence of data-feedback loops that allow large firms to grow even larger (Farboodi et al., 2019) for a variety of reason. For instance, the use of big data in financial markets can lower the cost of capital for large firms, allowing them to grow even larger (Begenau et al., 2018). Moreover, large firms can process more data relative to small ones, enabling them to produce more efficiently and grow even bigger (Farboodi and Veldkamp, 2020; Hagiu and Wright, 2020). Data may also allow large firms to expand into new markets (Vives and Ye, 2021) more easily. Indeed, some of these theoretical effects have been verified empirically as well. For example, Babina et al. (2021) find that firms that invest more in AI experience faster growth in sales and employment, and the AI growth effects concentrate among the ex-ante largest firms, leading to higher industry concentration and reinforcing winner-take-most dynamics. Moreover, Hoberg and Phillips (2021) find that over the past years, U.S. firms have expanded their scope of operations. Increases in scope and scale were achieved largely without increasing traditional operating segments, but rather through scope expansion which were primarily realized through acquisitions and investment in R&D, but not through capital expenditures.

Data regulations could put an end to these winner-take-most effects, ameliorate market competition and avoid other harmful societal effects of unregulated AI (Acemoglu, 2021). Our paper shows that data restrictions have – in fact – the opposite effect: The CCPA hurts firms without in-house data the most, as it restricts the ability of firms to buy external data (for a model of data intermediation, see Bergemann et al., 2020). Firms with in-house data already have deep knowledge about their typical customers, so data restrictions do not hurt the performance of their AI-algorithms as much. As a result, firms with in-house data expand their market share. This is in line with the theoretical findings of Eeckhout and Veldkamp (2021) who warn that big firms are using data to reallocate production to the goods consumers want most. As large firms already have this data, it is easier for them, relative to young, small firms, to tailor their products to consumers' preferences. Our work is in line with Peukert et al. (2021) and Campbell et al. (2015), who show that data regulation can create barriers to entry and may thus hurt competition. In particular, Campbell et al. (2015) show that though privacy regulation imposes costs on all firms, it is small firms and new firms that are most adversely affected, particularly for goods where the price mechanism does not mediate the effect, such as the advertising-supported internet. Similarly, Peukert et al. (2021) find that while all firms suffer losses from the GDPR, the dominant vendor, Google, loses relatively less and can significantly increase market share in important markets such as advertising and analytics.

# 3 Motivation and Descriptive Statistics

This sections provides an overview of the CCPA and the conversational-AI industry, and presents our data. We construct the key variables in our empirical analysis from various *unstructured big data sets*. These variables are: (1) Firms' AI investments. We measure this by looking at firm entry in the voice-AI space and the consumer ratings of their voice-AI applications; we *scrape* this information from Amazon's U.S. website; (2) Firm valuations, sales, investments, and other financial information collected from CRSP and Compustat.

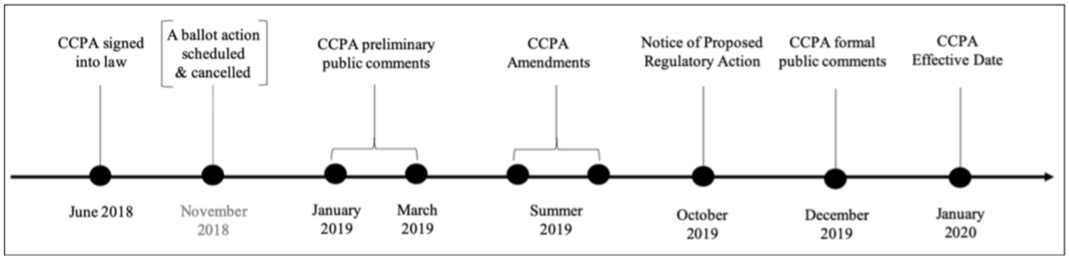
## 3.1 The California Consumer Privacy Act

We exploit a specific regional regulatory shock: the introduction of the California Consumer Protection Act in the United States. The California Consumer Privacy Act (CCPA), a broad based law protecting information that identifies California residents, was introduced June $28^{th}$, 2018 and became effective January $1^{st}$, 2020. Figure 1 shows the CCPA timeline. Having gone into effect in January 2020, the act applies to all companies that serve California residents, neglecting whether the company is based or has a physical presence in the state. In addition, these firms must have at least \$25 million in annual revenue, personal data on at least $50,000$ people or collect more than half of their revenues from the sale of personal data.[2] An amendment, made in April 2020,

[2]See, e.g.,the official CCPA website.

exempts "insurance institutions, agents, and support organizations" from the law due to the fact that they are already subject to similar regulations under California's Insurance Information and Privacy Protection Act (IIPPA).

Figure 1: **CCPA Timeline:** From Signing into Law to Effective Data




Similar to EU's GDPR, the CCPA dramatically alters the way U.S.-based companies process data. The law includes detailed disclosure requirements, provides individuals with extensive rights to control how their personal information is used, imposes statutory fines and creates a private right of action.[3] CCPA defines 'personal information' much more broadly than it is defined under most U.S. privacy laws. It is defined as any information that could reasonably be linked to a particular person or household, whether directly or indirectly. This includes real name, physical address, biometric information, IP address, online identifier, licence number, passport number, race, records of purchasing history or tendencies, internet browsing and search history, geolocation data, audio data, employment, or education data, as well as inferences drawn from these.

## 3.2 Conversational-AI and Customer Privacy

Amazon is the largest online retailer in the U.S., generating more than $457 billion in sales in 2021.[4] Founded in July 1994, the company first launched its digital assistant, the Alexa smart

[3]The CCPA gives all California residents the right to ask any business what personal information they have about them and what they do with that information, they can further ask businesses to delete their personal information or not to sell it to third parties. Additionally, they have to be notified before personal information is collected and they cannot be discriminated upon for exercising these rights, which importantly cannot be waived through contracting. California's Attorney General is in charge of acting against corporations in breach of these rules. Customers can further act directly against corporations when their private information is divulged following a data breach that the firm had not adequately operated to prevent.

[4]See, e.g., macrotrends.net In terms of sales, Amazon leads in global smart speaker sales and continues to expand its lead over Google and Apple, according to VoiceBot.AI. Amazon sold 16.5 million smart speakers and smart displays during the period 2019-2020, followed by Google with 13.2 million, Baidu at 6.6 million, and Alibaba with 6.3 million. Apple came in the fifth slot with 4.6 million smart speakers sold in the fourth quarter.

speaker, in 2014, which was made available to the general public in 2015, three years before CCPA was announced and five years before CCPA was passed. Amazon's Alexa dominates the market for voice-AI around the globe. In 2018, Amazon Alexa had a 72% market share, compared to 18.4% by Google in the U.S., where the market for smart speakers has been growing at a 30-40% annual rate. An analysis from Voicebot.ai shows that the number of adults using smart-speakers grew from 47 million in 2018 to 90 million in 2020, which is approximately 35% of the U.S. population.

Alexa's Skills, i.e., voice-AI products, allow customers to use their voices to perform everyday tasks such as checking the news, listening to music, playing games, shopping, accessing news services, scheduling transportation, or controlling smart home devices and other utilities (Gearbrain, 2018). Companies and individuals can publish Skills in the Alexa Skills Store to reach and entertain users of Alexa devices. Amazon made Alexa's Application Programming Interface (API) available to developers, allowing for integration in non-Amazon devices. Developers can interact with Alexa either through developing Alexa Skills, integrating Alexa with third-party hardware, or adding Alexa support to IoT hardware. Businesses can use Alexa by making use of existing Alexa capabilities to accomplish business tasks or build new Skills and integrations through the "Alexa for Business" platform. In short, Alexa platform allows businesses to "always listen" to customers from Amazon devices or third-party tablets or phones, and collect extremely valuable information about the customers. These data often is reflected in better products for the customers but is also often interpreted as violations of privacy.

## 3.3 Descriptive Statistics

We construct a panel of firms and voice-AI products (Alexa Skills) by scraping all of Amazon's Alexa universe between January 2017 and June 2020. Our U.S. sample contains 11,436 unique firms and 19,334 unique products excluding unrated Skills. We manually match this data to the CRSP-Compustat universe. In so doing, we identify 209 publicly traded corporations that utilize voice-AI products. Our data on Alexa skills is daily and contains 11,627,772 observations. For each Skill, we have a unique product identifier (i.e., ASIN number), the name of its manufacturer, a customer rating between zero and five, and the number of verified customer reviews.

Panel A of Table 1 reports summary statistics on the voice-AI products in our sample. Customer

Table 1: **Summary Statistics**

This table reports summary statistics on voice-AI products (Panel A) and firms (Panel B). In Panel A, Customer Satisfaction denotes customer rating of firm $i$'s product $j$ on day $t$. Customer Reviews denotes the number of customer reviews about firm $i$'s product $j$ on day $t$. In Panel B, Tobin's Q is assets total plus market value of equity minus book value of equity, all deflated by lagged book value of assets from the lottery year. Operating margin is sales minus cost of goods sold, all deflated by sales. Capital expenditures to assets denotes capital expenditures deflated by lagged book value of assets from the lottery year. Sales to assets is sales deflated by the book value of assets. Cash flow to assets is the sum of the income before extraordinary items, and depreciation and amortization deflated by lagged book value of assets from the lottery year. Debt to Assets refers to long-term debt to lagged book value of assets from the lottery year. Panel A (B) spans the period between January 20th, 2017, and May 7th, 2020 (2015-Q1 and 2021-Q1).

| | **Panel A: Voice-AI Products** | | | | | |
|---|---|---|---|---|---|---|
| | **Panel A.1: All Firms** | | | | | |
| | **N** | **Mean** | **Median** | **Std** | **P5** | **P95** |
| Customer Reviews | 11,627,772 | 43.35 | 2.00 | 731.02 | 1.00 | 71.00 |
| Customer Satisfaction | 11,627,772 | 3.65 | 3.90 | 1.33 | 1.00 | 5.00 |
| | **Panel A.2: Public Firms** | | | | | |
| | **N** | **Mean** | **Median** | **Std** | **P5** | **P95** |
| Customer Reviews | 409,399 | 145.37 | 4.00 | 1262.25 | 1.00 | 400.00 |
| Customer Satisfaction | 409,399 | 3.67 | 3.80 | 1.17 | 1.60 | 5.00 |
| | **Panel A.3: Private Firms** | | | | | |
| | **N** | **Mean** | **Median** | **Std** | **P5** | **P95** |
| Customer Reviews | 11,213,159 | 39.58 | 2.00 | 703.96 | 1.00 | 64.00 |
| Customer Satisfaction | 11,213,159 | 3.64 | 3.90 | 1.34 | 1.00 | 5.00 |
| | **Panel B: Firm-level Financial Data** | | | | | |
| | **N** | **Mean** | **Median** | **Std** | **P5** | **P95** |
| Tobin's Q | 1,629 | 1.83 | 1.57 | 0.84 | 0.95 | 3.60 |
| Operating Margin | 1,611 | 0.16 | 0.17 | 0.16 | -0.13 | 0.41 |
| Capital expenditures to assets | 1,613 | 0.02 | 0.01 | 0.02 | 0.00 | 0.06 |
| Sales to assets | 1,629 | 0.20 | 0.15 | 0.15 | 0.03 | 0.50 |
| Cash Flows to Assets | 1,554 | 0.02 | 0.02 | 0.03 | -0.04 | 0.06 |
| Debt to Assets | 1,556 | 0.28 | 0.27 | 0.21 | 0.00 | 0.66 |
| In-house Data | 1,629 | 0.47 | 0.00 | 0.50 | 0.00 | 1.00 |

Satisfaction denotes customer ratings of firm $i$'s product $j$ on day $t$. Customer Reviews denotes the number of customer reviews about firm $i$'s product $j$ on day $t$. As shown in the Panel A.1, the average product has 43.35 reviews and a customer rating of 3.65 out of five. Products of publicly traded firms have more customer feedback than those of private firms, but products of these two groups of firms have comparable customer ratings. As shown in Panels A.2 and A.3 of Table 1, a given product of public (private) firms has 145.37 (39.58) reviews and a rating of 3.67 (3.64) out of five on average.

Average consumer ratings provide important information about customer satisfaction, but their reliability depends on the sample size. The concept of statistical power suggests that, when inferring product qualities, customers should not only look at the average rating of the Skill but also the number of people who rated it, as well as the dispersion of the raters' judgments, which provide important information (Obrecht, Chapman and Gelman, 2007). For example, an average rating of two stars given by 400 consumers is less noisy than an average rating of two given by five consumers. We therefore utilize both of these measures when we analyze the influence of CCPA on businesses.

Panel B of Table 1 reports summary statistics on publicly traded firms that utilize voice-AI products. Our data on their financials is quarterly and spans the period between 2015-Q1 and 2021-Q1. As shown, mean Tobin's Q equals 1.83, operating margin equals 16%, capital expenditures to assets ratio equals 2% and sales to assets ratio equals 20%. In addition to these variables, we also report summary statistics on our proxy for a firm's in-house data advantage. To calculate this proxy we drop all observations after 2018-Q1 and take each firm's average number of reviews per voice-AI product. Our variable, In-house Data is equal to one if firm i has more customer feedback per voice-AI products than the sample median. Our measure therefore incorporates good and bad reviews, which can be utilized by firms as in-house data on consumer feedback.

# 4 Empirical Framework

This section provides information on the main empirical specification used in our analyses. To study the relation between product-level outcomes in the voice-AI space and the staggered adoption of the California Consumer Privacy Act (CCPA) in the United States, we run regressions on the below specification:

$$\begin{aligned} Y_{i,j,t} = \alpha \; + \; \beta_1 \; * \; \text{CCPA Introduced}_t * \text{In-House Data}_i \; + \; \beta_2 \; * \; \text{CCPA Effective}_t * \text{In-House Data}_i \; + \\ + \; \gamma_{i,j} \; + \; \phi_t \; + \; \epsilon_{i,j,t} \end{aligned} \tag{1}$$

where $Y_{i,j,t}$ denotes customer satisfaction with or customer feedback on voice-AI product $j$ of firm $i$ in calendar day $t$. Customer Satisfaction denotes customer rating of firm $i$'s product $j$ on day $t$. Customer Reviews denotes the number of customer reviews about firm $i$'s product $j$ on day $t$. In-house Data is equal to one if firm $i$ has more customer feedback per voice-AI products than the sample median before the introduction of the CCPA. $\text{CCPA Introduced}_t$ is equal to one between 2018Q3 and 2020Q1, and zero otherwise. $\text{CCPA Effective}_t$ is equal to one after 2020Q1 and zero otherwise. $\gamma_{i,j}$ denotes firm-product fixed effects and $\phi_t$ denotes year-month-day fixed effects. We cluster standard errors by firm id for private firms and industry (i.e., 2-digit sic codes) for public firms. Our results on products of public firms are generally robust to clustering at the firm-level.

There are multiple benefits to using the above empirical specification. With firm-product fixed effects we separate out the effects of the staggered CCPA adoption from the potential effects of contemporaneous shocks at firm and product levels. This is helpful, because CCPA can impact some firms more than others. By introducing daily fixed effects, we eliminate any daily trend that may be a confounder to the relation between CCPA laws and product outcomes. Related to this concern, the staggered adoption of CCPA helps us with the identification of privacy rules on voice-AI product outcomes.

# 5 Main Empirical Findings

This section presents the main empirical findings of our paper. We start with demonstrating the ramifications of CCPA on voice-AI products of U.S. businesses in subsection 5.1. In subsection 5.2, we report how CCPA influences firm level outcomes in the U.S. Subsection 5.3 presents how CCPA influences the accuracy of earnings predictions in the voice-AI space.

## 5.1 Customer Privacy Laws, Product Quality and Customer Satisfaction

We start our analyses with investigating the relation between staggered adoption of the CCPA and product outcomes following specification (1). The estimated coefficients of interest are the ones on $\text{CCPA Introduced}_t * \text{In-House Data}_i$ and $\text{CCPA Effective}_t * \text{In-House Data}_i$, which denote whether firm $i$ has built more in-house data than its peers before the adoption of the CCPA. We present our findings in Table 2.

As shown in Column 1 of Table 2's Panel A, after controlling for Firm × Voice-AI Product fixed effects, we find that a given U.S. firm without in-house data attains around 0.075 lower ratings and a given U.S. firm with in-house data attains around 0.030 unit higher ratings after the introduction of CCPA. These correspond to -2.5% and 0.82% changes relative to the sample mean of 3.65. After CCPA becomes effective, a given U.S. firm without in-house data attains around 0.095 lower ratings and a given U.S. firm with in-house data attains around 0.042 unit higher ratings, which correspond to 2.60% and 1.15% changes, respectively. As shown in Column 2 of Panel A, these findings are statistically robust to and economically larger after controlling also for daily fixed effects.

Columns 3-4 of Panel A present our findings on customer reviews. Firms with in-house data also attain 278.712 (102.091) additional reviews after CCPA is effective (introduced). These correspond to 542% and 135% increases relative to the sample mean of 43.35 reviews. Panels B and C of Table 2 provide further evidence on the subsamples of firms, i.e., private and public firms. As shown in columns 2 and 4 of these panels, after controlling for firm-product and daily fixed effects, we pin down increases of 0.314 and 0.526 units in customer ratings and 352.477 and 1,519.118 units in number of reviews for firms with in-house data.

Our findings in this subsection are consistent with the argument that the staggered adoption of the CCPA gave a strong advantage to firms with in-house data. Voice-AI products of these firms experience significant appreciations in customer ratings and additional feedback, i.e., data, from customers after the CCPA. These results are robust to firm-product and daily fixed effects, and subsample tests on public and private businesses. The following subsections delve into financial ramifications of voice-AI products and the adoption of the CCPA.

## 5.2 Firm-level Ramifications of the CCPA

In this subsection, we investigate the effects of CCPA adoption on firm-level metrics. To motivate the potential ramifications of CCPA on firms with in-house data advantage, we compare the valuations of firms with in-house data relative to synthetic control firms in calendar time. As shown in Figure 2, the average valuation of firms with in-house data strictly dominates synthetic control firms after the adoption of the CCPA.

Table 2: **In-house data advantage after the adoption of consumer privacy rules**

This table reports regressions of product-level outcomes on the staggered adoption of the California Consumer Privacy Act (CCPA) in the United States. We compare products of firms with and without in-house data on customers before the introduction of the CCPA. In Columns 1–4, we run regressions on the following specification

$$Y_{i,j,t} = \alpha \; + \; \beta_1 \; * \; \text{CCPA Introduced}_t * \text{In-House Data}_i \; + \; \beta_2 \; * \; \text{CCPA Effective}_t * \text{In-House Data}_i \; + \\ + \; \gamma_{i,j} \; + \; \phi_t \; + \; \epsilon_{i,j,t}$$

where $Y_{i,j,t}$ denotes customer satisfaction with or customer feedback on voice-AI product $j$ of firm $i$ in calendar day $t$. Customer Satisfaction denotes customer rating of firm $i$'s product $j$ on day $t$. Customer Reviews denotes the number of customer reviews about firm $i$'s product $j$ on day $t$. In-house Data is equal to one if firm $i$ has more customer feedback per voice-AI products than the sample median before the introduction of the CCPA. CCPA Introduced$_t$ (CCPA Effective$_t$) is equal to one after 2018Q3 (2020Q1) and zero otherwise. $\gamma_{i,j}$ denotes firm-product fixed effects and $\phi_t$ denotes year-month-day fixed effects. We study all firms, public firms, and private firms respectively in Panels A-C. The data spans the period between January 20, 2017, and May 7, 2020. Standard errors are clustered at the industry (firm) level in Panel B (A and C). $\star\star\star$, $\star\star$, or $\star$ indicates that the coefficient estimate is significantly different from zero at the 1%, 5%, or 10% level, respectively.

| | **Panel A: All Firms** | | | |
|---|---|---|---|---|
| | **Customer Satisfaction** | **Customer Satisfaction** | **Customer Reviews** | **Customer Reviews** |
| | (1) | (2) | (3) | (4) |
| CCPA Introduced | -0.075*** (-8.99) | ... | 4.322*** (3.51) | ... |
| CCPA Introduced x In-house Data | 0.030** (2.47) | 0.046*** (3.80) | 112.105*** (4.92) | 102.091*** (5.15) |
| CCPA Effective | -0.095*** (-8.18) | ... | 23.992*** (5.76) | ... |
| CCPA Effective x In-house Data | 0.042** (2.38) | 0.058*** (3.19) | 288.916*** (5.28) | 278.712*** (5.30) |
| *Fixed Effects* | | | | |
| Firm × Voice-AI Product | Yes | Yes | Yes | Yes |
| Day | No | Yes | No | Yes |
| Adj. R-squared | 0.929 | 0.930 | 0.362 | 0.362 |
| Observations | 8,428,309 | 8,428,309 | 8,428,309 | 8,428,309 |

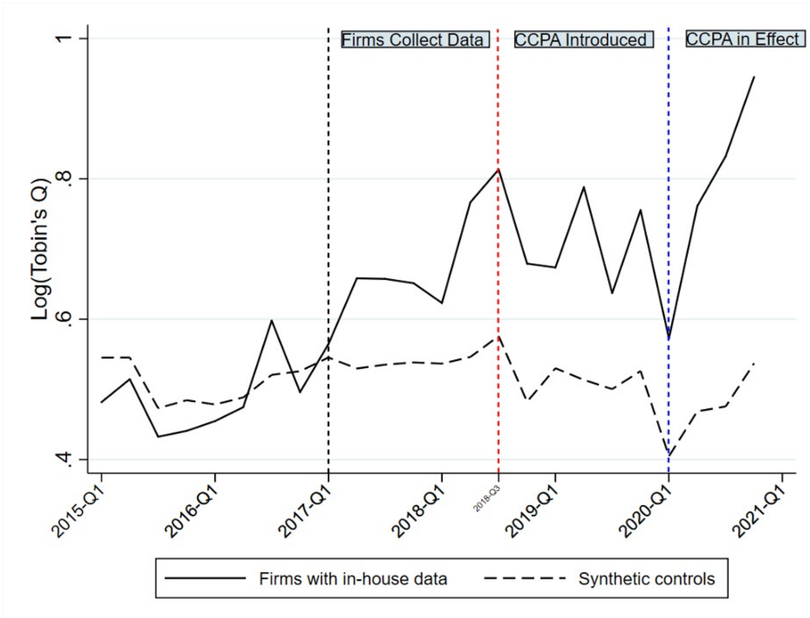
Table 2: **Cont. In-house data advantage after the adoption of consumer privacy rules**

| | **Panel B: Public Firms** | | | |
|---|---|---|---|---|
| | **Customer Satisfaction** | **Customer Satisfaction** | **Customer Reviews** | **Customer Reviews** |
| | (1) | (2) | (3) | (4) |
| CCPA Introduced | -0.107*** | ... | 82.372 | ... |
| | (-5.43) | | (1.26) | |
| CCPA Introduced x In-house Data | 0.279*** | 0.314*** | 430.524 | 352.477 |
| | (4.48) | (4.38) | (1.66) | (1.48) |
| CCPA Effective | -0.180*** | ... | 162.705 | ... |
| | (-6.78) | | (1.12) | |
| CCPA Effective x In-house Data | 0.492*** | 0.526*** | 1,598.832** | 1,519.118** |
| | (5.46) | (5.61) | (2.25) | (2.19) |
| *Fixed Effects* | | | | |
| Firm × Voice-AI Product | Yes | Yes | Yes | Yes |
| Day | No | Yes | No | Yes |
| Adj. R-squared | 0.903 | 0.903 | 0.315 | 0.316 |
| Observations | 364,959 | 364,959 | 364,959 | 364,959 |

| | **Panel C: Private Firms** | | | |
|---|---|---|---|---|
| | **Customer Satisfaction** | **Customer Satisfaction** | **Customer Reviews** | **Customer Reviews** |
| | (1) | (2) | (3) | (4) |
| CCPA Introduced | -0.073*** | ... | 4.346*** | ... |
| | (-8.64) | | (3.46) | |
| CCPA Introduced x In-house Data | 0.022* | 0.038*** | 100.748*** | 92.432*** |
| | (1.78) | (3.05) | (4.34) | (4.55) |
| CCPA Effective | -0.090*** | ... | 24.728*** | ... |
| | (-7.67) | | (5.72) | |
| CCPA Effective x In-house Data | 0.031* | 0.047** | 252.303*** | 243.874*** |
| | (1.72) | (2.52) | (4.68) | (4.70) |
| *Fixed Effects* | | | | |
| Firm × Voice-AI Product | Yes | Yes | Yes | Yes |
| Day | No | Yes | No | Yes |
| Adj. R-squared | 0.931 | 0.931 | 0.369 | 0.369 |
| Observations | 8,064,490 | 8,064,490 | 8,064,490 | 8,064,490 |

Figure 2: **Valuations of firms with in-house data and synthetic controls**

This figure shows the valuations of firms with in-house data relative to synthetic control firms. The x-axis denotes year-quarters around the staggered adoption of the CCPA. The y-axis shows the average logarithm of Tobin's Q of firms with in-house data and synthetic control firms in a given year-quarter. The synthetic match is done using logged Tobin's Q and logged book value of assets between 2015-Q1 and 2016-Q4 following the methodologies of Abadie and Gardeazabal (2003) and Abadie, Diamond, and Hainmueller (2010, 2014).




Motivated by Figure 2, we provide more direct evidence on financial characteristics of voice-AI companies. To do so, we run regressions on the below specification:

$$Y_{i,t} = \alpha \ + \ \beta_1 \ * \ \text{CCPA Introduced}_t * \text{In-House Data}_i \ + \ \beta_2 \ * \ \text{CCPA Effective}_t * \text{In-House Data}_i \ + \\ + \ \gamma_i \ + \ \phi_{j,t} \ + \ \epsilon_{i,j,t} \qquad (2)$$

where $Y_{i,j}$ denotes Tobin's Q, Operating Margin, Capital expenditures to assets, or Sales to assets of firm $i$ in year-quarter $t$. $\gamma_i$ denotes firm fixed effects and $\phi_{j,t}$ denotes industry-year fixed effects, where $j$ refers to firm $i$'s 2-digit SIC code. The estimated coefficients of interest are $\beta_1$ and $\beta_2$.

We present our findings in Table 4. The results in Column 1 suggest that firms with in-house data attain 0.323 units higher Tobin's Q after the introduction of CCPA and 0.326 units higher Tobin's Q after the full adoption of CCPA. Compared with sample means reported in Table 2, the

Table 4: **Firm-level ramifications of in-house data advantage**

This table reports regressions of firm-level outcome variables on the staggered adoption of California Consumer Privacy Act (CCPA) in the United States. In Columns 1–4, we run regressions on the following specification:

$$Y_{i,t} = \alpha \ + \ \beta_1 \ * \ \text{CCPA Introduced}_t * \text{In-House Data}_i \ + \ \beta_2 \ * \ \text{CCPA Effective}_t * \text{In-House Data}_i \ + \\ + \ \gamma_i \ + \ \phi_{j,t} \ + \ \epsilon_{i,j,t}$$

where $Y_{i,t}$ denotes *Tobin's Q*, *Operating Margin*, *Capital expenditures to assets*, or *Sales to assets* of firm $i$ in *year-quarter t*. *In-house Data*$_i$ is equal to one if firm $i$ has more customer feedback per voice-AI products than the sample median before the introduction of the CCPA. CCPA Introduced$_t$ (CCPA Effective$_t$) is equal to one after 2018Q3 (2020Q1) and zero otherwise. $\gamma_i$ denotes firm fixed effects and $\phi_{j,t}$ denotes industry-year fixed effects, where j refers to firm i's 2-digit SIC code. *Tobin's Q* is assets total plus market value of equity minus book value of equity, all deflated by lagged book value of assets from the lottery year. *Operating margin* is sales minus cost of goods sold, all deflated by sales. *Capital expenditures to assets* denotes capital expenditures deflated by lagged book value of assets from the lottery year. *Sales to assets* is sales deflated by the book value of assets. *Cash flow to assets* is the sum of the income before extraordinary items, and depreciation and amortization deflated by lagged book value of assets from the lottery year. *Debt to Assets* refers to long-term debt to lagged book value of assets from the lottery year. The data spans the period between 2015-Q1 and 2021-Q1. Standard errors are clustered at the industry level. $\star\star\star$, $\star\star$, or $\star$ indicates that the coefficient estimate is significantly different from zero at the 1%, 5%, or 10% level, respectively.

| | **Tobin's Q** | **Operating Margin** | **Capital Expenditures to Assets** | **Sales to Assets** |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| CCPA Introduced x In-house Data | 0.323*** | 0.042 | 0.007* | 0.027** |
| | (3.42) | (1.22) | (2.04) | (2.18) |
| CCPA Effective x In-house Data | 0.326*** | 0.047** | 0.004** | 0.041** |
| | (3.37) | (2.30) | (2.69) | (2.08) |
| *Controls* | | | | |
| Cash flows to Assets | 2.576*** | 1.720*** | -0.020 | 0.625** |
| | (3.03) | (6.86) | (-0.34) | (2.74) |
| Debt to Assets | 0.604 | 0.142* | -0.017* | -0.049 |
| | (1.13) | (1.93) | (-2.04) | (-0.53) |
| *Fixed Effects* | | | | |
| Firm | Yes | Yes | Yes | Yes |
| Industry × Year-Quarter | Yes | Yes | Yes | Yes |
| Adj. R-squared | 0.863 | 0.895 | 0.682 | 0.942 |
| Observations | 1,500 | 1,485 | 1,500 | 1,500 |

estimated coefficient values reflect 18% appreciations in firm value. As shown in Columns 2, 3, 4, and 5 of Table 4, U.S. firms attain up to a 0.047 unit increase in *Operating Margin*, a 0.007 unit increase in capital expenditures to assets, and a 0.041 unit increase in asset utilization, i.e., sales to assets ratio.

These findings provide additional evidence consistent with the argument that CCPA benefits firms with in-house data. Such firms exhibit higher valuations, profitability, asset utilization, and they invest more after the adoption of customer privacy laws.

## 5.3 Earnings Forecasts and CCPA

The previous subsections provide evidence supporting the hypothesis that CCPA benefits firms with in-house data. These firms have better products (proxied with higher customer ratings), are able to collect more data from customers (due to more reviews, on top of their higher product ratings), exhibit higher valuations, higher profitability and higher sales fueled by increases in capital expenditures. In this subsection, we study the ramifications of CCPA on the predictability of firm earnings. In particular, we examine the impact of CCPA adoption on the accuracy of the consensus earnings predictions for firms with and without in-house data. To do so, we merge our data with IBES dataset and run regressions on the following specification:

$$\begin{aligned} Y_{i,t} = \alpha \; + \; \beta_1 \; * \; \text{CCPA Introduced}_t * \text{In-House Data}_i \; + \; \beta_2 \; * \; \text{CCPA Effective}_t * \text{In-House Data}_i \; + \\ + \; \gamma_i \; + \; \phi_{j,t} \; + \; \epsilon_{i,j,t} \end{aligned} \tag{3}$$

where $Y_{i,t}$ denotes *Absolute Consensus Error*, or *St dev of Earnings Forecasts* of firm $i$ in *year-quarter* $t$. *Absolute Consensus Error* denotes the absolute value of the difference between consensus analyst EPS forecasts for firm $i$ in year-quarter $t$, deflated with firm $i$'s stock price from the previous quarter. *St dev of Earnings Forecasts* denotes the standard deviation of analyst EPS forecasts for firm $i$ in year-quarter $t$. Once again, the control variables are as in the next section and the estimated coefficients of interest are $\beta_1$ and $\beta_2$.

We present our results in Table 5. As shown in columns 1 and 2, we find a negative relation between CCPA's introduction in the U.S. and the absolution consensus errors for current earnings

Table 5: **In-house data advantage and the accuracy of firm earnings forecasts**

This table reports regressions of consensus earnings forecast errors on the staggered adoption of California Consumer Privacy Act (CCPA) in the United States. In Columns 1–4, we run regressions on the following specification:

$$Y_{i,t} = \alpha \ + \ \beta_1 \ * \ \text{CCPA Introduced}_t * \text{In-House Data}_i \ + \ \beta_2 \ * \ \text{CCPA Effective}_t * \text{In-House Data}_i \ + \\ + \ \gamma_i \ + \ \phi_{j,t} \ + \ \epsilon_{i,j,t}$$

where $Y_{i,t}$ denotes *Absolute Consensus Error*, or *St dev of Earnings Forecasts* of firm $i$ in *year-quarter* $t$. *Absolute Consensus Error* denotes the absolute value of the difference between consensus analyst EPS forecasts for firm $i$ in year-quarter $t$, deflated with firm $i$'s stock price from the previous quarter. *St dev of Earnings Forecasts* denotes the standard deviation of analyst EPS forecasts for firm $i$ in year-quarter $t$. *Absolute Consensus Error* $t+1$, or *St dev of Earnings Forecasts* $t+1$ refer the next quarter's errors and standard deviations. In-House $\text{Data}_i$ is equal to one if firm i has more customer feedback per voice-AI products than the sample median before the introduction of the CCPA. Voice AI $\text{Introduced}_t$ is equal to one between 2017Q1 and 2018Q2 and zero otherwise. CCPA $\text{Introduced}_t$ (CCPA $\text{Effective}_t$) is equal to one after 2018Q3 (2020Q1) and zero otherwise. $\gamma_i$ denotes firm fixed effects and $\phi_{j,t}$ denotes industry-year fixed effects, where j refers to firm i's 2-digit SIC code. The data spans the period between 2015-Q1 and 2021-Q1. Standard errors are clustered at the industry level. $\star\star\star$, $\star\star$, or $\star$ indicates that the coefficient estimate is significantly different from zero at the 1%, 5%, or 10% level, respectively.

| | **Absolute Consensus Error** | **Absolute Consensus Error+1** | **St dev of Earnings Forecasts** | **St dev of Earnings Forecasts +1** |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Voice AI Introduced x In-house Data | -0.004 | -0.004 | -0.022* | -0.019* |
| | (-1.56) | (-1.41) | (-1.97) | (-1.75) |
| CCPA Introduced x In-house Data | -0.004** | -0.004** | -0.041 | -0.055* |
| | (-2.32) | (-2.28) | (-1.60) | (-1.77) |
| CCPA Effective x In-house Data | -0.004 | -0.005 | -0.086 | -0.144* |
| | (-0.47) | (-0.63) | (-1.55) | (-1.82) |
| *Fixed Effects* | | | | |
| Firm | Yes | Yes | Yes | Yes |
| Industry × Year-Quarter | Yes | Yes | Yes | Yes |
| Adj. R-squared | 0.479 | 0.678 | 0.523 | 0.830 |
| Observations | 1,310 | 1,322 | 1,351 | 1,379 |

predictions and following quarter's earnings predictions. In particular, the absolute consensus EPS prediction error decreases by 0.04% ($t-stat = 2.32$) and 0.04% ($t-stat = 2.28$). As shown in column 4, the standard deviation of consensus EPS forecasts decline by 0.055 units after CCPA is introduced and adopted, although with marginally significant t-stats of $-1.77$ and $-1.82$, respectively. These findings collectively suggest that market participants are able to predict earnings of firms with in-house data more accurately after the adoption of CCPA.

# 6 Theoretical Framework

Our empirical analysis suggests that the staggered adoption of the CCPA gave a strong advantage to firms with in-house data as opposed to firms that rely on external data acquisition. Voice-AI products of firms with lots of in-house data prior to CCPA experience significant appreciations in customer ratings and additional feedback from customers after the CCPA. Moreover, such firms exhibit higher valuations, profitability, asset utilization, and they invest more after the adoption of CCPA. Lastly, their earnings become easier to predict as uncertainty about the effects of CCPA gets resolved once the CCPA is signed into law.

In this section, we build a simple theoretical model to rationalize these findings. We present a simple static model to think about the effects of privacy regulation, such as the CCPA, on the data economy. We focus particularly on restrictions on the buying and selling of data which could be used to improve firm business operations (e.g., training of marketing algorithms that allow to reach potential customers more efficiently).

## 6.1 Households and Consumption

There is a unit mass of households with log utility

$$U = \ln(C), \tag{4}$$

where $C$ is a composite final good and equal to aggregate output $Y$. Households supply each one unit of labor inelastically and earn a wage $W$.

The final good $Y$ consists of two separate goods that are produced by two different types of firms,

$$Y = \left(\gamma\left(Y_A\right)^{\frac{\sigma-1}{\sigma}} + (1-\gamma)\left(Y_B\right)^{\frac{\sigma-1}{\sigma}}\right)^{v\frac{\sigma}{\sigma-1}}, \tag{5}$$

where $Y_A$ is the symmetric output of $A$-firms, $Y_B$ is the symmetric output of $B$-firms and $\gamma$ pins down their relative size. $\sigma > 0$ is the constant elasticity of substitution. $v$ governs the returns to scale. If $v < \frac{\sigma-1}{\sigma}$, the aggregation technology for the final good features diminishing returns, such that firms of one group profit from a reduction in production in the other group. As we view $A$-firms and $B$-firms as competitors, we will focus on that case. The final consumption good $C$ also serves as the numéraire and its price is normalized to one.

## 6.2 Firms

There is a $\gamma$ mass of $A$-firms and a $1-\gamma$ mass of $B$-firms . Since inside each group firms are identical, firm-specific variables will be indexed by $j$ and firm-specific subscripts are dropped. Firms have a Cobb-Douglas production function,

$$Y_j = l_j^{\alpha}\left(k_j^I\right)^{\beta}, \tag{6}$$

where $l_j$ is labor input, $k_j^I$ is intangible capital. Firms that have more intangible capital can produce more using the same amount of labor. Intangible capital may be understood as data that allows to better target marketing campaigns, which increases sales without otherwise needing more input. On the firm level, returns to scale are governed by $\alpha + \beta \leq 1$.

### 6.2.1 Intangible Capital Production

Firms are endowed with a technology for the production of intangible capital subject to a quadratic cost,

$$c^j(k_j^{I,P}) = \frac{\phi^j}{2}(k_j^{I,P})^2. \tag{7}$$

In the following, we will refer to $A$-firms as (data)-*sophisticated* firms and to $B$-firms as (data)-*naive* firms. The crucial and only difference between both types is that *sophisticated* firms are better at producing or gathering data ($\phi^A < \phi^B$) in the form of intangible capital than *naive* firms. Note that $k_j^{I,P}$ stands for the amount of capital *produced* by a $j$-firm, whereas $k_j^I$ stands for the intangible capital *employed* by firm a $j$-firm.

There is a market for intangible capital. Firms can trade intangible capital at a price $p^I$ subject to an iceberg transportation cost $\tau \in (0,1)$, such that for each unit of intangible capital shipped only a fraction $(1-\tau)$ arrives at its destination. The trading friction $\tau$ reflects a number of realistic features of the data economy. For example, firms may not able to employ data as efficiently as the data originator. Moreover, data regulation may make buying and selling data more costly and complicated than using in-house data. Finally, we will assume at the moment that the usage of intangible capital is rival, such that sold intangible capital cannot be used anymore by the data-originating firm.[5]

## 6.3 Firm Problems

We conjecture that $A$-firms sell intangible capital, whereas $B$-firms buy. Taking this into account, the maximization problem of $A$-firms is

$$\max_{l_A, k_A^{I,B}, k_A^{I,P}} \quad p_A l_A^\alpha \left(k_A^{I,P} - k_A^{I,S}\right)^\beta - w l_A + p^I k_A^{I,S} - c(k_A^{I,P}), \tag{8}$$

$$s.t. \quad k_A^{I,P} - k_A^{I,S} \geq 0 \tag{9}$$

where $k_A^{I,S}$ stands for the amount of intangible capital that $A$-firms sell. $A$-firms decide how much labor $l_A$ to employ given a wage $w$, how much intangible capital $k_A^{I,S}$ to sell given its price $p^I$, and how much intangible capital $k_A^{I,P}$ to produce. Finally, $A$-firms can at most sell $k_A^{I,P}$ units of intangible capital.

[5]Although data by itself is non-rival (Jones and Tonetti, 2020), it may be that the use of data to increase profits is at least partially rival. For example, using data to better target advertisement may give a competitive advantage if handled by one firm, but lose its effect if every firm employs similar techniques. A model in which intangible capital is partially non-rival is in the making.

Since $B$-firms buy intangible capital, their maximization problem is

$$\max_{l_B, k_B^{I,B}, k_B^{I,P}} \quad p_B l_B^{\alpha} \left(k_B^{I,P} + k_B^{I,B}\right)^{\beta} - w l_B - \frac{p^I}{1-\tau} k_B^{I,B} - c(k_B^{I,P}), \tag{10}$$

$$s.t. \quad k_B^{I,P} - k_B^{I,S} \geq 0. \tag{11}$$

Due to the iceberg transportation costs $\tau$, $B$-firms have to buy $\frac{1}{1-\tau}$ units of intangible capital to receive one unit, effectively increasing its price. Otherwise the maximization problem for $B$-firms is symmetric to $A$-firms.

## 6.4 Equilibrium

The price of each intermediate good is given by

$$p_A = \frac{\partial Y}{\partial Y_A} = \gamma v Y^{\alpha_Y} \left(Y_A\right)^{-\frac{1}{\sigma}} \tag{12}$$

$$p_B = \frac{\partial Y}{\partial Y_B} = (1-\gamma) v Y^{\alpha_Y} \left(Y_B\right)^{-\frac{1}{\sigma}}, \tag{13}$$

where we denote $\alpha_Y = \frac{\sigma v - \sigma + 1}{\sigma v}$. Note that whenever $v < \frac{\sigma - 1}{\sigma}$, an increase in aggregate output $Y$ has a negative effect on the price of either intermediate good. After plugging the prices of intermediate goods into the firms' problems, the first order conditions for intangible capital production and selling for $A$-firms are

$$\gamma v Y^{\alpha_Y} \hat{\beta} l_A^{\hat{\alpha}} \left(k_A^{I,P} - k_A^{I,S}\right)^{\hat{\beta}-1} = \phi^A k_A^{I,P} \tag{14}$$

$$\gamma v Y^{\alpha_Y} \hat{\beta} l_A^{\hat{\alpha}} \left(k_A^{I,P} - k_A^{I,S}\right)^{\hat{\beta}-1} = p^I, \tag{15}$$

where $\hat{\alpha} = \frac{\sigma-1}{\sigma}\alpha$ and $\hat{\beta} = \frac{\sigma-1}{\sigma}\beta$. Hence, capital production is pinned down to

$$k_A^{I,P} = \frac{p^I}{\phi^A}. \tag{16}$$

Following the same steps, the capital production decision of $B$-firms is equal to

$$k_B^{I,P} = \frac{p^I}{(1-\tau)\phi^B}. \tag{17}$$

Therefore, the price of intangible capital is crucial for the intangible capital production decision. Labor demand for each firm is given by

$$l_j = \left[w^{-1}\hat{\alpha}\gamma v Y^{\alpha_Y}\left(k_j^I\right)^{\hat{\beta}}\right]^{\frac{1}{1-\hat{\alpha}}}. \tag{18}$$

Finally, the market clearing conditions for the capital market and labor market are given by

$$\gamma k_A^{I,S} = (1-\gamma)k_B^{I,B} + \tau\gamma k_A^{I,S} \tag{19}$$

$$\sum_{j\in\{A,B\}} l_j = 1. \tag{20}$$

The amount of intangible capital that is shipped by $A$-firms equals what $B$-firms are buying plus the iceberg transportation costs. Moreover, labor demand by firms needs to equal labor supply that is normalized to one.

## 6.5 Results

In the following, we focus on a calibration with $v < \frac{\sigma-1}{\sigma}$, i.e., $A$- and $B$-firms compete for the same market. In particular, the parameters for the simulations are given by $\gamma = 0.5$, $v = 0.1$, $\sigma = 2$, $\delta = 0.3$, $\phi^A = 0.1$, $\phi^B = 3$. We are primarily interested in the effects of changes in the iceberg transportation costs $\tau$. Privacy protection regulation may make buying and selling data from data brokers more complicated and expensive. Such regulation will have different effects on *sophisticated* and *naive* firms, depending on their ability to produce data in-house. This leads directly to our first result

**Proposition 1** *An increase in iceberg transportation costs $\tau$ leads to*

*(i) a fall in the price of intangible capital $p^I$.*

*(ii) higher production and lower employment of intangible capital by $B$-firms.*

*(iii) lower production and higher employment of intangible capital by A-firms.*

The sketch of the proof goes as follows. An increase in iceberg transportation costs $\tau$ reduces the demand for intangible capital by $B$-firms, which leads to a fall in the price of intangible capital $p^I$. Still, $B$-firms face a higher price for intangible capital. As a result, they decide to produce more themselves at a higher cost, which subsequently decreases how much intangible capital $B$-firms employ. The reverse is true for $A$-firms. Because selling intangible capital becomes less attractive, they decide to produce less but employ more of it inside their own firm. This result is illustrated in Figures 3 and 4.

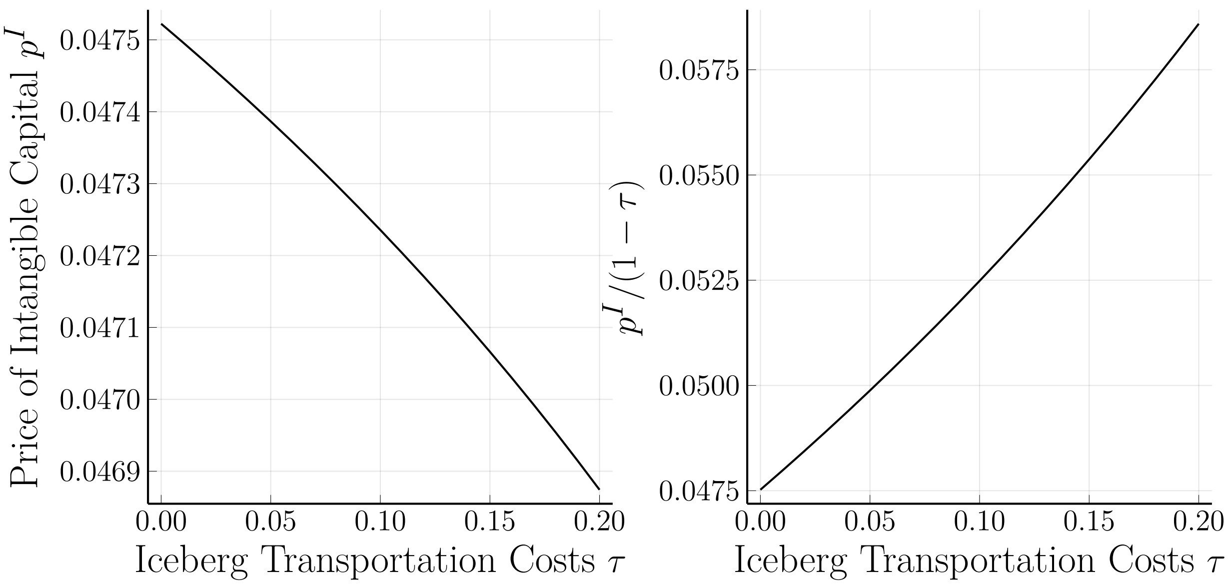
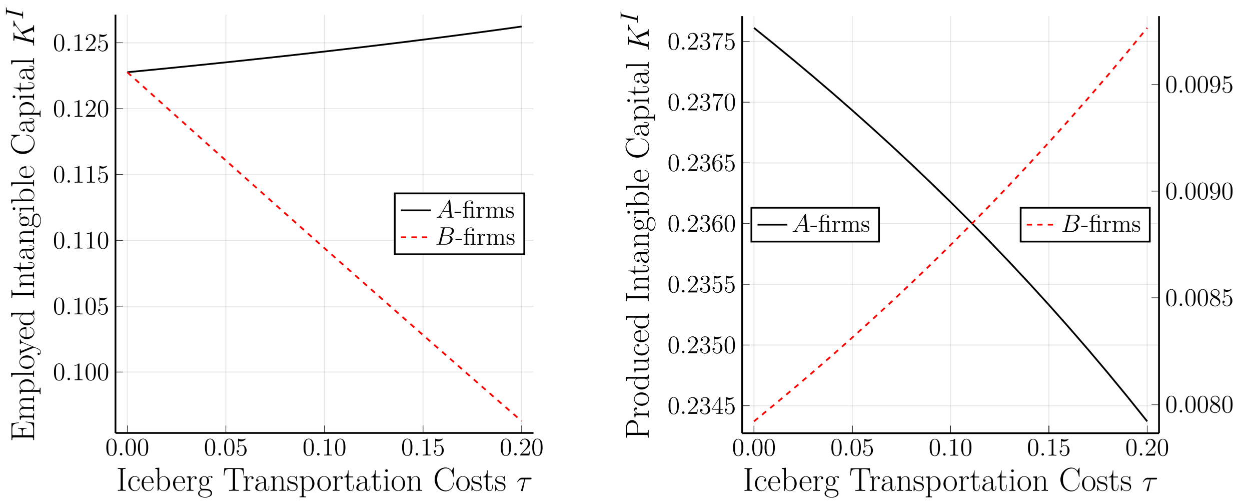



Figure 3: **Price of Intangible Capital:** The introduction of ice berg trading costs $\tau$ introduce a wedge between the price that sellers and buyers of intangible capital face. While $p^I$ falls, the price that $B$-firms face increases.

Although an increase in trading frictions makes the economy less efficient as a whole, it may be that one type of firm profits at the expense of the other. In particular, firms experience two opposing effects. The *direct* effect is negative for both firms. While the price for intangible capital $p^I$ falls, which harms $A$-firms as the sellers of intangible capital, the price that $B$-firms face ($\frac{p^I}{1-\tau}$) increases, which harms $B$-firms as the buyers of intangible capital. If both firms compete for the same market ($v < \frac{\sigma-1}{\sigma}$), then there is a positive *general equilibrium* or *indirect* effect. As decreased aggregate output $Y$ leads to an increase in the price of intermediate goods $p_j$, both type of firms




(a) **Employed Capital** (b) **Produced Capital**

Figure 4: **Iceberg Transportation Costs and Intangible Capital:** Higher iceberg transportation costs lead to a higher use of intangible capital in sophisticated versus naive firms. At the same time, sophisticated firms produce less intangible capital, whereas naive firms produce more.

would like to expand their production. In this respect, $A$-firms are in a better position to expand their output, as they face lower costs. This leads to our first hypothesis:

**Hypothesis 1** *If the* general equilibrium *effect is sufficiently strong and and the difference between* sophisticated *and* naive *firms is large enough, then tighter data regulations (higher iceberg transportation costs $\tau$) lead to an increase in profits of* sophisticated *firms.*

The corresponding simulation to the hypothesis is pictured in Figure 5, in which iceberg transportation costs are gradually increased.

To summarize, the model predicts that a tightening of data regulations may benefit *sophisticated* firms at the expense of *naive* firms. The main reason is that *sophisticated* firms are better at gathering data in-house, whereas *naive* firms face higher costs when trying to substitute externally acquired data with in-house data.

Finally, these results suggest that a tightening in data regulations can lead to an increase in market concentration, as already large firms become even larger. We seek to explore the implications for entry and exit in future work.

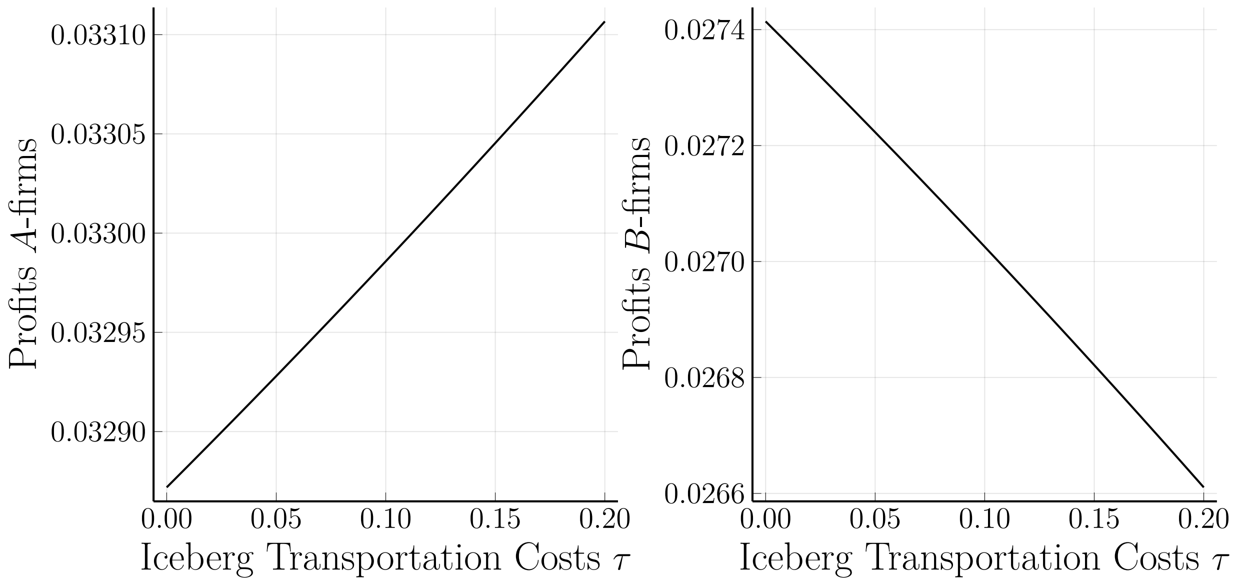



Figure 5: **Firm Profits:** Sophisticated firms increase their profits at the expense of naive firms.

# 7 Conclusion

Quantifying the value of consumer data and consumer privacy is crucial for understanding the rapid period of transformation of the modern data economy. In this project, we exploit the introduction of the California Consumer Protection Act (CCPA) which limits the amount of data firms can trade, but does not affect the amount of data firms can generate in-house to the same extent. We examine a unique and hand-collected data set of conversational-AI firms that rely on consumer data to grow their business and we found that personal consumer data is a source of competitive advantage for firms and the key to unlocking customer value.

Our empirical analysis suggests that the staggered adoption of the CCPA gave a strong advantage to firms with in-house data as opposed to firms that rely on external data acquisition. Voice-AI products of firms with lots of in-house data prior to CCPA experience significant appreciations in customer ratings and additional feedback, i.e., data, from customers after the CCPA. These results are robust to firm-product and daily fixed effects, and subsample tests on public and private businesses. We also examined the financial ramifications of voice-AI products before and after the adoption of the CCPA. We provided additional evidence consistent with the argument that CCPA benefits firms with in-house data. Such firms exhibit higher valuations, profitability, asset utilization, and they invest more after the adoption of CCPA. Moreover, their earnings become easier to predict as uncertainty about the effects of CCPA is resolved once the CCPA is signed into law.

Lastly, to rationalize our empirical findings, we built a simple theoretical model in which firms with different levels of sophistication use data in the form of intangible capital as an input into their production function. We showed that a tightening of privacy regulation in the form of an increase in the iceberg transportation costs of data may lead to an increase in the profits of sophisticated firms at the expense of naive firms, as the latter find it more costly to substitute externally acquired data. This theoretical prediction matched our empirical results, as we expected sophisticated firms with more in-house data prior to the increase in the transportation costs of data to benefit from this restriction. In future work, we plan on expanding our theoretical analysis to more closely match the partial non-rival nature of intangible capital such as data, and consider other aspects, such as dynamic effects or firm entry and exit.

# References


**Acemoglu, Daron**, "Harms of AI," Technical Report, National Bureau of Economic Research 2021.

**Acquisti, Alessandro, Curtis Taylor, and Liad Wagman**, "The Economics of Privacy," *Journal of Economic Literature*, 6 2016, *54* (2), 442–492.

**Aridor, Guy, Yeon-Koo Che, and Tobias Salz**, "The Effect of Privacy Regulation on the Data Industry: Empirical Evidence from GDPR," NBER Working Papers 26900, National Bureau of Economic Research, Inc 2020.

**Babina, Tania, Anastassia Fedyk, Alex Xi He, and JamesW Hodson**, "Artificial Intelligence, Firm Growth, and Industry Concentration," Technical Report 2021.

**Begenau, Juliane, Maryam Farboodi, and Laura Veldkamp**, "Big data in finance and the growth of large firms," *Journal of Monetary Economics*, 2018, *97*, 71–87.

**Bergemann, Dirk, Alessandro Bonatti, and Tan Gan**, "The economics of social data," 2020.

**Bessen, James, Stephen Impink, Lydia Reichensperger, and Robert Seamans**, "GDPR and the Importance of Data to AI Startups," SSRN Working Paper 3576714 2020.

**Campbell, James, Avi Goldfarb, and Catherine Tucker**, "Privacy Regulation and Market Structure," *Journal of Economics & Management Strategy*, March 2015, *24* (1), 47–73.

**Cong, Lin William, Danxia Xie, and Longtian Zhang**, "Knowledge Accumulation, Privacy, and Growth in a Data Economy," *Management Science*, 2021.

**de Matos, Miguel Godinho and Idris Adjerid**, "Consumer Consent and Firm Targeting After GDPR: The Case of a Large Telecom Provider," *Management Science*, 2021.

**Eeckhout, Jan and Laura Veldkamp**, "Data and Market Power," Columbia University Working Paper Sep 2021.

**Farboodi, Maryam and Laura Veldkamp**, "Long-Run Growth of Financial Data Technology," *American Economic Review*, August 2020, *110* (8), 2485–2523.

**_ , Roxana Mihet, Thomas Philippon, and Laura Veldkamp**, "Big Data and Firm Dynamics," *AEA Papers and Proceedings*, May 2019, *109*, 38–42.

**Hagiu, Andrei and Julian Wright**, "Data-enabled learning, network effects and competitive advantage," NUS Working Paper 2020.

**Hoberg, Gerard and Gordon M. Phillips**, "Scope, Scale and Competition: The 21st Century Firm," SSRN Working Paper 3746660 2021.

**Jia, Jian, Ginger Zhe Jin, and Liad Wagman**, "The Short-Run Effects of GDPR on Technology Venture Investment," *Marketing Science*, Mar 2021, *40* (4), 661–684.

**Jones, Charles I and Christopher Tonetti**, "Nonrivalry and the Economics of Data," *American Economic Review*, 2020, *110* (9), 2819–58.

**Khansa, Lara, Deborah Cook, Tabitha James, and Olga Bruyaka**, "Impact of HIPAA provisions on the stock market value of healthcare institutions, and information security and other information technology firms," *Computers & Security*, Sep 2012, *31*, 750–770.

**Peukert, Christian, Stefan Bechtold, Michail Batikas, and Tobias Kretschmer**, "European Privacy Law and Global Markets for Data," Technical Report 2021.

**Rydning, David Reinsel-John Gantz-John**, "The digitization of the world from edge to core," *Framingham: International Data Corporation*, 2018, p. 16.

**Tang, Huan**, "The value of privacy: Evidence from online borrowers," Technical Report, Technical report, Working Paper, HEC Paris 2019.

**Vesset, Dan and Jebin George**, "Worldwide Big Data and Analytics Spending Guide," 2021.

**Vives, Xavier and Zhiqiang Ye**, "Information Technology and Bank Competition," Technical Report 2021.