INSTITUTE FOR INNOVATION
AND PUBLIC PURPOSE

Developing Dynamic Capabilities through Acquisitions

A patent lens on M&A's impact on Big Tech's technological profile

Ilan Strauss
Senior Research Associate
| UCL Institute for Innovation and Public Purpose

Jangho Yang
Assistant Professor
| Management Sciences in the Faculty of Engineering, University of Waterloo

Institute for Innovation and Public Purpose

WORKING PAPER
2024/01

About the Institute for Innovation and Public Purpose

The Institute for Innovation and Public Purpose (IIPP) at University College London (UCL) aims to develop a new framework for creating, nurturing and evaluating public value in order to achieve economic growth that is more innovation-led, inclusive and sustainable. This requires rethinking the underlying economics that has informed the education of global civil servants and the design of government policies. Our work feeds into innovation and industrial policy, financial reform, institutional change and sustainable development. A key pillar of IIPP's research is its understanding of markets as outcomes of the interactions between different actors. In this context, public policy should not be seen as simply fixing market failures, but also as actively shaping and co-creating markets. Re-focusing and designing public organisations around mission-led, public purpose aims will help tackle the grand challenges facing the 21st century. IIPP is a department within UCL – and part of The Bartlett, which consistently ranks in the top two faculties for architecture and the built environment in the world.

www.ucl.ac.uk/iipp

ISSN 2635-0122

This working paper can be referenced as follows:

Strauss, I. and Yang, J. (2024). Developing Dynamic Capabilities Through Acquisitions: A patent lens on M&A's impact on Big Tech's technological profile. UCL Institute for Innovation and Public Purpose, Working Paper Series (IIPP WP 2024-01). Available at: https://www.ucl.ac.uk/bartlett/public-purpose/wp2024-01

Developing Dynamic Capabilities Through Acquisitions: A patent lens on M&A's impact on Big Tech's technological profile

Ilan Strauss and Jangho Yang *

January 2024

Abstract

Using a unique dataset covering all patents ultimately owned by Big Tech, including through M&A, we describe the dynamic capabilities acquired and developed by Big Tech. We examine the nature, evolution, and differences in the technological capabilities of the five major Big Tech companies, highlighting the vital role of M&A in this process. Our analysis, combining M&A and patent data, reveals that M&A has been crucial in Big Tech developing integrated hardware-software ecosystems. Big Tech's evolving capabilities closely track their competitive potential and market entry strategies. Our analysis of patent data, including through a logit regression, reveals diverse strategies and motivations behind Big Tech's external patent acquisitions.

Keywords: Patents, capabilities approach, Big Tech, M&A, dynamic competition.
JEL Codes: K21, L41, L44, M13, E14.

---

*Ilan Strauss is a Senior Research Associate at the Institute for Innovation and Public Purpose (IIPP), University College London (UCL), London, UK, WC1E 6BT. Jangho Yang is Assistant Professor of Management Sciences in the Faculty of Engineering, University of Waterloo. We would like to acknowledge grateful funding from The Omidyar Network and in particular The Economic Security Project (ESP) for this project grant. A big thanks to Becky Chao at ESP for grant assistance. We would like to express our deep gratitude to Nigel Swycher for access to Cipher's patent dataset (acquired by LexisNexis) and to Patrick Curry and others at Cipher for working through tedious dataset construction together. A big thank you to our two incredible research assistants at University of Waterloo: Elisha Matharu for data and programming, and Fuyang Deng for graphing and case studies. Thank you to Tim O'Reilly for edits and comments. Thank you to those who provided helpful comments at two workshops: at the UCL Centre for Law, Economics and Society on 10th October 2023, and at the UK CMA on 12th October, 2023, where a more antitrust focused version of this paper was presented. All errors are our own. Contacts: i.strauss@ucl.ac.uk (Corresponding Author) and j634yang@uwaterloo.

1 Introduction

There is a growing sense among regulators and scholars that existing antitrust theory has been excessively focused on regulating competition as a static struggle for existing market share within a stable industry (Evans et al. 2008). The impact of this has been to permit Big Tech to acquire hundreds of smaller, highly innovative companies, each with minimal existing market share. Our merger and acquisitions (M&A) dataset shows at least 995 majority acquisitions by the five Big Tech companies of Apple, Amazon, Alphabet, Meta, and Microsoft between 2000 and 2022.[1] These are young firms being acquired with novel, often, unproven technologies.[2] But they have been the essential building blocks in Big Tech creating diversified product ecosystems (Doctorow 2023; Wikipedia 2023). Concerns over this acquisitive strategy are magnified by the technological shift underway towards artificial intelligence and generative language models, which are covered by machine learning, image processing, and speech recognition in our patent dataset. Such a technological shift can dethrone incumbents and foster competition, yet Big Tech may remain dominant due to high barriers to entry, minimal disclosures (Mazzucato et al. 2023; O'Reilly et al. 2023), and their continued acquisitions (Sharma 2023).

FTC Commissioner Rebecca Kelly (De Vynck et al. 2021) argues that: "Each individual merger, viewed independently, may not be seen to have significant impact [...] But the collective impact of hundreds of smaller acquisitions can lead to a monopoly behemoth." This rests on the idea that competitive advantage arises from a company's ability to continuously innovate over time (D. J. Teece et al. 1997), and that this in turn is a cumulative capability (Posner 2006),[3] involving combinations of thousands of complementary assets working together (Sidak et al. 2009). In this approach, the competitive threat facing the monopolist is from new potential product and technological markets in the future (Bryan et al. 2020a).

Despite calls for antitrust authorities to make a more concerted effort to apply a dynamic approach to M&A oversight, premised on capabilities (Sidak et al. 2009; McSweeny et al. 2018; Bryan et al. 2020b; Petit and D. J. Teece 2021; Cadman 2023), no clear theoretical approach has emerged – despite harms to innovation greatly reducing economic welfare in the long-run (Hovenkamp 2008). One reason for this absence is empirical, since (ibid., p. 3): "measuring innovation or its impact from an ex ante

[1] Most transactions are below the size thresholds required in the Hart-Scott-Rodino (HSR) Act.

[2] Around half of the acquisitions not reported to the FTC between 2010-2019 were firms younger than five years old (FTC 2021). Big Tech reported 819 non-HSR reportable transactions over the 10-year period 2010-2019 alone (ibid.). But this includes 87 minority stakes, non-corporate interests, and non-acquisition licensing agreement which we try to exclude from our dataset.

[3] Posner (2006):"most intellectual property builds upon intellectual property. Most of it is incremental"

perspective" is almost impossible, whereas "Most changes in price and output are continuous and related to one another".

Our analysis, although preliminary and descriptive, offers one attempt to empirically illustrate how a dynamic capabilities analysis can be undertaken by scholars and practitioners on Big Tech using patent data – including limitations to this approach. Our empirical approach is largely descriptive, *focusing on how M&A has contributed to Big Tech's evolving technological capabilities*, even though a more normative and theoretical approach is possible. In our analysis, patents are not just a bundle of intellectual property rights (Morton et al. 2013), but a collection of capabilities working in combination to produce products, services, and competitive ecosystems (D. J. Teece 2012). Patents are a proxy for Big Tech's ability to innovate – for their capabilities (Pavitt 1982; Levin et al. 1987; Griliches 1998) – even though they are inherently uncertain, and so may reflect failed innovations as much as real innovative potential. We use fine-grained machine learning classified technological categories from Cipher (now LexisNexis Patents), instead of product categories, as a more intuitive way to categorise Big Tech's capabilities. Our core findings are that:

1. Quantitatively, M&A has been central to Big Tech developing its capabilities. At least 10.3%, 13.1%, and 10.8% of total patent counts on an unweighted, forward-citation-weighted, and forward-citation-weighted adjusted for patent-age basis, respectively, have been acquired externally by Big Tech through majority stake M&A activity (Figure 1). These are lower bound estimates because Big Tech also innovates internally based on external technologies. We are also unable to dissect the assignment history of each patent to know the ultimate source of every single patent (external vs. M&A). Qualitatively, acquisitions have given Big Tech essential, but often young, technology, which is not always captured in the patent data, since patenting takes time and reflects innovation *ex post* (Kim et al. 2016)[4] – and because software may also be less patentable after the 2014 Alice Corp. v. CLS Bank International ruling (Saltiel 2019).

2. Patent (forward) citations show that external acquisitions provide the Big Tech firms with access to technology that is often more productive than their own internally developed technology, judged by various measures of forward citations of patents acquired through M&A compared to internally developed patents (Figure 1 and Table 3).

[4] Kim et al. (2016) find that:"the later the timing of the patent, the higher the innovation performance, while under low uncertainty there is an early-mover advantage".

3. Big Tech show a diverse range of motivations in the capabilities that they have acquired. For most acquisitions, capabilities acquired tend to be extremely young (judged by low median patent counts) with few if any proven technologies (judged by few patents with citations). This makes quantitative analysis of M&A's impact on Big Tech's capabilities difficult, especially on a case by case basis. Many of the patents acquired by Big Tech have also expired, either due to age or product failure. However, an overall pattern emerges where Big Tech uses M&A to expand their ecosystems, notably through the acquisition of a range of hardware, location & satellite technologies. Individual, highly advanced, technological acquisitions are also evident (judged by very high median citations in key firms acquired) in technologies that have been vital to Big Tech, such as in touch screens, voice recognition, advertising technologies, and other software capabilities.

4. Descriptively, there appears to be a tight link between diversification in capabilities (including through M&A) and diversification and entry into product markets. Big Tech have pivoted notably to integrated hardware-software ecosystem by way of their capabilities after 2010, which involves a much greater emphasis on the user interface, social media, wireless networks, and to a lesser extent storage. Machine learning shows the largest jump in patents held after 2010, relative to 2010 and prior. Alphabet shows the greatest use of M&A to build out a range of capabilities across technological categories, showing the most integrated and diverse suite of ecosystem products.

In contrast to the emphasis on assets being acquired by Big Tech to be excluded, "killed", or deter entry, as in the dominant "killer acquisition" story (Cunningham et al. 2021; Affeldt et al. 2021), our analysis highlights that acquired capabilities are essential – and highly uncertain – things which allow for market entry, adaptation to constantly changing market conditions (D. J. Teece et al. 1997), and for the erection of deeper ecosystems moats (Petit and D. Teece 2020). Our findings indicate circumstantially that when Big Tech buy companies, they are buying capabilities with which to compete through furthering their own innovation directly (Gugler et al. 2023), rather than indirectly through exclusionary tactics. This does not preclude anti-competitive effects, though, since a real or potential competitor is being removed from the market (Areeda et al. 2023, Section 701).

By looking at the combined effects of past acquisitions, we provide evidence that may frame an individual future acquisition as potentially more harmful, presumptively, if it can be shown to contribute to an established pattern of behaviour, to dominate or acquire key competitors within a given technological ecosystem. This gives expression to the U.S.'s Merger Guidelines.[5]

A capabilities approach, focusing on the dynamics of competition, has arguably become more salient as the gap from technology to market entry (Argente et al. 2020) has greatly shortened, given Big Tech's pre-existing access to mass distribution online and their ability to integrate multiple services on a single platform.[6]

Our paper provides the first publicly available dataset on all patents owned by Big Tech, adjusting for changes in ownership and assignee status. Typically, patent transactions involve the acquiring firm either retaining the original assignee name or reassigning it to themselves. Changes in a patent's ultimate ownership are not tracked publicly.[7] Our dataset instead makes adjustments for changing patent ownership based on careful work done in generous collaboration with Cipher – a machine-learning powered patent data company recently acquired by LexisNexis. (For more information see Data Appendix.) Our data approach, linking Big Tech's M&A data to patent data, is similar to Gugler et al. (2023), but they focus on the time-series dimension. Our dataset has significant advantages since it is organized at the (consolidated) patent *family* level, which is global and eliminates redundant patents filed, since the same invention might be patented in multiple countries. Our data also tracks changes in patent ownership more comprehensively, including through changes in ownership, and we use a more extensive M&A dataset. Our present dataset does not, however, include time-series of patent counts and citations since these are covered partially in Gugler et al. (ibid.).

Section 2 provides an overview of our data, exploring potential motivations for M&A through a patenting lens. Section 3 examines the relationship between M&A and competition, focusing on the technological patents which it has given Big Tech access to. Section 4 conducts a brief logit regression, predicting the probability of a patent being externally acquired through M&A to assess if capability acquisition strategies differ by Big Tech firm. Section 5 concludes briefly with potential policy implications.

[5] Which propose that (FTC and DoJ 2023, p. 23):"the Agencies will consider individual acquisitions in light of the cumulative effect of related patterns or business strategies". For critical discussion see Hovenkamp (2023a).

[6] Notes Areeda et al. (2023, p. 701b), this acquisitive behaviour is more harmful when the firm being acquired shows more potential for growth or more capacity for expansion, as online markets with low marginal costs to distributions are characterized by.

[7] The USPTO has a database that tries to do this at a very basic level.

2 Data and Overview

2.1 Capabilities, Patents, and Dynamic Competition

Our paper uses patents as a proxy for a firm's capabilities. Patents are often used in academic research as indicators of a firm's technological capabilities, innovation output, and overall competitive strength (Akcigit et al. 2023). This is because patents represent legally protected innovations that can provide a firm with a competitive edge in final product markets (Argente et al. 2020). The number and quality of patents held by a firm are typically seen as a reflection of its research and development (R&D) intensity, innovative capacity, and ability to create new products or processes (Pavitt 1982; Levin et al. 1987; Griliches 1998). As a result, in markets subject to dynamic competition, capabilities from patents may drive competitiveness in goods and services. Forward citations in particular can reflect the quality of the patent and in turn its degree of innovativeness or patent centrality (Trajtenberg 1990).

Our analysis, integrating patents as capabilities with M&A activity, emphasizes all three aspects of a dynamic capabilities approach together (D. J. Teece 2007): the recognition of new opportunities or threats; the capacity to respond to these; and the ability to continuously renew one's capabilities. From a dynamic capabilities perspective (D. J. Teece 2018), what has made Big Tech's competitive strategy so potent is precisely its ability to always be renewing and re-configuring its assets in order to innovate and continuously respond to changes in present or potential market demand (D. J. Teece 2018; Pundziene et al. 2019).

From a competition perspective, a dynamic assessment (Cadman 2023) involves focusing on the threat from innovation, from outside the market (Caffarra et al. 2023a,b). This has recently gained prominence among several established competition scholars and practitioners (Sidak et al. 2009; Petit and D. J. Teece 2021) - even if differences exist in Europe and the U.S. (Gifford et al. 2011).[8]

Applying a dynamic competition approach to Big Tech's M&A activity has seen emphasis on

[8] Dynamic competition is defined by Sidak et al. (2009) as: "a style of competition that relies on innovation to produce new products and processes and concomitant price reductions of substantial magnitude. Such competition improves productivity, the availability of new goods and services and, more generally, consumer welfare." For Petit and D. J. Teece (2021) it is:"a situation in which firms compete for future rents. In dynamic competition, firms use innovation to introduce new products, processes and services. Rivalry results in product differentiation, integration, diversification, or platformisation. It is a type of competition animated not by firms that compete head-on with similar products but by heterogeneous competitors, complementors, suppliers and customers [...]. Such competition improves long-term factor productivity, raises consumer welfare and supports higher wages." Both cited in Cadman (2023).

several similar concepts: "nascent competition" (FTC 2018),[9] "ecosystem" competition (Jacobides, Cennamo, et al. 2018) or "ecosystem theories of harm" (Caffarra et al. 2023a,b), and in U.S. antitrust law the "potential competition" doctrines (Glick et al. 2019; FTC 2018; Areeda et al. 2023), as was argued successfully in the FTC's revised complaint against Facebook involving their acquisition of Instagram.[10] Such acquisitions are increasingly being modelled too (Bryan et al. 2020a; Chen et al. 2023; Benkert et al. 2023). Dynamic competition has a long history in Austrian economics, perhaps most prominently with Schumpeter (2003, p. 87), who emphasises that the impact of innovation over time is the ultimate uncertainty which monopolists try to mitigate through their business decisions. Schumpeter argued that the real threat to competition came from the struggle for market share for tomorrow's good and services premised on new technologies and new good and services. Not from monopolistic practises aimed at "conserving established positions and at maximizing the profits accruing from them."

2.2 Data

Our key dataset is at the patent level – showing all patents owned by Big Tech – and combines two datasets: a dataset of firms acquired by Big Tech (Refinitiv, Wikipedia, Web Scraping) and a dataset on all patents owned by Big Tech (Cipher now owned by LexisNexis). We collaborated with Cipher to improve the linking of organizations to patents. This is crucial for tracking changes in patent ownership ("assignee"), which can change when companies are bought and sold, but might instead remain the same. This makes linking the patent to the new owner difficult. Our patent data is structured hierarchically, with an ultimate organization linked to a patent owner, the legal assignee of the patent, an original historical owner, and assignee history.

We merge & match patent data with M&A data (by Big Tech company). Matching is done based one of the following variables: the patent's ownership, its assignee, or its historical ownership. 10% of patents (13,196 patents) match to external assignees / organisztions / historical owners; or 227 firms from our M&A dataset are present in our patent dataset (22% match). This underestimates the true

[9] Defined as "verticals" by Susan Athey in remarks to FTC. See also remarks by Paul Denis (FTC 2018, p. 187): "So my suggestion for this afternoon is that we try to focus on "nascent competition" being limited to competition that's presently being felt but not yet fully realized. Used in this sense, the acquisition of a nascent competitor by one of its rivals would be seen as extinguishing not only current competition between those firms but also either extinguishing or perhaps amplifying the prospect for greater competition in the future."

[10] See *FTC v. Facebook, Inc.* (2022). Notes Areeda et al. (2023, p. 701d):"But when one of the merging firms is a monopolist and the other is a potential entrant into the same market in which the monopolist has its power, anticompetitive concerns [under the "potential competition" doctrines] are much less speculative [...] it is important to preserve that unique prospect for competition in the future.". See also Hovenkamp (2023b).

extent though of patents acquired externally though since patents developed internally might rely on externally acquired patents. Google might buy an external company along with all its patents and integrate that company as a separate subsidiary. But all new patents invented from that subsidiary can be assigned to Google itself instead of the subsidiary, for example.

M&A Data, 2000 - 2022. The M&A dataset is a firm-level merger and acquisitions dataset of 995 majority-stake acquisitions undertaken by the five Big Tech companies from Refinitiv, Wikipedia, and AI webcrawlers, all of which have been verified by humans (with corresponding verification source provided). See Table 1 for the number of acquired firms by each Big Techs. The dataset ranges from 2000 to 2022 for the year completed of the M&A deal. Share repurchases, minority stakes, and property acquisitions are excluded. Acquisitions of subsidiaries are included, for example Google buying Motorola Mobility. Software is the main 'mid' level category of acquisition[11], with "Internet Software & Services" being a close second for Amazon. The 'macro' industry classification of the acquired firms given by Refinitiv are almost entirely in High Technology.

Table 1. Firms Acquired by Big Tech, 2000 - 2022

Acquirer	No. of M&A	Share of Sample
Alphabet	341	34.3
Microsoft	272	27.3
Amazon	144	14.5
Apple	122	12.3
Meta	116	11.7

Note: *Based on 995 firms acquired by Big Tech in Refinitiv database, Wikipedia, and AI Crawlers between 2000-2022. Cleaned to include only majority stake, removing share repurchases, real estate deals, and deals frozen or rejected by regulatory authorities (e.g. Meta/Giphy). We include Microsoft's 2009 deal with Yahoo! (amended 2015) since at the time it provided Microsoft with an exclusive 10-year license to Yahoo!'s core search technologies.*

Patent Dataset, 2000 - 2022. The second dataset is a patent-level dataset from Cipher contained all patent families owned by the five Big Tech companies, containing 127,298 patent (families). See Table 2 for the number of patents externally obtained by each Big Techs. Cipher is a private patent

[11] Only available for our Refinitiv sub-sample or three quarters of the M&A dataset.

data company (recently acquired by LexisNexis, to be integrated into their PatentSight service), which uses publicly available patent data and applies machine learning classifiers to discern useful technology categories. The patent data is at the patent family level - this is global and avoids multiple similar patents filed in different jurisdictions removing considerable redundancies. We can filter by geography though based on where the patent was granted. It includes expired, inactive, and pending patents, but excludes design patents. Since we want to capture previously acquired patents from past acquisitions - even if expired. All citation data is as of the present (December 2022). Other patent-level variables in our dataset are a patent's forward citations, patent title, patent abstract, year granted, PVIX score, technology categories from Cipher / LexisNexis ML classifiers, and standard Cooperative patent classification (CPC) codes.

Table 2. Merged Patent-level Dataset, 2000 - 2022

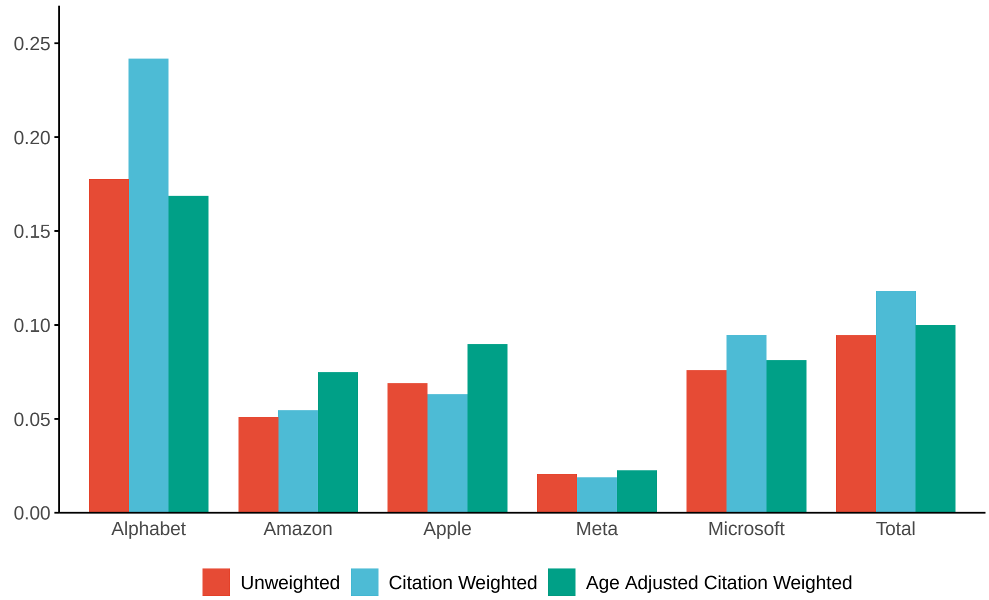
Big Tech	Total Patents	M&A Patents	M&A Patent (%)	M&A w. Patents
Alphabet	32,855	5,835	17.7	60
Amazon	15,175	775	5.1	35
Apple	27,489	1,896	6.5	32
Meta	7,582	161	2.1	21
Microsoft	44,197	3,417	7.7	77
Total	127,226	12,084	9.5	225

Note: *Patents acquired by Big Tech. Removing a jointly owned patent between Meta and Microsoft. Final column showing number of acquired firms in sample which contains patents in the merged datasets.*

225 firms from our M&A dataset are present in our patent dataset, or a 22.6% match percentage indicating a high share of recognized technology in the acquired companies. Apple and Microsoft have the highest capabilities motivation in their acquisitions, with 26.2% and 28.3% respectively of their acquired firms having matched patents. Alphabet has the lowest at 17.6% - even though it has the highest share of patents arising from M&A.

Figure 1 shows that at least 10.3%, 13.1%, and 10.8% of patents unweighted, citation weighted, and citation weighted adjusted for age, respectively, are acquired externally by Big Tech.

Figure 1. Share of Patents from M&A: Unweighted vs Citation Weighted



Note: *10.3%, 13.1%, and 10.8% of patents unweighted, citation weighted, and citation weighted adjusted for age, respectively, are acquired externally by Big Tech. "Weighted" means that we multiply patents by forward citations. age-adjusted involves dividing this measure by the age of the patent, defined by the year it was granted.*

This is certainly a lower bound given that we are unable to dissect the assignment history of each patent, and so we cannot know the source of each patent (external vs. M&A) with any certainty. And more generally, internally developed patents might crucially rely on externally acquired ones (including through complementary human capital). In addition, our matching algorithm is imperfect. Finally, qualitatively, many acquired technologies are not always patented, especially when in their infancy or involving software. For all Big Tech firms except for Alphabet, acquired patents are more innovative (ex-post), since we using citations as of the present (2022) rather than at the time of acquisition. Even after adjusting for patent age, than external patents, since forward citations are higher for these. This is shown by the unweighted column being smaller than the age-adjusted citations column for all Big Tech firms, except Alphabet. This is especially true for Amazon and Apple. Alphabet's acquired patents are older due to all the Motorola Mobility patents they hold, making age adjustment important for them. After accounting for this, their externally acquired patents are not more innovative than their internally developed ones.

Although this does not adjust for self-citation (by the same firm), or for the fact that Big Tech might make these patents more productive through integration with their technology, this shows that when Big Tech buy companies they are in fact buying capabilities with which to further their own innovation. This is somewhat contrary to the "killer acquisition" story. Two key variables of interest in our patent dataset are:

- **Forward citations**: Indicating how many times the patent has been cited by other patents (Hall et al. 2005).

- **ML Technology category**: The ML technology classifier is an unsupervised learning based multi-class classifier. The ML technology classifier is a Universal Technology Taxonomy ("UTT"), meaning it classifies the entire patented world through a common lens, based on 10 major technological categories and 122 subcategories (LexisNexis 2023). This contrasts with the 300,000 CPC codes which is for classifying prior art when examining a patent application. CPC codes are less useful for classifying existing patenting technology in relation to one another since they are given without a view to how the patent is actually used or the technology evolves in relation to other technologies.

2.3 Motivations for Acquisitions

Based on the broad industry classification in Refinitiv of the firms acquired (high-technology), acquiring advanced capabilities (including patents and human capital, but also data) is clearly a key motivation. Capabilities are often in their infancy when acquired though, complicating any quantitative analysis. Table 3 shows that median patents held among all the companies acquired are as low as 2 (Microsoft, Alphabet, and Amazon), and with a high of 4 only for Apple (Table 3) - column four. In other words, the companies which Big Tech are acquiring tend to be even too young to hold any recognized, registered, technological assets.

Table 3. How Innovative are the Acquired Companies (by Citations & Patents)?

Big Tech	Median Cite (Non-M&A)	Median Cite (M&A)	Median Patent Total	Max Patent Total	Max Firm
Microsoft	15	20	2	1882	Nuance Communications
Alphabet	9	21	2	4992	Motorola Mobility
Apple	9	12	4	1582	Intel (Modem chip)
Amazon	5	5	2	547	Zoox
Meta	5	4	2	79	WhatsApp

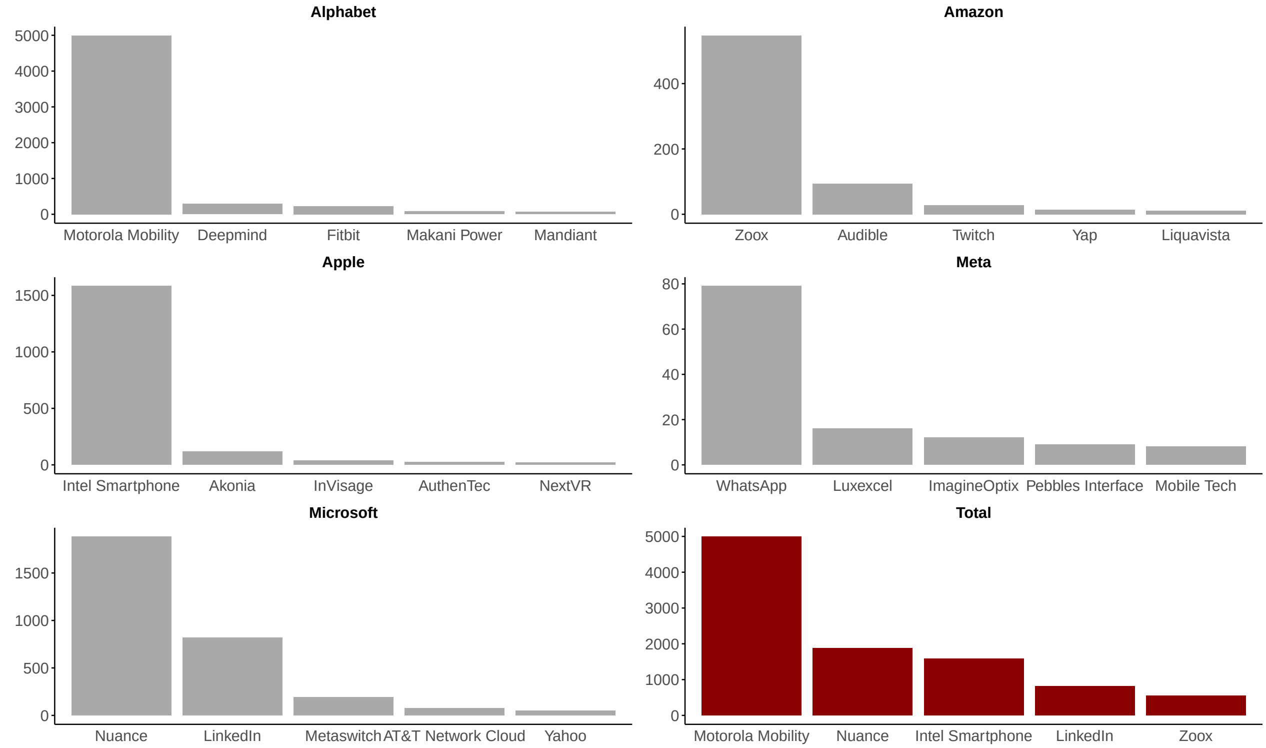
Note: *Showing number of M&A firms acquired with patents; median forward citations among pooled acquired patents; median no. of patents acquired considering the total patents acquired in each firm; maximum patents acquired from any one firm; and the firm containing the most amount of patents acquired.*

A key finding shown in Table 3 is just how much some Big Tech firms rely on acquisitions for innovation. Median citations among all acquired patents, compared to all non-acquired patents, is twice as high in Alphabet (21 vs. 9), one-third higher for Microsoft (20 vs. 15), and one-third higher for Apple (12 vs. 9). Though for Amazon and Meta citations for internally developed and externally acquired patents are equal or a bit less than equal.

A key motivation for Big Tech's acquisitions has been building out new, more expansive, product ecosystems (Jacobides, Cennamo, et al. 2018; Jacobides and Lianos 2021). This reflects vertical (different or unrelated production stages or products) rather than horizontal (same market) acquisitions into complementary hardware, software, cloud, and emerging (sometimes speculative) technologies. This is evident in Figure 2, which shows the largest five acquisitions, by total number of patents, by each Big Tech.

It highlights such vertical and technological deepening acquisitions. For Alphabet, this is about shifting from software to hardware and selling goods. For Amazon, acquisitions have supported pure diversification in capabilities in ways which it thinks supports its core retail business, such as Zoox (self-driving cars) and Audible (audiobooks and podcasts), or its web service, such as Twitch (gaming

Figure 2. Top Five Acquired Companies with Most Patents by each Big Tech



Note: *Top five companies, by total patents, for each Big Tech company. Including Yahoo! Search and Advertising technologies acquired by Microsoft since it was an exclusive technology license for 10 years (the terms of which were changed in 2015).*

streaming).[12] And this is highlighted at the patent level, as a proxy for capabilities. Integration of AI by Big Tech is evident with large acquisitions for Alphabet (DeepMind), Microsoft (Nuance Communications), Meta (Mobile Technologies) (TechCrunch 2013) and Amazon (Yap) (TechCrunch 2011). Apple's purchases were often used to bolster its iPhone (InVisage - improved cameras) while for Amazon, the AI was historically for voice recognition (Amazon's Echo).

Ecosystem diversification is not always evident in the patent data at the individual firm-level though (Doctorow 2023). Patterns often only emerge at the aggregate technological level, taking into account all firms acquired. It is well known that Facebook bought Whatsapp and Instagram, that Google bought DoubleClick and YouTube, and that Apple bought Beats. In fact, most of the beloved products of Big Tech grew out of serial acquisitions. Google's acquisition of Docverse in 2010 helped it build out its online collaborative suite of Office-like products (Google Docs, Google Slides, etc.). Google bought a one and a half year old start up called Android in 2005, to expand its core search and ads business beyond the PC platform (Callaham 2022). Today, Android is the most popular

[12] Amazon's two largest acquisitions are not represented in our data: Wholefoods and Metro-Goldwyn-Mayer (MGM).

mobile operating system (OS) and provides Google, in conjunction with its suite of apps (its "Google Mobile Services"), with considerable leverage over OEMs to make its apps default. Yet no patents are registered under Android at Google. The company Android was too young to have patents at the time of its acquisition. In most of these instances, the acquisitions replaced flailing internal products at Big Tech, thereby reducing competition in the market, as in Maps, or Youtube displacing Google Video, or Beats Music displacing Apple iTunes purchase-to-own model.

The major acquisitions in Figure 2 also highlight a pure patent motivation (for defensive reasons and to enter protected markets, such as with Alphabet buying Motorola Mobility and Apple buying Intel Smartphones). Alphabet acquired patents and phones from Motorola Mobility (for $12.45bn). This is by far Alphabet's largest acquisition to date - with Nest (2014) coming in second at $3.2 billion. Google executives at the time acknowledged patents played a role in the original Motorola deal (Kopytoff 2014). When Google sold Motorola Mobility to Lenovo for $2.91bn in 2014, it kept the majority of patents, for example (which our data tries to track and account for) (Google 2014). Google tried to use the acquisition to better enter into the smartphone market. But this failed and it sold off these assets to Lenovo.

Similarly, Apple's major patent acquisition (for $1 billion) comes from Intel's smartphone modem division, which was motivated by a desire to acquire patents and expertise to pursuit chip Independence from Qualcomm (which is defended by its patent empire which Apple must license at substantial cost) (Reuters 2022). This remains Apple's second largest acquisition to date (after Beats for $3 billion in 2014).[13] As part of the deal, Apple took over 17,000 wireless technology patents, including protocols for cellular standards, modem architecture, and modem operations (Spiceworks 2022). Apple argues that it is impeded in developing its 5G smartphone modems by two Qualcomm patents (PhoneArena 2022), forcing it to continue to be Qualcomm's largest customer until at least 2026 (Spiceworks 2022).

Microsoft's acquisition of Nuance Communications ($19.7 billion) - third to LinkedIn and now Activision Blizzard - was driven by acquiring cross-cutting artificial intelligence and voice recognition technologies, with applications not just in Healthcare.[14]

Meta's acquisition strategy has been to expand into various communication and social media domains, with a notable number of patents acquired from WhatsApp - and more recently hardware (VR). Amazon's pattern shows a large number of patents acquired from Zoox (self-driving cars),

[13] Beats have only six patents in our Apple dataset which can be traced back to Beats.
[14] Compare Microsoft (2021) press release, with the market definition adopted by UK Competition and Markets Authority (2021).

followed by Audible and Twitch, pointing towards a diversification strategy, enhancing their portfolio in autonomous vehicles, audio content, and gaming respectively.

Another insight into the motivation and nature of acquisitions comes from their over-representation of expired patents in previously acquired companies. Acquisitions account for 10% of patents, yet 20% of expired patents. This may be because the assets and patents are from mature companies with technology necessary for Big Tech to compete against established firms. Or it could be due to risky investments made in young companies often with unproven technologies. Their assets may also compete with internal assets, rendering one set of assets redundant. Although Microsoft holds the most number of expired patents, this is only because it has the most number of total patents. In fact, Alphabet has by far the largest proportion of expired patents coming from acquisitions at 40% (3,319 patents), followed by Meta at 24% (78 patents). This is driven by Alphabet's acquisition of patents from Motorola Mobility. Although Google sold Motorola Mobility to Lenovo in 2014, it kept the majority of patents (Google 2014). If Google used these shelved patents to sue rivals to prevent market entry, for example, then this is a variation of the killer acquisition story (Areeda et al. 2023, Section 701c).[15]

48 patents from Google's acquisition of Makani Power (airborne wind turbine kites) have expired (out of 80). Our dataset indicates that those which originated with Google's "x development" as the owner remain active, even if still registered as the assignee to Makani. The expired patents were all granted between 2009 and 2018. It is tempting to say they likely reflect outdated or unsuccessful technologies. But several of the expired patents, for example, US8800931B2 and US20150251754A1, have dozens of forward citations, reflecting an advanced efficient energy extraction model. However, the company itself appears to have been a drain on Google's resources, and it is unclear if it was a good fit for Google in the first place, highlighting the non-strategic nature of the acquisition. In February 2020, Alphabet decided to shut down Makani, as part of Alphabet's broader effort to wind down its "Other Bets" and focus on projects with a clearer path to commercial viability (TechCrunch 2020). Was this a "killer acquisition" because the assets were shut down at Google's X (Areeda et al. 2023, Section 701c)? No. These and other assets were not intentionally shut down to squash a competing technology.

A different conclusion might be warranted for Alphabet's purchase of Fitbit, a highly innovative

[15] ibid:"The externally acquired but later unpracticed patent is a variation on the killer acquisition story, which dates back to the Supreme Court's 1908 Paper Bag decision."

company (226 patents now held by Alphabet) with an average of 61 forward citations per patent today. Only three of Fitbit's 226 patent families Google has let expire, including patent family containing the highly innovative patents US7641124B2 and US20070241201A1. Further research is needed to assess the motivation for this. But if it contained innovative features which could potentially undermine Google's existing capabilities then this was a (partial) killer acquisition story. We do know that Alphabet slowly killed FitBit by limiting functionality and transferring its core tech to the Google Pixel Watch. This removed a competitor from the market and with it price and product competition.

The existing literature provides additional insight into the motivation of Big Tech's acquisitions. One-third of the (unreported - non-HSR) majority-stake corporate acquisitions by Big Tech between 2010-2019 were motivated, by acquiring patents (12.5%) and (non human-capital) assets (20.6%), according to firms' self-reporting to the FTC (2021).[16] Though this categorization suffered from definitional issues. Evidence on a "killer acquisition" motivation by Big Tech is mixed (Cunningham et al. 2021; Rinehart 2023). From a capabilities perspective, active integration of acquired technologies by Big Tech – instead of "killing' them – is supported at the patent-level by Gugler et al. (2023). Using a time-series of patent citations in technological classes, the authors find that patent citations increase by 36% for Big Tech's acquisitions (except for Apple's) after 2010 (compared to a control group), in marked contrast to acquisitions before then.

3 Acquired Capabilities and Potential Competition: Evidence from Big Tech's Patents

This section presents evidence showing a strong link between capabilities acquisition and entry in the market for goods and services. Individual acquisitions do not always make a difference but in combination they do. Excessive concentration in capabilities, especially cross-cutting ones (ML) or complementary combinations (product ecosystem), might be viewed with caution in a competitive framework.

Empirically, what have been the key capabilities acquired by Big Tech? Our patenting approach requires using technological categories. Various machine learning approaches have been proposed, including applying BERT to patents (Srebrovic et al. 2020). The US Patent Office (USPTO) (A.

[16] Likely underestimated. Figure 1 and footnote 49 noting patent motivation only if no other motivation present. Noting majority control as its own motivation separate to acquiring assets or patents. We exclude the 45 minority interest stakes, 39 non-corporate interests, and 8 licensing agreements from this calculation.

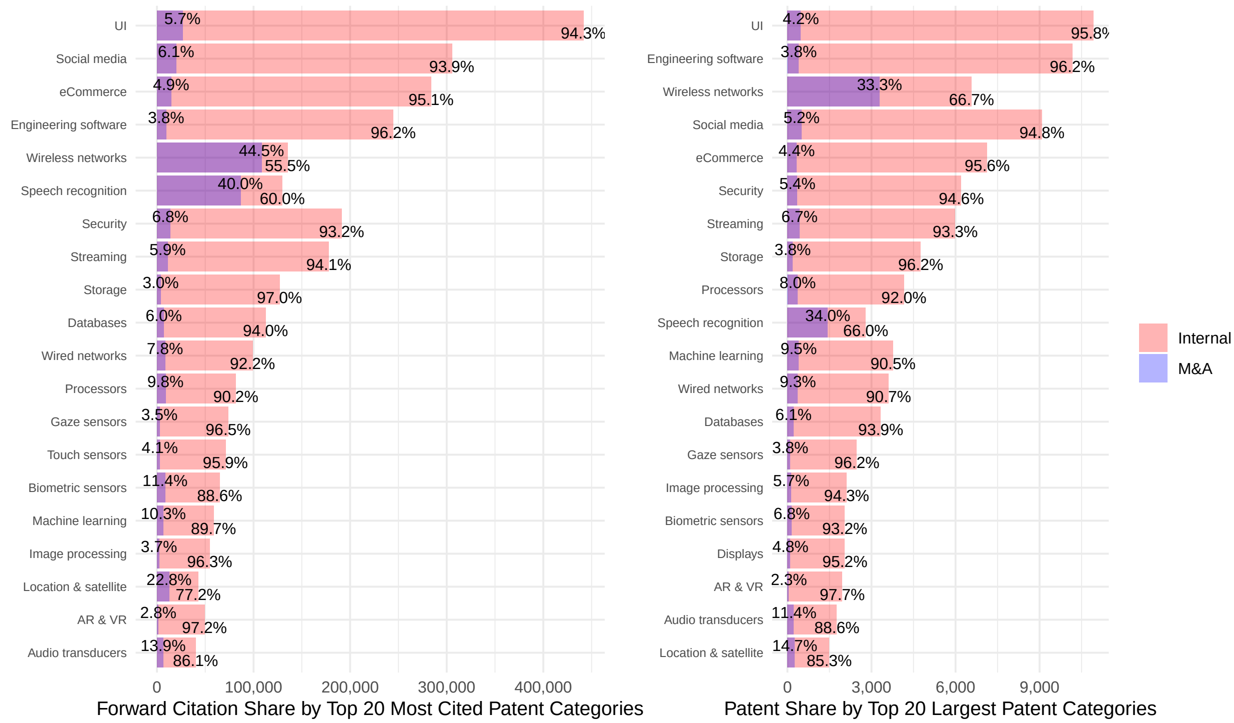
Toole et al. n.d.) uses the machine learning approach outlined in A. A. Toole et al. (2019) and Abood et al. (2018) - also known as "patent landscaping". We use the ML technology classifier from Cipher / LexisNexis Patent Research which is, as noted previously, an unsupervised learning based multi-class classifier, using 10 major technological categories and 122 subcategories, on a global basis. Within this, we focus on key technology categories. These same technology categories have been used by Big Tech who historically have used Cipher's patent dataset.

Figure 3 shows that, across all of Big Tech, user interface (UI), social media, and eCommerce related technologies dominates the total patents held by Big Tech (adjusted for citations), and unadjusted they are UI, engineering software, and wireless networks. The Figure highlights the importance of M&A to building out Big Tech's capabilities in speech recognition (40% of all citation adjusted patents), wireless networks (44.5%), location & satellite (22.8%) i.e. Google and Apple Maps (left hand graph, adjusted for patent importance i.e. forward citations). But it has also been important to other hardware, including audio transducers (13.9%) used for speakers, processors (9.8%), and biometric sensors (11.4%). Machine learning patents also feature (10.3%).

Looking at the above by each Big Tech company separately (Figure 4), shows considerable differences in the use of M&A to acquire capabilities and compete in new and uncertain markets. Alphabet shows the greatest reliance on M&A for its capability development. This tracks its products fairly closely. For example, 72% of wireless networks and 46% of location & satellite citation adjusted patents come from M&A. The latter underpins Google Maps. Such patents have been important to Amazon too. Amazon has used M&A to gain robotics capabilities (Zoox) but also streaming (Twitch) and machine learning (14%). As has Alphabet (12%), Microsoft (10%), and Meta (4%). Processors (6%), wireless networks (33%), and antenna (9%) have been important areas where Apple has used M&A to develop its capabilities. Similar to Microsoft, except for Microsoft's speech recognition (its acquisition of Nuance Communications).

The above technology categories are the largest patent categories held by Big Tech. These categories are different to the capabilities most acquired through M&A by Big Tech, which have tended to focus on practical hardware technologies to build out the 'things' which their software has filled. Looking at technology categories where at least 20% of forward citations are from M&A patents and total patent citations in that category are above the median of 2210 (so as to avoid very small categories): amplifiers (52.5%, total forward citations 12,498)*, Wireless Networks (44.5%, 243,468)*,

Figure 3. Capabilities Acquired by Major Technology Category (Adjusted for Forward Citations vs. Patent Totals)



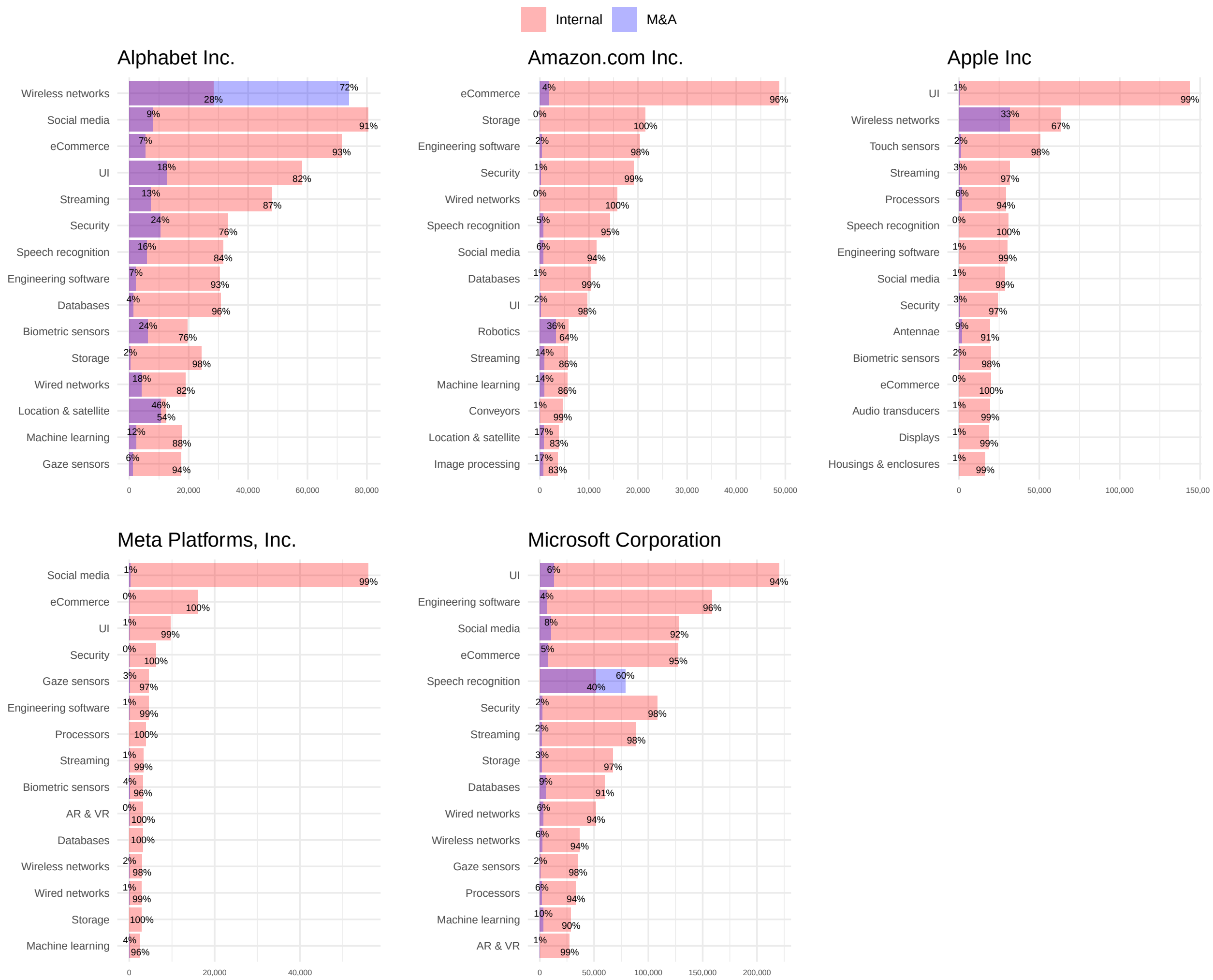
Note: *Top 20 UTT Technology categories held by Big Tech as a whole, by categories with largest total patent count. Showing share due to M&A and share to internal development.*

Photovoltaics (41.1%, 2,629), Speech recognition (40%, 216,322)*, Antennae (29.2%, 42564), ADC & DAC(27.1%, 8,419)*, Hinges (25.1%, 4,385), Batteries (23.5%, 9,277), Location & Satellite (22.8%, 55,357), and Inductors (21.2%, 13,877). A star indicates that the technology category is also meaningful when looking at total patent counts, since the share of M&A patents is large and there are a large enough number of patents in the category too.[17] In terms of what these capabilities mean practically:

- Amplifiers (52.5%): Used in smartphones (e.g., iPhone's audio system), laptops (like MacBook's speakers), and smart speakers (such as Amazon Echo or Google Home) to enhance audio output.
- Wireless Networks (44.5%): Essential in smartphones (like Samsung Galaxy or Google Pixel), smart home devices (e.g., Google's Nest Thermostat), and wearable technology (like Apple Watch) for connectivity.
- Photovoltaics (41.1%): Used in powering data centers (like those of Amazon Web Services) and in consumer products like solar-powered charging devices.

[17] Must be above median total patent count of at least 166 and at least 20% of total patents in the given category are from M&A.

Figure 4. Capabilities Acquired by Major Technology Category by Big Tech (forward citations)



Note: *Top UTT Technology categories held by each Big Tech, by categories with largest total patent count by forward citations. Showing share due to M&A and share to internal development.*

- Speech Recognition (40%): Integral to voice assistants (like Apple's Siri, Amazon's Alexa, or Google Assistant) used in smartphones, smart speakers, and smart home devices.
- Antennae (29.2%): Crucial in smartphones, wireless routers, and internt of things (IoT) devices for communication and signal reception.
- ADC & DAC (Analog-to-Digital and Digital-to-Analog Converters) (27.1%): Found in smartphones, audio equipment (like Apple's AirPods), and high-definition televisions for processing audio and video signals.
- Hinges (25.1%): Used in foldable smartphones, laptops (like Macbook Airs) and convertible

laptops (like Chromebooks), for flexible device designs.

- Batteries (23.5%): Power smartphones, laptops, tablets (like iPads), electric vehicles, and smartwatches.
- Location & Satellite Technology (22.8%): Utilized in GPS services in smartphones, vehicle navigation systems, and in technology for global communication networks.
- Inductors (21.2%): Incorporated in electronic circuits of various devices such as smartphones, laptops, and power supplies to manage current flow and filter signals.

Next, Figure 5 explores what Big Tech's patents look like in 2010 or earlier compared with after 2010. The closer the correspondence between capabilities and product markets, the greater the potential for acquisitions of capabilities to warrant a competitive threat. Given Big Tech's access and control of existing platforms with large user (and producer) bases, as well as complementary technologies, we would expect to see a greater correspondence as the timeline from technology acquisition to consumer facing product can be dramatically shortened.

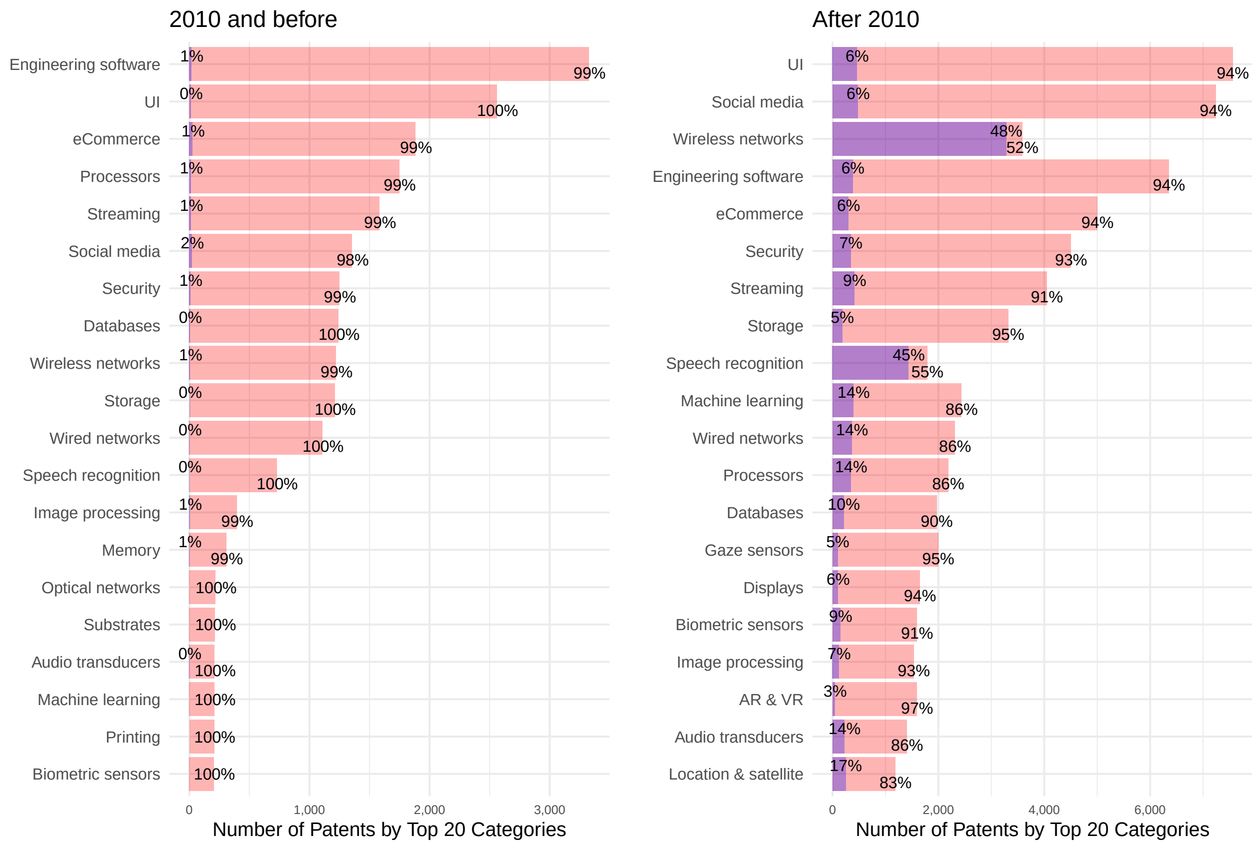
Figure 5 shows that from 2010 and before (patent) capabilities were focused on engineering software overwhelmingly (defined here as purpose-built computer code to design and document a product).[18] After 2010 much greater emphasis is put on building out engaging integrated hardware-software product ecosystems, with UI, social media, and wireless network patents dominating the top three. Storage also becomes far more important. User interfaces (UI) are when a human user interacts with a computer, website or application. Machine learning, gaze sensors (for eye tracking and gaming), displays, AR & VR, and location & satellite patent technology all become far more important post-2010. Machine learning especially shows the largest jump in patents held after 2010, relative to 2010 and prior.

M&A also becomes quantitatively central to all capabilities development after 2010 (shaded purple). Before 2010 little importance is attached to M&A in Big Tech's technological development in our dataset. But in practise, its qualitative impact was considerable. Consider, for example, Google's acquisition of DoubleClick, which helped Google dominate the advertising technology stack (New York Times 2020).

Lastly, it is useful to adopt a more qualitative approach to capabilities acquired (focusing not on total patents acquired but on the most valuable patents acquired). Practically, we can look at median

[18] This class includes electronic design automation (EDA, ECAD) for semiconductor design, Logic simulation, In-circuit emulation (ICE), Software compilers and Computational fluid dynamics (CFD). It excludes printed circuit board design software which is captured by the PCB class.

Figure 5. Diversification in Capabilities through M&A Leads to Ecosystems



Note: *Showing Top 20 Patent Categories owned by Big Tech by total patent count before or equal to 2010 and after 2010. Year is calculated.*

forward citations by company acquired. The top four Big Tech acquisition on this metric have been:

- Apple - FingerWorks (acquired 2005): With a median forward citation in acquired patents of 871.5, the high citation count indicates its technology's significant impact. The expertise from FingerWorks played a pivotal role in the development of the multi-touch interface for the iPhone and later products.

- Amazon - Evi (acquired 2012): Median forward citation is 220. Evi was a voice-powered assistant. Its technology was integrated into Amazon's Alexa platform, enhancing its capabilities.

- Alphabet - Teracent Corp (2009): With median forward citations of 209, Teracent specialized in personalized display advertising. Google utilized Teracent's technology to refine and personalize display ads on its platform.

- Microsoft - Groove Networks Inc (acquired 2005): Median Forward citations of 175. Groove Networks developed collaboration software which was later integrated into "Microsoft Office

Groove" and eventually became part of "SharePoint Workspace."

3.1 Regulating "Maps" or Location & Satellite Technology?

Below we explore the role of M&A in Google gaining and sustaining an advantage in Mapping capabilities and product markets. Google Maps grew out of three acquisitions in 2004: Keyhole, ZipDash, and especially Where 2 Technologies (Vox 2015; Gilbert et al. 2019). These were all young firms with few if any registered patents. But then Google's acquisition of Israeli mapping company Waze in 2013 - ultimately approved by the FTC, and the UK and Israel competition authorities[19] - gave Google access to further technological dominance, through crowd sourced real-time traffic data.[20]

Using Cipher's UTT ML classifier, we can see that Alphabet has twice the number of patent families (758) in "Location & satellite" technology than any other Big Tech firm in our dataset (Microsoft with 375, followed by Apple with 347).[21] Though Apple has engaged in dozens of acquisitions to try and compete with Google in Maps, including acquiring Placebase, Poly9, C3 Technologies, WiFiSlam, Locationary, HopStop.com, Embark, BroadMap, Spotsetter, Coherent Navigation, Mapsense, and Indoor.io (Wikipedia 2023).[22]

One could include the other leaders in the market from outside of Big Tech to see their relative shares of "Location & satellite" technology. One could also narrow this down further by those who compete in a given product. We can apply this same UTT classifier, now also used by LexisNexis (2023), to the competitive landscape in this technological field for patents registered in the U.S. market. We see that Alphabet, Microsoft, Apple do feature but are outside of the top three. Alphabet is fourth, behind Qualcomm, Boeing, and Toyota. Microsoft is 15th and Apple 17th (not shown). This highlights that using technology alone to classify markets increases the potential size of the market and in turn the potential competition which the market could experience. But not all competitors are as likely to enter a given product market with their technology competencies. However, an acquisition which is complementary might allow this rapidly if they have pre-existing related competencies.

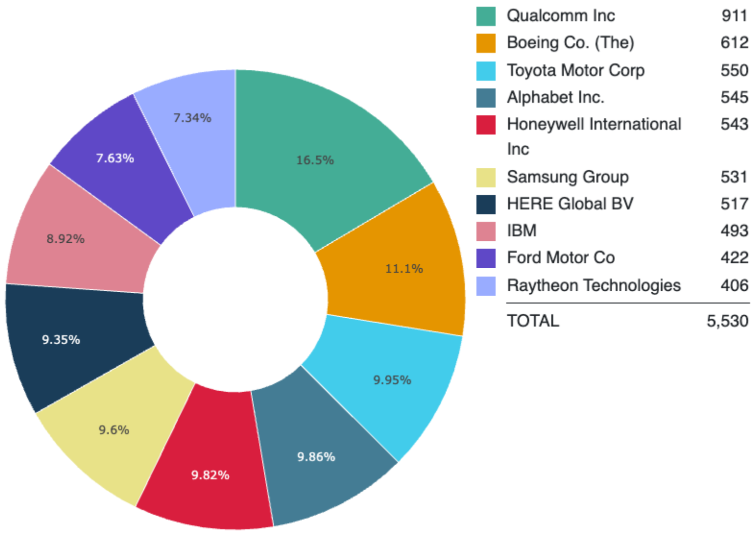
[19] In 2020 the FTC announced that it would re-examine Google's 2013 acquisition of Waze.

[20] An analysis of Google's patent assignment history shows at least 14 individual patents which Google acquired from Waze. While some of these patents were rejected others are active with dozens of forward citations, implying high levels of innovation. A closer analysis of the CPC codes of these patents (linked to patent families) could reveal the extent to which Waze's technology provided Google with complementary advantages or reinforced existing ones.

[21] Adjusting for forward citations does not change this advantage much: Alphabet's lead diminishes only slightly, while Amazon's share falls considerably.

[22] Of these we can trace easily the (few) patents Apple has to three of them in our database: Mapsense, Coherent Navigation Inc, and Embark Inc. Many of the patents have been reassigned with little trace unless one digs into a full assignee history for each patent. HopStop.com navigation patents, for example, are alive and well with Apple under patent family C0048964758 cipher family id (patent number US7957871B1), and for Locationary patent family C0003299803 (patent US9767137B2).

Figure 6. Location and Satellite Patent Ownership (Capabilities) in USA, Top 10 Shares



Note: *Showing active patent family owners for USA for location and satellite technologies using a UTT ML patent classifier. This share excludes the next 5,000 patent owners from the market. This includes: Technologies related to satellite communication, geocentric orbit type satellites, remote sensing satellites and global positioning satellite (GPSS) and terrestrial based location technologies. For more information see LexisNexis (2023).*

4 Do Capability Acquisition Strategies Differ by Big Tech Firm?

Using a logit regression, this section estimates what drives the probability of each Big Tech firm acquiring a patent externally. This builds on the dynamic capabilities literature, with its emphasis on "the firm's ability to integrate, build, and reconfigure internal and external competences to address rapidly changing environments" (D. J. Teece et al. 1997). The binary outcome variable is whether the patent is "externally" (i.e. via M&A) or internally developed. This treats each patent as largely independent from one another, focusing on each patent as a capability, potentially separable from the firm and capable of combination with other patents - both internal and external to the firm. The analysis is not at the level of acquired firms then but at the level of acquired patents, or capabilities.

We have two main sets of predictors for this binary outcome, which correspond to two distinct approaches to acquiring capabilities:

- *Hypothesis 1*: Big Tech firms acquire high performing, often proven, assets. This is proxied by patents which have high age-adjusted forward citations being more likely to be acquired. (Note that we use log age-adjusted citations as of December 2022 rather than at the time of acquisition.)

- *Hypothesis 2*: Big Tech firms acquire assets based more on their technology field, either to complement specific assets and/or to enter into completely new fields of production. These assets might have a wide variability in their proven commercial applications and viability. We proxy for this by use of technology field.

The patent-level model predicting the probability of a patent being acquired through M&A can most simply be expressed as follows:

$$P(\text{External} = 1) \sim \text{Bernoulli}(\pi) \tag{1}$$

$$\text{logit}(\pi) = \ln\left(\frac{\pi}{1-\pi}\right) = \beta_1 \times \text{Citation\_log} + \beta_2 \times \text{Technology\_field}, \tag{2}$$

where π represents the probability of `External` being 1. The first equation indicates that the patent response variable `External` follows a Bernoulli distribution with success probability π. The second equation, representing the logit link function, transforms this probability into a linear combination of the predictors: `Citation_log` and `Technology_field`. We use the `glm` function in R with a `binomial` family and `logit` link function. These predictors are in practise estimated at the patent level, subscript β_i, and are estimated in separate regressions for each Big Tech firm β_j (of which there are five), and so could be written as $\beta_{j[i]}$ for clarity - even though estimated separately. Patent sample sizes vary by each Big Tech firm, as shown in the results Table 4. We are unable to control for the acquired firm from which the patents originate, as their number is disproportionately large compared to the patents.

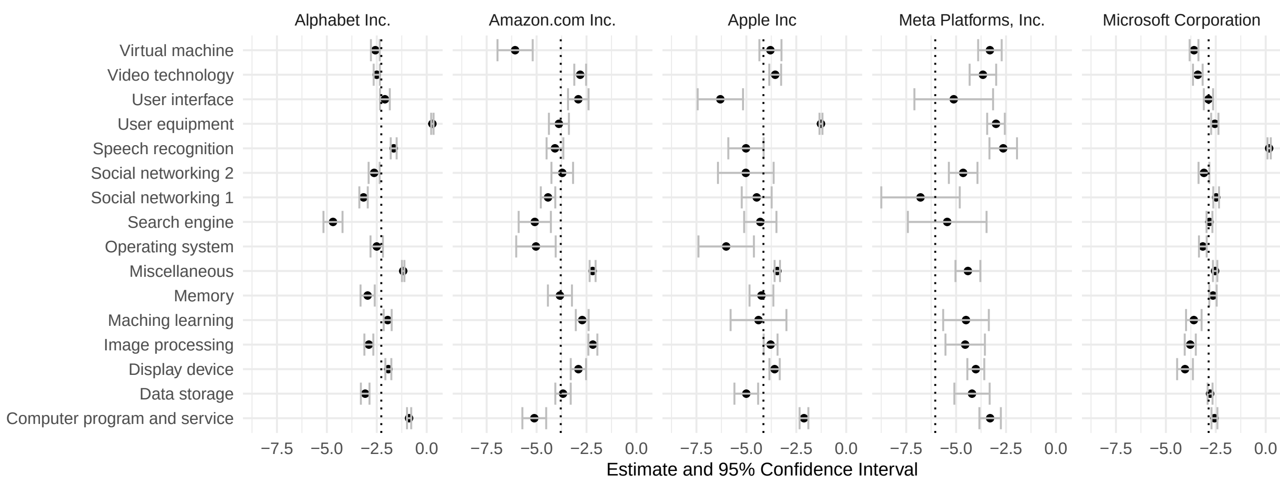
Patent data is from Cipher/LexisNexis, including the UTT ML technology category used in the regression. The M&A indicator based on a matching from our M&A database combining Refinitiv, Wikipedia, and Webscraping. Results are shown below in Table 4 for hypothesis 1.

Hypothesis 1 seems to capture Amazon (.48), Apple (.22), and to a lesser extent Meta's (0.19 but with greater uncertainty) acquisition behaviour of capabilities, since for these firms a patent's age-adjusted citations predicts well the probability that the firm will acquire a given patent. This is not true for Alphabet and Microsoft. For Alphabet the estimate is simply too uncertain. While for Microsoft the opposite seems to be true (as the coefficient is negative). The results for Alphabet may be driven by the speculative (or simply the varied) nature of many of its past acquisitions, with over 17% of its patents coming from external acquisition. For both Alphabet and Microsoft, hypothesis 2 instead seems to better capture their capabilities development through M&A, as shown in Figure 7.

Table 4. Comparison of GLM Models: Predicting patent being acquired through M&A

	DV: Probability of a patent being from M&A				
	Alphabet	Amazon	Apple	Meta	Microsoft
log(Patent Citations)	-0.003	0.48^{***}	0.22^{***}	0.19^{**}	-0.16^{***}
	(0.014)	(0.036)	(0.022)	(0.076)	(0.016)
Observations (n)	29,273	14,670	23,605	6,618	41,297
Note:	*p<0.1; **p<0.05; ***p<0.01				

Figure 7. How Important is Technology Category to Acquisition Motivation?



Note: *Vertical dotted line is total (global) average effect across all technology categories showing the average tendency (probability) for a patent to be externally acquired. This is highest for Alphabet followed by Microsoft (though still negative given most patents are internally developed in our dataset). Showing estimates from separate logit regressions. Omitting two technology field estimates from Meta due to the confidence interval being too wide (smaller sample size). Patent data is from Cipher/LexisNexis including UTT ML technology category. M&A indicator based on a matching from our M&A database combining Refinitiv, Wikipedia, and Webscraping.*

Alphabet has the highest baseline probability of buying a company across all technology categories (dotted vertical line crossing at -2.4), followed by Microsoft. Unlike the other Big Tech firms, Microsoft and Alphabet also have very strong technological preferences, buying patents in specific technology areas with positive probability, in user equipment for Alphabet and in speech recognition for Microsoft. Apple also has a strong preference for buying capabilities in user equipment. This highlights that for these firms, their dynamic capabilities are best enhanced through acquiring specific technological areas, irrespective of the proven efficacy of those capabilities in isolation from Big Tech.

5 Conclusion and Policy Implications

Our empirical framework, utilizing an extensive patent dataset combined with M&A data, differentiates between Big Tech's acquired and internally developed patents. This approach offers insight into Big Tech's strategic behaviour, underscoring the role of acquired technologies in shaping future competition. Our study underscores the empirical feasibility of focusing on capabilities, as patents, within a dynamic competition framework. The interplay between capability acquisition, market entry, and dominance is complex (Argente et al. 2020) and highlights the influence of M&A in sectors like digital mapping, where Alphabet has fortified its position through multiple small and one major strategic acquisitions.

One implications of our research is that patent data can, in many instances, be useful in undertaking a capabilities-based competition assessment (D. J. Teece 2012; Cadman 2023). Our empirical findings are consistent with several arguments on dynamic competition. First, dynamic competition is often disrupted more by innovative small firms (from outside the market) than by traditional competitors in stable markets (Hovenkamp 2008). The challenge for regulators lies in assessing the impact of these smaller firms, often pioneers in emerging technologies, on future markets. Our patent data provides limited insights into these entities though, underscoring the need for nuanced regulatory approaches that consider potential market impacts from nascent technologies.

Second, markets with rapid innovation and growth are high-risk areas where market definitions might be better framed using broader technological categories – though still connected to actual or potential product markets (Cadman 2023).

Third, a dynamic competition perspective based on technological categories suggests that existing market leaders might not be as dominant as presumed. External firms, especially those operating in high-innovation sectors, could pose substantial competitive threats. A thorough examination of market entry barriers and factors restricting market contestability is crucial for a comprehensive market analysis. In conclusion, our findings advocate for a more dynamic, capabilities-focused analysis of Big Tech's acquisitions, both present and historical. This approach contributes to a deeper understanding of the competitive landscape and the strategic decisions shaping Big Tech's market influence.

References

Abood, Aaron and Dave Feltenberger (2018). "Automated Patent Landscaping." *Artificial Intelligence and Law* 26.2, pp. 103–125.

Affeldt, Pauline and Reinhold Kesler (2021). "Big Tech acquisitions — towards empirical evidence." *Journal of European Competition Law & Practice* 12.6, pp. 471–478.

Akcigit, Ufuk and Sina T Ates (2023). "What Happened to US Business Dynamism?" *Journal of Political Economy* 131.8, pp. 2059–2124.

Areeda, Phillip and Herbert Hovenkamp (2023). *Antitrust Law: An analysis of antitrust principles and their application*. Online Edition.

Argente, David et al. (2020). "Patents to Products: Product innovation and firm dynamics." *CEPR Discussion Paper No. DP14692*.

Benkert, Jean-Michel, Igor Letina, and Shuo Liu (2023). "Startup Acquisitions: Acquihires and Talent Hoarding." *arXiv preprint*. URL: https://arxiv.org/pdf/2308.10046.pdf?preview=preview.

Bryan, Kevin A and Erik Hovenkamp (2020a). "Antitrust Limits on Startup Acquisitions." *Review of Industrial Organization* 56, pp. 615–636.

— (2020b). "Startup Acquisitions, Error Costs, and Antitrust Policy." *The University of Chicago Law Review* 87.2, pp. 331–356.

Cadman, Richard (2023). "Dynamic Competition and Big Tech Mergers Report." *British Institute of International and Comparative Law*. URL: https://www.biicl.org/publications/dynamic-competition-and-big-tech-mergers?cookiesset=1&ts=1704894474.

Caffarra, Cristina, Matthew Elliott, and Andrea Galeotti (June 2023a). *'Ecosystem' Theories of Harm in Digital Mergers: New insights from network economics, part 1*. URL: https://cepr.org/voxeu/columns/ecosystem-theories-harm-digital-mergers-new-insights-network-economics-part-1.

— (June 2023b). *'Ecosystem' theories of harm in digital mergers: New insights from network economics, part 2*. URL: https://cepr.org/voxeu/columns/ecosystem-theories-harm-digital-mergers-new-insights-network-economics-part-2.

Callaham, John (2022). *Google made its best acquisition nearly 17 years ago: Can you guess what it was?* Android Authority. URL: https://www.androidauthority.com/google-android-acquisition-884194/.

Chen, Jun, Matthew Elliott, and Andrew Koh (2023). "Capability Accumulation and Conglomeratization in the Information Age." *Journal of Economic Theory* 210.

Cunningham, Colleen, Florian Ederer, and Song Ma (2021). "Killer Acquisitions." *Journal of Political Economy* 129.3, pp. 649–702.

De Vynck, Gerrit and Cat Zakrzewski (Sept. 2021). *Tech Giants Quietly Buy up Dozens of Companies a Year. Regulators are Finally Noticing*. Washington Post. URL: https://www.washingtonpost.com/technology/2021/09/20/secret-tech-acquisitions-ftc/.

Doctorow, Cory (May 2023). *Google's AI Hype Circle*. URL: https://doctorow.medium.com/googles-ai-hype-circle-6158804d1299.

Evans, David S and Keith N Hylton (2008). "The Lawful Acquisition and Exercise of Monopoly Power and its Implications for the Objectives of Antitrust." *Competition Policy International* 4.2, p. 203.

FTC (2018). "Susan Athey Remarks at FTC Hearings on Competition & Consumer Protection in the 21st Century." URL: https://www.ftc.gov/system/files/documents/public_events/1413712/ftc_hearings_session_3_transcript_day_3_10-17-18fullupdated.pdf.

— (2021). "Non-HSR Reported Acquisitions by Select Technology Platforms, 2010–2019: An FTC Study." *FTC Study*, p. 246. URL: https://www.ftc.gov/system/files/documents/reports/non-hsr-reported-acquisitions-select-technology-platforms-2010-2019-ftc-study/p201201technologyplatformstudy2021.pdf.

FTC and DoJ (2023). *2023 Merger Guidelines*. URL: https://www.ftc.gov/news-events/news/press-releases/2023/12/federal-trade-commission-justice-department-release-2023-merger-guidelines.

FTC v. Facebook, Inc. (2022). F. Supp. 3d. 2022 WL 103308 (D.D.C. Jan. 11, 2022).

Gifford, Daniel J and Robert T Kudrle (2011). "Antitrust approaches to dynamically competitive industries in the United States and the European Union." *Journal of Competition Law and Economics* 7.3, pp. 695–731.

Gilbert, Ben and David Rosenthal (2019). *Google Maps*. Acquired Podcast: Season 5, Episode 3. URL: https://www.acquired.fm/episodes/google-maps.

Glick, Mark and Catherine Ruetschlin (2019). "Big Tech Acquisitions and the Potential Competition Doctrine: The case of Facebook." *Institute for New Economic Thinking Working Paper Series* 104.

Google (2014). *Lenovo to Acquire Motorola Mobility*. URL: https://blog.google/inside-google/company-announcements/lenovo-to-acquire-motorola-mobility/.

Griliches, Zvi (1998). "Patent Statistics as Economic Indicators: a survey." *R&D and productivity: the econometric evidence*, pp. 287–343.

Gugler, Klaus, Florian Szücs, and Ulrich Wohak (2023). *Start-up Acquisitions, Venture Capital and Innovation: A Comparative Study of Google, Apple, Facebook, Amazon and Microsoft*. Tech. rep. Tech. rep., Vienna University of Economics and Business, Department of Economics.

Hall, Bronwyn H, Adam Jaffe, and Manuel Trajtenberg (2005). "Market value and patent citations." *RAND Journal of economics*, pp. 16–38.

Hovenkamp, Herbert (2008). "Schumpeterian Competition and Antitrust." *U Iowa Legal Studies Research Paper* 08-43.

— (July 2023a). *Competitive Harm and the 2023 Draft Merger Guidelines*. Promarket. URL: https://www.promarket.org/2023/07/27/herbert-hovenkamp-competitive-harm-and-the-2023-draft-merger-guidelines/.

— (Feb. 2023b). *Reclaiming the Antitrust Law of Potential Competition Mergers*. Promarket. URL: https://www.promarket.org/2023/02/27/reclaiming-the-antitrust-law-of-potential-competition-mergers/.

Jacobides, Michael G, Carmelo Cennamo, and Annabelle Gawer (2018). "Towards a Theory of Ecosystems." *Strategic management journal* 39.8, pp. 2255–2276.

Jacobides, Michael G and Ioannis Lianos (2021). *Regulating Platforms and Ecosystems: An introduction*.

Kim, Bongsun et al. (2016). "The Impact of the Timing of Patents on Innovation Performance." *Research Policy* 45.4, pp. 914–928.

Kopytoff, Verne (2014). *Motorola is a Gargantuan Mistake Only Google Could Afford to Make*. Time Magazine. URL: https://time.com/2650/motorola-is-a-gargantuan-mistake-only-google-could-afford-to-make/.

Levin, Richard C et al. (1987). "Appropriating the Returns from Industrial Research and Development." *Brookings papers on economic activity* 1987.3, pp. 783–831.

LexisNexis (Oct. 2023). *Universal Technology Taxonomy*. URL: https://www.lexisnexisip.com/wp-content/uploads/2023/10/Universal-technology-taxonomy-download.pdf.

Mazzucato, Mariana et al. (2023). "Regulating Big Tech: The role of enhanced disclosures." *Oxford Review of Economic Policy* 39.1, pp. 47–69.

McSweeny, Terrell and Brian O'Dea (Feb. 2018). "Data, Innovation, and Potential Competition in Digital Markets — Looking Beyond Short-Term Price Effects in Merger Analysis." *Antitrust Chronicle 7, 11 (Competition Policy International)*.

Microsoft (Apr. 2021). *Microsoft accelerates industry cloud strategy for healthcare with the acquisition of Nuance*. URL: https://news.microsoft.com/2021/04/12/microsoft-accelerates-industry-cloud-strategy-for-healthcare-with-the-acquisition-of-nuance/.

Morton, Fiona M Scott and Carl Shapiro (2013). "Strategic Patent Acquisitions." *Antitrust LJ* 79, p. 463.

New York Times (Sept. 2020). *This Deal Helped Turn Google Into an Ad Powerhouse. Is That a Problem?* URL: https://www.nytimes.com/2020/09/21/technology/google-doubleclick-antitrust-ads.html.

O'Reilly, Tim, Ilan Strauss, and Mariana Mazzucato (2023). *Algorithmic Attention Rents: A theory of digital platform market power*. Tech. rep. Working Paper Series (IIPP WP 2023-10). URL: https://www.ucl.ac.uk/bartlett/public-purpose/wp2023-10.

Pavitt, Keith (1982). "R&D, Patenting and Innovative Activities: A statistical exploration." *Research Policy* 11.1, pp. 33–51.

Petit, Nicolas and David Teece (2020). "Taking Ecosystems Competition Seriously in the Digital Economy: A (Preliminary) Dynamic Competition/Capabilities Perspective." *Capabilities Perspective (December 9, 2020)*.

Petit, Nicolas and David J Teece (2021). "Innovating Big Tech Firms and Competition Policy: Favoring dynamic over static competition." *Industrial and Corporate Change* 30.5, pp. 1168–1198.

PhoneArena (2022). *Qualcomm patent puts wrench into Apple's plan to build 5G modem*. URL: https://www.phonearena.com/news/qualcomm-patent-puts-wrench-into-apples-plan-to-build-5g-modem_id141078.

Posner, Richard (2006). "Do We Have Too Many Intellectual Property Rights?" *The Eighth Annual Honorable Helen Wilson Nies Memorial Lecture in Intellectual Property Law*.

Pundziene, Asta, Sohvi Heaton, and David J Teece (2019). "5G, Dynamic Capabilities and Business Models Innovation in Healthcare Industry." *2019 IEEE International Symposium on Innovation and Entrepreneurship (TEMS-ISIE)*. IEEE, pp. 1–8.

Reuters (Oct. 2022). *Apple loses second bid to challenge Qualcomm patents at U.S. Supreme Court*. URL: https://www.reuters.com/technology/apple-loses-second-bid-challenge-qualcomm-patents-us-supreme-court-2022-10-03/.

Rinehart, William (2023). "Are 'Killer Acquisitions' by Tech Giants a Real Threat to Competition?" *The Center for Growth and Opportunity*.

Saltiel, Joseph (2019). *Five Years After Alice*. WIPO Magazine. URL: https://www.wipo.int/wipo_magazine/en/2019/04/article_0006.html.

Schumpeter, Joseph A (2003). *Capitalism, Socialism and Democracy*. Routledge.

Sharma, Ruchir (2023). *What's Wrong with Tech Giants Riding the AI Wave*. URL: https://www.ft.com/content/31dc322a-b2a1-4ef3-9a9d-d7268d58fac5.

Sidak, J Gregory and David J Teece (2009). "Dynamic competition in antitrust law." *Journal of Competition Law and Economics* 5.4, pp. 581–631.

Spiceworks (2022). *Qualcomm-Apple 5G Modem Deal*. URL: https://www.spiceworks.com/tech/innovation/news/qualcomm-apple-5g-modem-deal/.

Srebrovic, Rob and Jay Yonamine (2020). "Leveraging the BERT algorithm for Patents with TensorFlow and BigQuery." *White paper*.

TechCrunch (Nov. 2011). *Amazon Snaps Up Yap and Its Voice Recognition Technology*. URL: https://techcrunch.com/2011/11/09/amazon-snaps-up-yap-and-its-voice-recognition-technology/.

— (Aug. 2013). *Facebook Acquires Mobile Technologies, Speech Recognition, and Jibbigo App Developer*. URL: https://techcrunch.com/2013/08/12/facebook-acquires-mobile-technologies-speech-recognition-and-jibbigo-app-developer/.

— (Feb. 2020). *Alphabet Takes the Wind out of its Makani Energy Kites*. URL: https://techcrunch.com/2020/02/18/alphabet-takes-the-wind-out-of-its-makani-energy-kites/.

Teece, David J (2007). "Explicating dynamic capabilities: the nature and microfoundations of (sustainable) enterprise performance." *Strategic Management Journal* 28.13, pp. 1319–1350.

Teece, David J (2012). “Next-Generation Competition: New concepts for understanding how innovation shapes competition and policy in the digital economy.” *JL Econ. & Pol’y* 9, p. 97.

— (2018). “Profiting from Innovation in the Digital Economy: Enabling technologies, standards, and licensing models in the wireless world.” *Research Policy* 47.8, pp. 1367–1387.

Teece, David J, Gary Pisano, and Amy Shuen (1997). “Dynamic capabilities and strategic management.” *Strategic management journal* 18.7, pp. 509–533.

Toole, Andrew et al. (n.d.). “Inventing AI: Tracing the diffusion of artificial intelligence with US patents.” *US Patent and Trademark Office, Alexandria* 5 (), p. 2020.

Toole, Andrew A et al. (2019). “The Promise of Machine Learning for Patent Landscaping.” *Santa Clara High Tech. LJ* 36, p. 433.

Trajtenberg, Manuel (1990). “A Penny for Your Quotes: Patent citations and the value of innovations.” *The Rand journal of economics*, pp. 172–187.

UK Competition and Markets Authority (2021). *Microsoft-Nuance: Decision for Publication*. URL: https://assets.publishing.service.gov.uk/media/6244608fd3bf7f32ac2b9425/Microsoft-Nuance_-_Decision_-_for_publication_.pdf.

Vox (Feb. 2015). *Ten Years of Google Maps: From Slashdot to Ground Truth*. URL: https://www.vox.com/2015/2/8/11558788/ten-years-of-google-maps-from-slashdot-to-ground-truth.

Wikipedia (2023). *List of Mergers and Acquisitions by Apple*. Accessed: December 2023. URL: https://en.wikipedia.org/wiki/List_of_mergers_and_acquisitions_by_Apple.

Appendix

A Further Descriptive Statistics

Figure 8. Big Tech's Top 5 Technology Sectors with Most Patents + Machine Learning

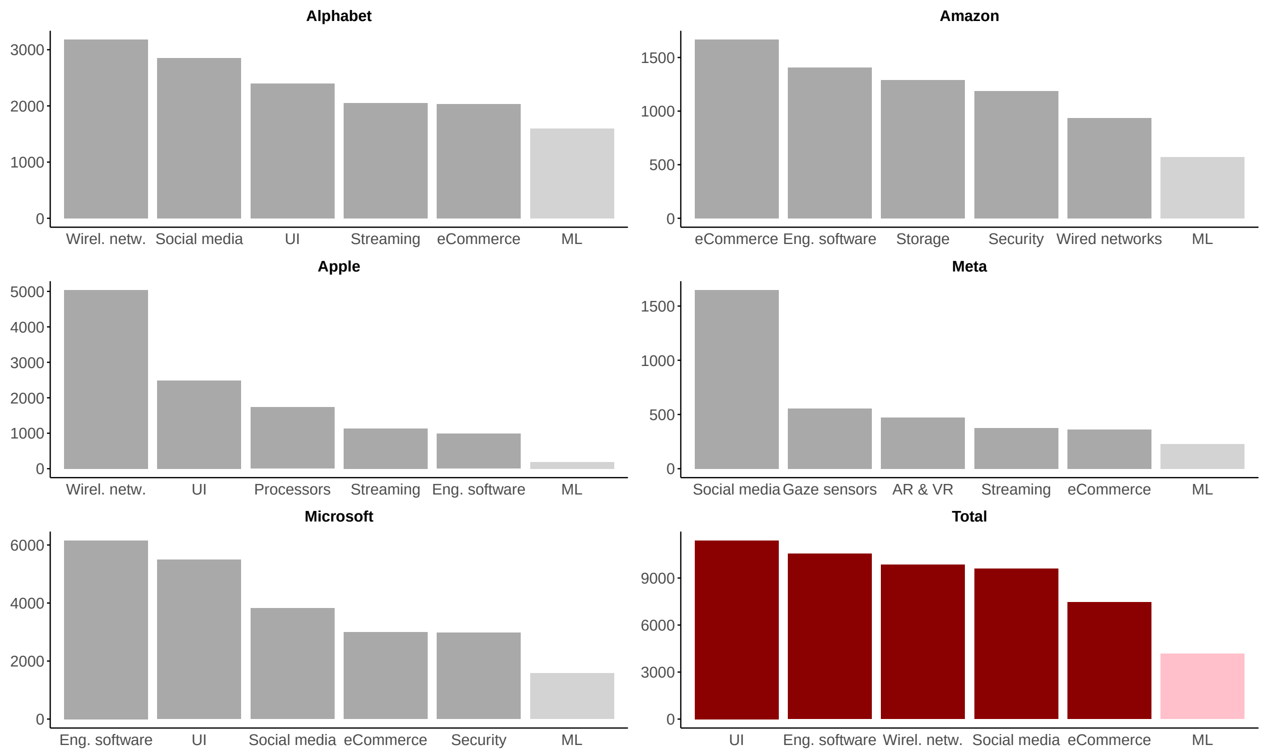


Figure 9. Big Tech's Top 5 M&A Technology Sectors with Most Patents + Machine Learning

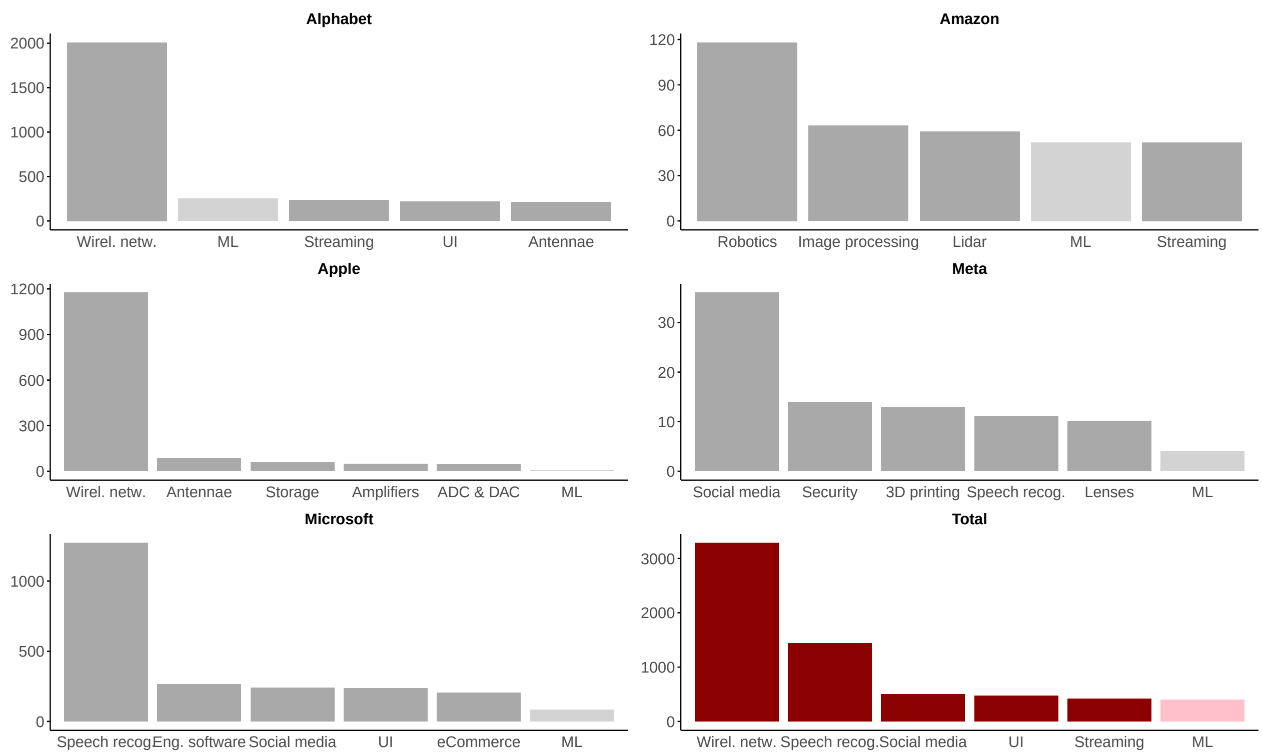
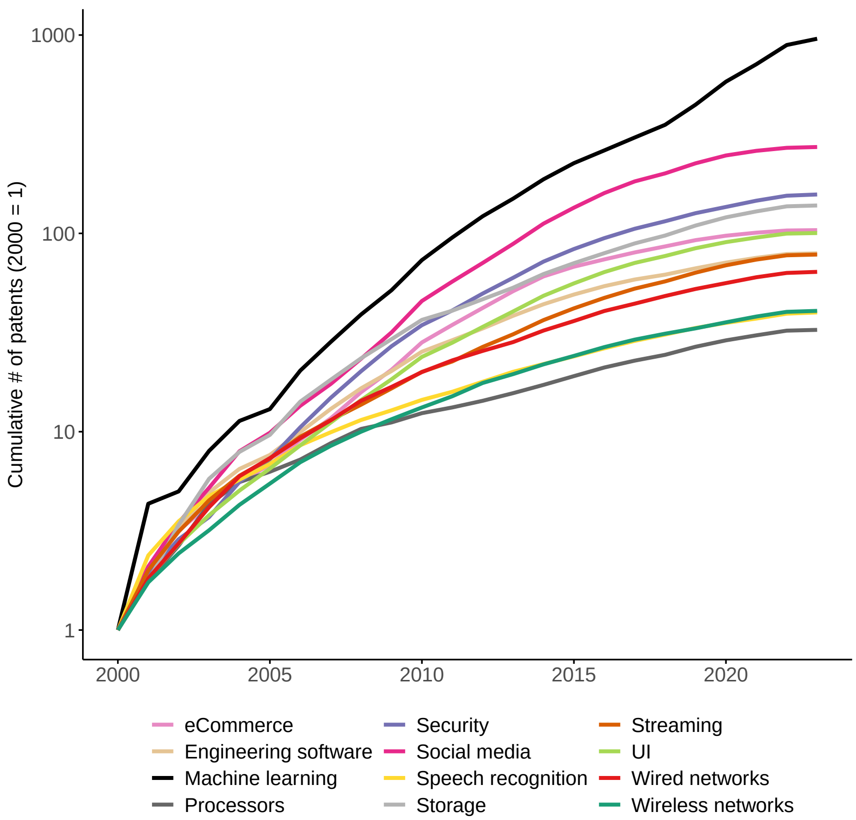


Figure 10. Cumulative Big Tech Stock of Patents by Sector



B Cleaning Cipher Data

To construct our analytical dataset on a patent-family level, we utilize Cipher - a dataset from global patent databases which offers a wealth of information such as patent owners, assignees, technology areas, original assignees, and status. Recently bought by LexisNexis, the company manually creates the owner and parent categories. They also use ML to classify technology types. The Cipher dataset also includes expired patents. This appendix describes the steps we took to clean the data.

B.1 Identifying Patent Originators

Our aim is to identify the patents that Big Tech owns as a result of acquisitions of other firms. However, we encounter three main issues in matching names in our acquisition-level dataset (Refinitiv) to Cipher. Firstly, our patent data does not clearly indicate if the patents that Big Tech owns come from another company. Secondly, it does not track assignee data, which means it may not recognize certain patents that Big Tech owns. Finally, the firm names are inconsistent between datasets, making it challenging to correctly identify each firm.

To overcome these obstacles, we merge the patent dataset and the acquisition dataset by first algorithmically creating a list of subsidiaries of Big Tech and the resulting patents assigned to them. However, we discovered three distinct scenarios that made the process more complex.

Firstly, many patents that are internally developed rely on externally acquired patents and human capital, which our study underestimates. To account for this, we ensure that any patents assigned to the acquired firm are properly matched with Big Tech, even if the assignee's name has not been formally changed. Secondly, we found that Big Tech owns patents that are not formally registered to them but have been acquired through M&A. After acquiring a company, Big Tech may retain the acquired company's patents as being registered under the acquired firm's name (assignee), with no formal change made to whom the owner of the patent is registered to. To address this scenario, we use our M&A list to match the acquired firm's patents with Big Tech, noting that the acquired firm and their patents are ultimately under the control of a Big Tech firm (even if not formally registered to them). Thirdly, we encountered scenarios where Big Tech owned patents that were registered under an internally developed subsidiary's name. A patent may be registered as belonging to a seemingly unrelated firm with no apparent links to Big Tech, but in fact, the seemingly unrelated firm was a subsidiary of Big Tech developed internally with no direct M&A links. To address this scenario, we applied our knowledge of Big Tech's corporate structure to discover the company was owned by Big Tech but not through acquisitions but instead through internal development.

We then thoroughly clean this list to address inconsistencies between the firms and other datasets we are using. For example, if a firm has "LLC" as part of its name in one dataset and not in another, we must ensure that the firms correctly match to avoid missing important data. Finally, we check this cleaned list against our M&A dataset from Refinitiv to trace the origins of the patents that Big Tech owns. This enables us to adequately demonstrate that all the companies which Big Tech has bought are actually listed as subsidiaries of Big Tech and which of the resulting patents are then assigned to Big Tech due to the acquisitions of those firms.

B.2 Merging Cipher with Acquisition Data

One of the major challenges we faced in using Cipher was to standardize owner, assignee, organization and originator names to reconcile firm names for matching with other datasets, specifically M&A events. We did so by developing a multistep cleaning algorithm which cleaned and standardized all company names. The raw data had 146, 664 patents and after processing, we were left with 127, 299 patents in the dataset.

First, we filter the dataset to contain patent information from 1980 and onward. Then, we remove all instances of ".com" from the dataset. After, the misspellings of various firms and extra words are corrected (e.g., "amazong" needs to be "Amazon" and "Mela" needs to be "Meta"). This ensures that we can accurately identify and target the appropriate firms during the cleaning process. Then, we translate foreign firm names into English to facilitate accurate detection. The names are also then stripped of punctuation (commas and periods) and capitalization. We then remove words after the key Big Tech names (e.g., microsoft corporation changes to just microsoft, apple inc changes to just apple). This isolates the big tech company's stem name (the main body of the firm name) excluding any extraneous suffixes. A challenge that arose with this step is that in a few instances, some firms could be unintentionally removed from the data, so we have to manually fix them. For example, "metaswitch networks ltd" was left to be "meta", which is not an accurate representation of the original firm name so it must be reverted to "metaswitch networks ltd". There are other firms that need manual correcting as well. After this, we then find and take out extra words, legal entity endings, abbreviations, and redundant characters. After all these steps, there are still firms that may need specific alterations to ensure proper cleaning. Finally, we split firms with multiple owners or assignees into multiple columns, which are denoted using dashes, vertical lines, and brackets.

After all the cleaning, we then run the matching script and check the matching results manually to confirm accuracy.

C Acquisition Data

We collect firm acquisition-level data from two sources. We first extract all announced and completed M&As (with complete information on acquirer and target firms) and announced and effective dates from Refinitiv which is provided by Refinitiv Desktop. This comprehensive dataset contains firms, their acquirers, ultimate parents, and deal status. However, upon further inspection, we found that this dataset was incomplete and missing many M&A firms, so we manually develop an additional dataset using the information found from Big Tech Mergers & Acquisitions Wikipedia pages. We manually compile the missing data from the Wikipedia pages into a new dataset with similar columns as Refinitiv. This dataset includes details such as the deal confirmation source, the acquirer ultimate parent, the deal status, and the ultimate parent. Furthermore, we believe some additional firms were not captured by Refinitiv and Wikipedia, so we manually search for them using various sources including news articles, company reports, and other publicly available data sources. We format the data in the same way as the Wikipedia data.

First, we clean the Refinitiv dataset. Since the same firm could appear in different databases under slightly different names, we create standardized and homogeneous names by removing extraneous words as well as stripping the names of capitalization. The process for cleaning this dataset is similar to that used for cleaning the patent-level data, but there are some key differences. With this dataset, we first add a few missing firms and then filter out deals made before 2000. Then, we simplify the names of all big tech firms (e.g., “Amazon.com Inc” becomes “Amazon”, “Facebook Inc” becomes “Meta”). Next, joint buyouts where the ultimate acquirer is not listed as big tech is filtered out. We then remove instances of “.com” and split firms that have brackets and dashes into multiple columns. After, firms where the “Deal Type” is repurchased are extracted. Firms Expedia Inc, Giphy Inc, Delta Airlines Inc and WestJet Airlines Ltd are removed. Then, commas and periods are removed from the company names, and they are all changed to lowercase. We find and take out extra words, legal entity endings, abbreviations, and redundant characters. We also remove any special characters from the end of the firm names. Again, after all the cleaning, there are still numerous firms that may need specific alterations to ensure they are correctly retained and standardized. The raw data had 918 firms before processing and afterwards had 770 firms.

Next, we standardize and refine the data from Wikipedia. To create the dataset, we copied over the tables from the Wiki pages and format the information in a spreadsheet with similar formatting to the Refinitiv dataset. We removed redundant and unnecessary columns, such as "Country", "Talent Acquired", and "Related to". Similar to how we process the Refinitiv data, we remove the deals made in 2023, extract instances of ".com", and split firms with brackets and dashes. Then, we remove the punctuation, make everything lowercase and remove extraneous words. We also correct specific firms manually as they could be incorrectly manipulated in the cleaning process. Then we match these cleaned firms with the cleaned Refinitiv data and remove duplicates. Before merging the non-duplicate firms with the Refinitiv file to create a larger M&A file, we manually check the Deal Status and deal control variable to ensure that the deals are correct and actually occurred. Confirmation sources are also included in the data. The raw data had 865 firms before processing and afterwards had 210 firms.

Additionally, we cleaned other publicly available acquisition data in the same method as the Wikipedia dataset. This dataset was created by gathering information from reliable online resources that provided the firm and acquiror details, and then formatting the data in a very similar way to the Wikipedia data. We process this third dataset the exact way we did previously with Wikipedia. All deals before 2000 were filtered out, instances of ".com" were removed and firms were split from brackets and dashes. All capitalization, punctuation and unnecessary words were removed from the names. Manual corrections are made to specific firms to ensure they are not wrongly manipulated. These fully cleaned firms are then matched with the cleaned Refinitiv and Wikipedia data and any duplicates are removed. We then manually check the Deal Status and deal control variable to make sure these events are correct. The confirmation sources are recorded and included in the data. The non-duplicates are merged with the M&A file with the deals from the Refinitiv and Wikipedia datasets. The raw data had 457 firms before processing and afterwards had 21 firms that we added to the total M&A dataset.

After we add the missing data from the other two datasets into Refinitiv to create one complete M&A dataset, the dataset has a total of 995 acquired firms until 2022 (acquisition date). To the best of our knowledge, this combined process provides the most comprehensive database of acquisitions.

With the cleaned acquisition data compiled from both sources, we can accurately link acquisition events to their respective target firms and can begin the process of merging the patent and acquisition datasets. We combine our acquisition database with the patent data through a name-matching algorithm combined with manual checks. The merged patent and acquisition data show acquisition activities in our analytical dataset with 22.5 percent of acquisitions recorded in our patent database. This means that 225 M&A events from our acquisition dataset are also present in our patent dataset.

D Merging Patent and Acquisition Data

In this section, we describe the process to merge patent and acquisition data with the Cipher patent database by matching company names with the owner, assignee, and original assignee names in the Cipher patent database. To minimize potential problems introduced by the minor discrepancies between different versions of the patent database and the M&A dataset, we run the cleaning algorithm to source the most standardized firm information. After this step, each company in the patent and acquisition database will have its original firm name and the target standardized name.

D.1 Name Standardization and Cleaning

We begin by standardizing company names in our patent and acquisition database using our developed name standardization algorithm that is described in the appendices above. As some names are misspelled or include additional characters that prevent exact matching, this cleaning algorithm homogenizes these firm names and helps to isolate the company's stem name by removing redundant words, stripping punctuation, and making all into lowercase.

D.2 The Matching Procedure

With these standardized and stem company names, we match the patent and M&A databases with the following procedure:

1. We first create a list of the columns we want to use to find matches in the Cipher patent dataset. As we are looking to match the M&A firms in the standardized target name ("target_no_brackets_undashed" column) with the potential patents they might hold, we must look for matches in Cipher's cleaned ownership, assignee, or original assignee columns. Specif-

ically, these columns are “owner1”, “owner2”, “owner3", “owner4”, “assignee1”, “assignee2”, “assignee3”, “assignee4”, “original_assignee1_no_bracket”, and “original_assignee2_bracket”.

2. Then, we initialize two data frames with empty columns to hold the matched data. For each row in Refinitiv, the algorithm searches each column in Cipher for a match.
   (a) If a match is found, it sets the value of the "match" column in the matched Refinitiv data frame to TRUE for that row in Refinitiv. It also sets the value of the "matches" column in the matched Cipher data frame to TRUE for the row(s) in Cipher that match the value in Refinitiv. Additionally, it populates the year of acquisition, the original target full name, the data source (Refinitiv, Wikipedia or other publicly available) and the cleaned target name we matched from in the matched Cipher data frame with the corresponding values from the M&A dataset.
   (b) If a match is not found, then the next row is searched.
3. After the exact matching process is complete, we calculate the percentage of matches found. We also manually check if the matches are correctly identified and if there are not false positives or false negatives.

Ex-post duplicate matches were removed to ensure that only the Big Tech acquirers patents was matched with the target firm, rather than other Big Tech companies who also might own patents from the target firm but which they did not acquire through M&A.

UCL Institute for Innovation and Public Purpose
11 Montague Street, London, WC1B 5BP

@IIPP_UCL
ucl.ac.uk/IIPP