ETLA Working Papers | 72 19.11.2019

Data Markets in Making

THE ROLE OF TECHNOLOGY GIANTS

Heli Koski
The Research Institute of the Finnish Economy and Aalto University, Finland
heli.koski@etla.fi

Mika Pantzar
University of Helsinki, Finland
mika.pantzar@helsinki.fi

Suggested citation:
Koski, Heli & Pantzar, Mika (19.11.2019). "Data Markets in Making: The Role of Technology Giants".

ETLA Working Papers No 72.
http://pub.etla.fi/ETLA-Working-Papers-72.pdf

Abstract

This paper focuses on the role of large technology companies' entry and expansion to the data-intensive market areas via their technological development and strategic acquisitions of companies. We analyze the evolvement of personal data related innovation in various data-intensive domains. We find that the ideas related to personal data are increasingly protected by patents. The growth in the numbers of personal data related patents was relatively modest from 2005 to the early 2010s, but it has intensified since 2011.

Large technology companies' entry to various new market areas is reflected in an exponential increase in patent applications particularly in the artificial intelligence domain. Furthermore, we find that the number of artificial intelligence/data analytics companies acquired by the data giants has escalated during the 2010s. Patent and acquisition data further echo technology giants' intentions to expand their activities into the financial and personal health services. Overall, the data show the data giants' buyouts are frequently targeted to companies active in the markets outside their core business. Our analysis illustrates how the divergencies in the data giants' innovation activities and strategic acquisitions have led them to each conquer their specific areas of dominance in the global markets for data.

Tiivistelmä

Datamarkkinoiden synty: teknologiajättien rooli

Tämä tutkimus keskittyy globaalien teknologiajättien tuloon dataintensiivisille markkinoille ja toiminnan laajentamiseen niillä teknologisen kehityksen ja strategisten yritysostojen kautta. Tarkastelemme henkilödataan liittyviä innovaatioita useilla dataintensiivisillä aloilla. Aineisto osoittaa, että henkilötietoihin liittyviä ideoita suojataan kasvavissa määrin patenteilla. Henkilötietoihin liittyvien patenttien määrän kasvu oli vaatimatonta vuodesta 2005 aina 2010-luvun alkuun asti, mutta on kiihtynyt vuoden 2011 jälkeen.

Teknologiajättien uusien markkinoiden valtaaminen heijastuu eksponentiaalisena kasvuna erityisesti niiden tekoälyyn liittyvien teknologioiden patentoinnissa. Lisäksi datajättien tekoäly- ja data-analytiikkayrityksiin kohdistuneiden ostojen määrässä on tapahtunut huomattavaa kasvua 2010-luvun aikana. Patentti- ja yritysostosdata heijastelevat teknologiajättien toiminnan laajentumista rahoitus- ja terveysaloille. Kaiken kaikkiaan, data osoittaa teknologiajättien kohdistaneen ostojaan usein yrityksiin, jotka ovat aktiivisia niiden ydinliiketoiminta-alueiden ulkopuolella. Aineistoanalyysimme havainnollistaa, kuinka datajättien erilaiset painopisteet innovaatiotoiminnassa ja strategisissa yritysostoissa ovat johtaneet kunkin yrityksen tiettyjen erityisalueiden valloitukseen ja dominanssiin globaaleilla datamarkkinoilla.

Dr.Sc. (Econ) **Heli Koski** is a Research Director at The Research Institute of the Finnish Economy and a Research Director at Aalto University.

Dr.Sc. (Econ) **Mika Pantzar** is a Professor at the University of Helsinki.

KTT **Heli Koski** on tutkimusjohtaja Elinkeinoelämän tutkimuslaitoksessa ja tutkimusjohtaja Aalto yliopistossa.

KTT **Mika Pantzar** on professori Helsingin Yliopistossa.

Acknowledgement: The authors are grateful for Business Finland for its financial support.

Kiitokset: Kiitämme Business Finlandia tutkimuksen rahoittamisesta.

Keywords: Data economy, Innovation, Patents, Acquisitions, Technology giants

Asiasanat: Datatalous, Innovaatiot, Patentit, Yritysostot, Teknologiajätit

JEL: G34, L12, L25, O33

1. Introduction

In 2000, only about one quarter of the world's data resources were stored in a digital format, while in 2013, over 98 % of the data were digitalized (Mayer-Schönberger and Cukier, 2013). This change, alongside with today's massive possibilities for firms to use real-time data concerning, e.g., consumer preferences, shopping behavior, location and health, to guide their actions has resulted in dramatic changes in firm strategies and how markets function. We analyze the developing of the markets for data in the following four domains with an exceptionally large potential to generate value from data: financial services/fintech, health, location-based services and artificial intelligence (i.e., AI) and data analytics. Our study investigates how the technology giants have paved their way to the emerging data-intensive markets. We shed light on the innovation patterns in data-intensive markets, and further explore the data giants' entry or expansion of activities to the new market areas by the acquisition of companies.

The financial services domain is estimated to have the largest digital datasets stored per firm, on average (see, e.g., McKinsey Global Institute, 2011). *Health care* forms another domain holding massive personal data stores. In the health sector, image data (e.g. X-ray and CT scans) dominate stored data volume-wise. Our analysis focuses, instead of the whole health sector, on the wearables domain comprising tracking applications and other technological solutions for individuals to monitor, collect and analyze their own personal fitness and health data. Our third domain of interest concerns markets for *personal location data,* the amount of which has increased rapidly alongside with the diffusion of smartphones. Individual location data generates value for companies in various sectors, e.g., via location-based advertising and service provision. Our fourth domain, *artificial intelligence* (AI) techniques or solutions, particularly machine learning, and data analytics provides ways to extract more value from personal data for firms in all sectors of the economy (e.g., via the analysis of consumers' online behavior).

We analyze the roles of the six U.S.-based technology giants, i.e. Google, Apple, Facebook, Amazon and Microsoft and IBM (i.e., GAFAMI), shaping the markets for data. Google (owned by Alphabet Inc. since Oct. 2015) offers Internet products (e.g., search, ads, maps) and provides digital content, applications and cloud offerings and hardware products. Apple focuses on consumer electronics, personal comput-

ers and portable devices in addition to selling related software, services, accessories, networking solutions, and third-party digital content and applications. Facebook is the major provider of online social network services and messaging. Amazon is the dominant e-commerce company that also manufactures and sells electronic devices. Microsoft develops, licenses, and supports various software products, applications, services and devices, and further offers could-based solutions as well as devices/hardware it designs and manufactures. IBM has five major segments: cognitive solutions, global business services, technology services and cloud platforms, and systems and global financing. The data giants' further extract value from personal data via on-line advertisement in different platforms such as search engines (e.g. Google), application stores (e.g. Apple's App Store), online market platforms (e.g., Amazon.com), personal social media platforms (e.g., Facebook), professional social media platforms (e.g. Microsoft) and health data platforms (e.g., IBM). In 2018, the total turnover of the six companies was about USD 881 billion.

We find that the ideas related to personal data are increasingly protected by patents. The growth in the numbers of personal data related patents was relatively modest from 2005 to the early 2010s, but it has intensified since 2011. Large technology companies' entry to various new market areas is reflected in their investments in R&D and resulting exponential increase in patent applications particularly in the AI analytics domain. Furthermore, we find that the number of AI/data analytics companies acquired by the data giants has escalated during the 2010s. Patent and acquisition data further echo technology giants' expansion of their activities into the financial and personal health services and various other markets outside their own core market areas. Our data illustrate how the divergencies in the data giants' innovation activities and strategic acquisitions have led them to each conquer their specific areas of dominance in the global markets for data.

The rest of the paper is organized as follows. Section 2 introduces the selected technology/market domains and the data. Section 3 first discusses the insights that patents provide to the emerging markets and then gives an aggregate picture of the evolvement of patenting of personal data related innovation over the years. Section 4 sheds light on the role of data giants in shaping the markets for data via their innovation activities and acquisitions of companies. Section 5 portrays some notable product launches of GAFAMI to illuminate how the described developments have materialized in the consumer markets. Section 6 concludes.

2. Data

2.1 Selected domains

This section gives a brief description of the four data-intensive domains (see also Koski and Luukkonen, 2018). The health sector manages highly personal and often sensitive customer data. We restricted our analysis concerning health domain to wearable health and fitness technologies as they represent emerging technologies that collect, often real time, large quantities of personal data. Indeed, the wearable device industry continues its rapid growth and its market size is expected to exceed $ 50 billion by 2022[1]. Currently, three main types of wearable devices and applications collect personal health data: i) those monitoring physiological attributes associated with certain diseases (e.g., diabetes), ii) those tracking fitness activities in real time and iii) personal assistants tracking certain variables of interest such as calories consumed (Olshanksy et al., 2016).

Location-based services (or LBS) exploit information using the location of mobile devices. Markets for LBS such as geotargeted advertisements and offers (e.g., for local restaurants and shops) have emerged alongside the widespread adoption of smartphones. Location-based services can be defined as (mobile) computer applications "that deliver information tailored to the location and context of the device and the user" (Huang et al., 2018). Although the first location-based services appeared in the early 1990s, the main impetus for the field took place in the early 2000s when President Bill Clinton extended Global Positioning System (GPS) from the US military purposes to the larger public audience. Importantly at the same time smart phones with GSP were introduced. During the last two decades there have been many changes in the field. Among the many advances are an expansion from outdoors to indoors positioning, rapid advances in the enabling technologies, e.g. smartwatches and augmented reality; more diverse applications, e.g. in healthcare, transportation, and gaming; many new opportunities in traffic management and urban planning, e.g. in developing smart city solutions; and most importantly using GPS in daily living has become normalized among a large audience (Huang et al., 2018).
The finance sector manages and analyzes vast quantities of customer data such as financial records and credit card information. Fintech is often seen as a uniquely recent marriage of information tech-

[1] See, e.g., http://www.marketsandmarkets.com/PressReleases/wearable-electronics.asp, accessed April 7, 2017; https://www.businessinsider.com/apples-wearables-segment-now-almost-as-big-as-mac-business-2019-10?r=US&IR=T, accessed November 4, 2019.

nology and the finance sector, but it is not an inherently novel development for the financial services industry:[2] The term fintech originated already in the early 1990s; Fintech was the name of Citycorp's (predecessor of Citygroup) project initiated to facilitate technological cooperation efforts (Arner et.al, 2015). Today, some examples of innovation in emerging financial technologies or fintech are, e.g., cryptography and blockchain, new digital advisory and trading systems, chatbots, smart (crypto-)wallets, peer-to-peer lending and equity crowdfunding (Philippon, 2016). With data economy the market for financial services expands and develops to new directions. Globally, mobile banking or increasing use of cryptocurrencies, for instance, are transforming low-income people's access to financial services (Christi and Barberis, 2016). The proliferation of digital financial channels is not only raising customer expectations but also significantly changing the ways customers perceive and use banking channels and products.

Artificial intelligence or machine learning can be regarded as general purpose technology (Trajtenberg, 2019) that relates closely to all the above discussed technology domains. In digital health, AI algorithms can be exploited, for instance, to detect abnormalities learned from the data that wearables transmit from the bodies of their users. The AI system may then alert the user and/or medical personnel in real time (e.g., changes in blood sugar levels for those living with diabetes), improving the efficiency of the treatment of various diseases and promoting the prevention of life-threatening conditions such as heart attacks. For location-based services, AI algorithms are used to analyze user location data and often combined with other user-specific data. Such analyzed data may, for example, offer predictions about users' upcoming locations or guidance on content, advertisements and promotions to target to users. AI has changed financial services provision in many ways, e.g., via improving service security, fraud detection and providing new tools for forecasting and financial advising.

2.2 Patent and acquisition data

We use data on the filed patent applications from 2005 to 2016 in the selected four domains with the US Patent and Trademark Office (USPTO). We focus on the US markets as often personal data related

[2] Already the introduction of the telegraph with its first commercial use in 1838 provided the fundamental infrastructure for the first major period of financial globalization in the late 19th century (Arner et al., 2015).

innovation or innovation in data-intensive sectors comprise software that is patentable in the USPTO, but not in the European Patent Office. Also, the United States forms a single largest software market in the world. We collected data from the patenting activities in the IPC classes applicable to financial services, artificial intelligence and location-based services, all falling to the subcategories of the IPC classes G06 and H04 (see Annex 1 for a detailed description of the IPC classes of these domains). Technologies measuring, collecting and transmitting personal health, wellness and fitness data are patented under IPC category A61B5/00 (i.e., Detecting, measuring or recording for diagnostic purposes; Identification of persons). We inspect both the overall patenting in the selected data-intensive domains, and further take a closer look at new technologies that are described in the filed patent application to be targeted for personal data related purposes. We distinguished such patents by using a search criterion stating that the term "personal data" or "personal information" had to appear in the title, abstract or description of the filed patent application.

The data concerning acquisitions are extracted from Crunchbase, which is a proprietary platform for finding business information about both private and public companies. We collected data on the all acquisitions of the sample large technology companies for 2005-2017. The dataset further comprises the major product categories of both the acquirer and the acquired companies. We inspect the data giants' buyouts in the following four data-intensive market categories: financial services, health care, navigation and mapping and artificial intelligence/data analytics. Location-based services are imperfectly captured here with a rather narrow "navigation and mapping" category that does not include various important location-based service categories such as location-based advertising. Furthermore, we analyze data giants' acquisitions in advertising, which is one major market area in which data giants extract revenues from personal data[3].

3. Personal data related innovation and markets for technologies

3.1 Why patents provide insights to the emerging markets?

A patent provides its holder a right to exclude others from making, using or selling the invention for a limited time, typically 20 years. Firms do not use patents only as a means of protecting innovative ideas

[3] The patent data do not allow us to distinguish patentable ideas related to (digital) advertising as they are not detectable easily by mere IPC classes and further relate to various more generally applicable IPC classes (e.g., artificial intelligence, software).

involved in their new products that they launch on the market and generate revenue from. In fact, various studies suggest that only a relatively small percentage of patented inventions yield substantial economic value for their owners (Harhoff et al., 1999). Patents are also used for various other commercial purposes such as market signals to attract investments, as a source of licensing revenues or as the means to obtain access to complementary technologies via cross-licensing. A firm's patenting in a certain technological field signals the firms' expectations on the future importance of the markets for these technologies. Patents may be used for supporting the creation of new markets but also as options, i.e. to manage an uncertain future and to ensure that the firm will have the freedom to commercially exploit its ideas in new, expanding market areas. The complexity and cumulative nature of innovation in ICT markets stresses the importance of a firm's pre-emptive patenting strategy, and it may lead to the patent rivalry and building up of patent portfolios in technological fields in which firms trace future potential.

In markets in which an invention or new commercial application requires a firm to bundle prior generations of related, complementary inventions, patent portfolios may form a major asset for companies. In case of cumulative innovation, patents are not only providing their holder with a monopoly right for a single invention but also a power to potentially block follow-on innovation. Galasso and Schankerman (2015) find empirical evidence that this, indeed, has taken place in the technology fields of computers, electronics and medical instruments. This has also resulted in global patent wars; the leading technology companies have spent huge resources on patent infringement lawsuits.[4] Patents are thus also applied for defending a firm against future patent infringement lawsuits or for providing the firm means to countersue its competitors in case of infringement. Indeed, Koski and Luukkonen (2017) find that large technology companies relatively rarely initiate aggressive patent infringement battles but rather tend to act as defendants in patent infringement lawsuits. Large technology companies have also attempted to establish goodwill, soften IPR competition and to signal that a firm is using patents rather as a defensive IPR strategy than aggressive means of competition.[5]

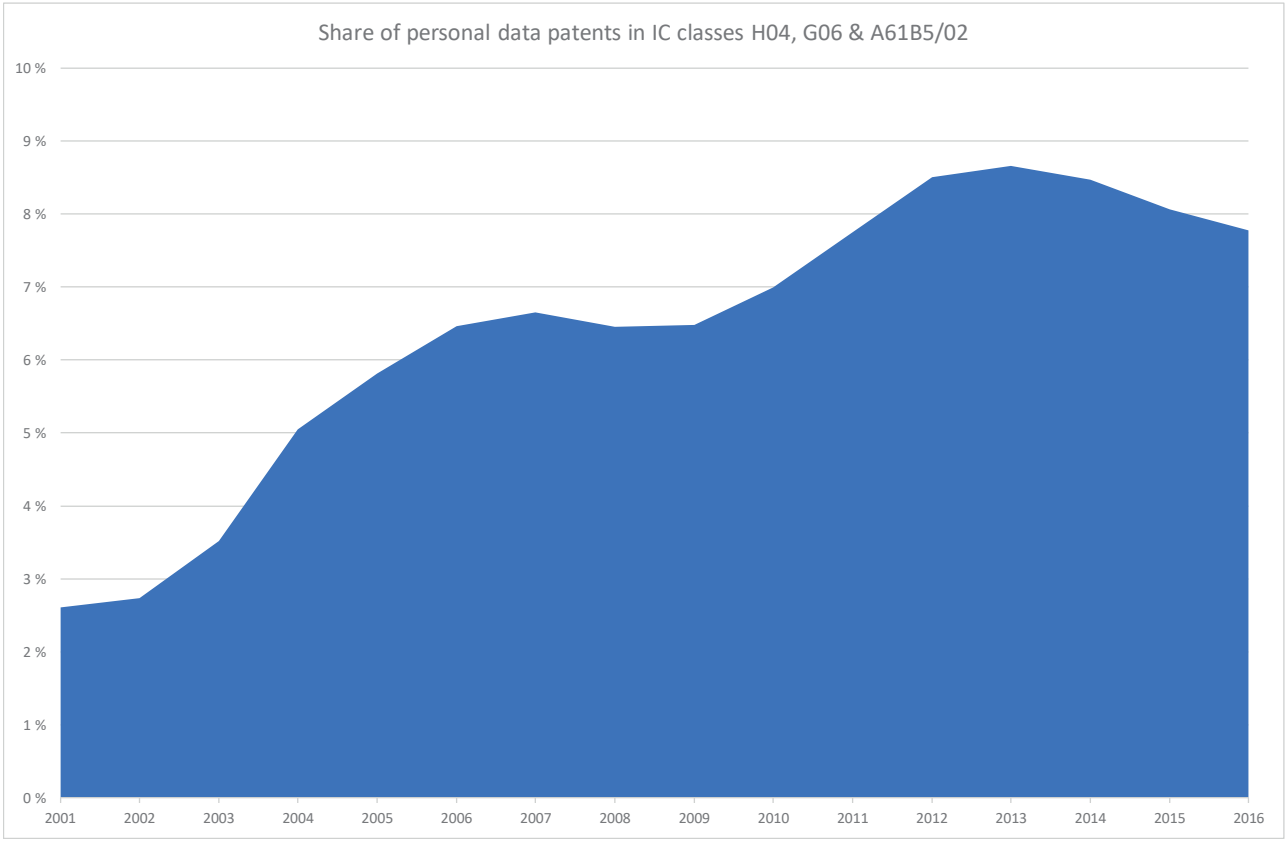
[4] An example of this is a long patent infringement battle fought between Apple and Samsung (see, e.g., http://fortune.com/2016/03/21/apple-samsung-supreme-court/).

[5] One example of this is Google's Open Patent Non-Assertion Pledge announced in 2013. Google has contributed 113 of its US patents (in total 245 global patents) to the pledge and promised it does not sue developers, distributors or users of these patents unless it is first attacked. Source: https://www.google.com/patents/opnpledge/pledge/; Accessed May 20, 2016.

3.2 Ideas related to personal data are increasingly protected by patents

In, 2016, the number of personal data related patent applications filed in the IPC classes G06, H04[6] and A61B6/00 was over 10,000. This was nearly double the respective patent applications filed with the USPTO in 2005, and an almost six-fold increase compared to the year 2001. These numbers illustrate how the technology companies' expectations on the growth and potential of the markets for technologies related to personal data use have increased during the past two decades. However, to acknowledge the possibility that the absolute numbers rather reflect a general increasing trend in ICT patenting, we draw a figure of the share of personal data patents of all patents published in IPC classes G06, H04 and A61B6/00 from 2001 to 2016.

Figure 1. Share of applied "personal data" related patent applications of all ICT patent applications filed with the USPTO



[6] Patented ideas related to collecting, transmitting, exchanging and analyzing personal data largely appear under IPC classes G06 (i.e., computing, calculating and counting devices) and H04 (e.g., electrical communication systems). See Koski and Luukkonen (2018) for a more detailed discussion.

Figure 1 reveals that not only the absolute numbers of personal data related patent applications filed annually with the USPTO have multiplied, but also the share of personal data patent applications of all filed ICT patent applications has surged from 2001 to the mid-2010s. During the early 2000s, personal data related patent applications filed with the USPTO covered only about 2-3 percent of all filed ICT patent applications, while in the 2010s the share had grown to 8-9 percent.

Figure 2. Total number of personal data related patent applications in the selected domains filed with the USPTO

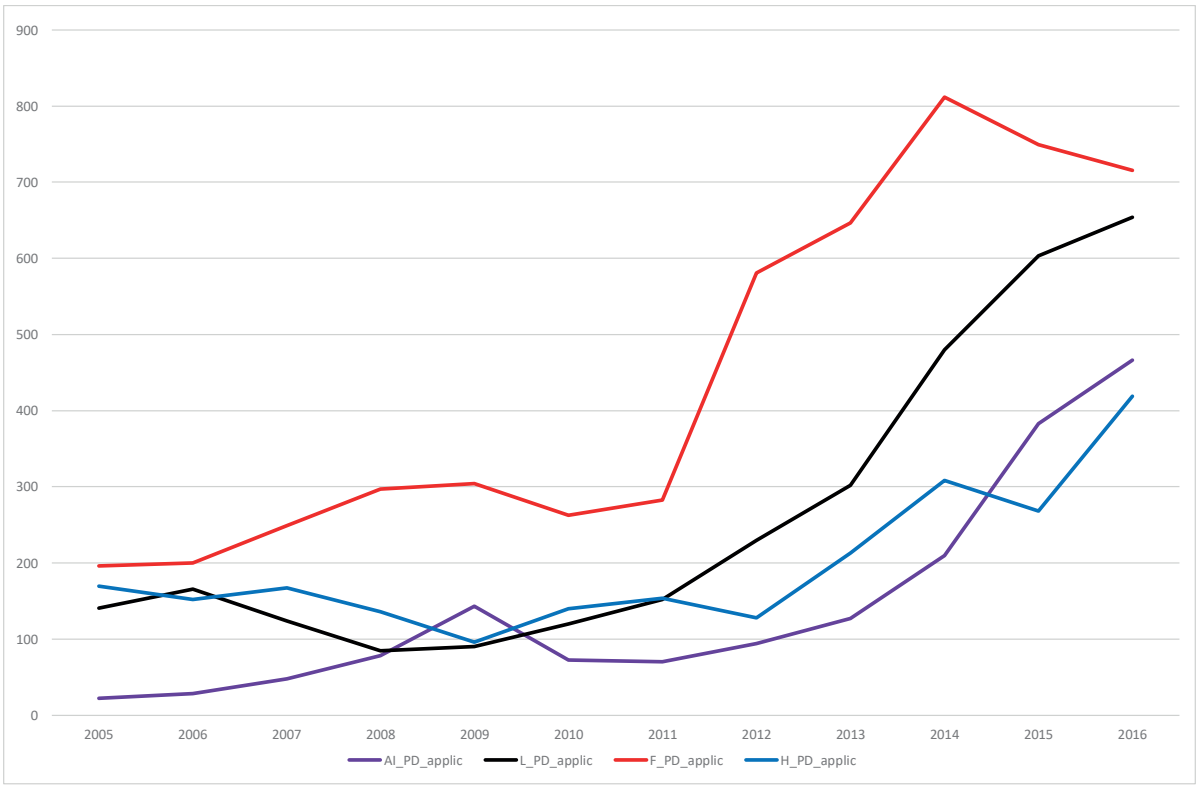


Figure 2 illustrates how the total number of personal data related patent applications filed in the four selected data-intensive domains has evolved from 2005 to 2016. There was a relatively modest growth in the number of patent applications in all four domains from 2005 to 2011 besides AI, in which the annual number of filed patent applications with the USPTO more than tripled from about slightly over 20 to about 70. The finance domain is clearly the most prominent one in terms of personal data pa-

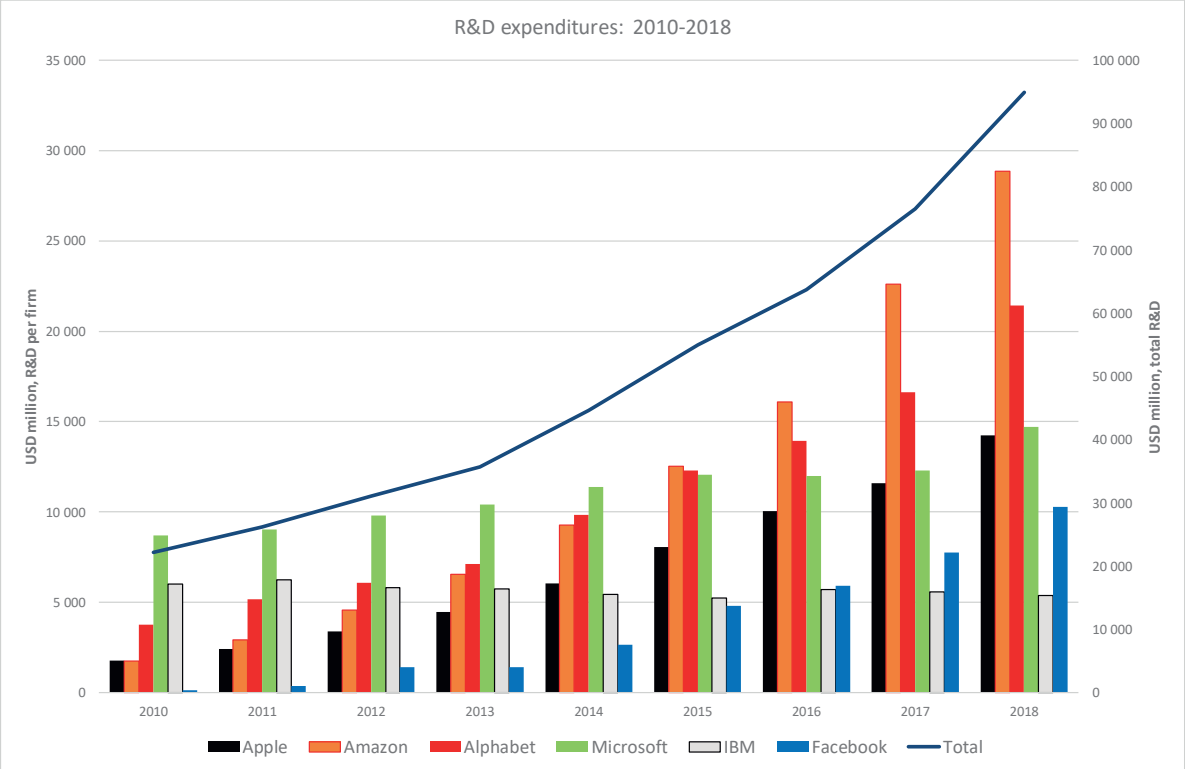
tenting among the sample ones. Since 2011, we observe exponential growth patterns in the filed personal data patent applications in all four domains. Relatively, the greatest growth has taken place in the technology domain concerning artificial intelligence.

4. How do technology giants shape the market?

4.1 Innovation

Our aim here is to shed light on the technology development paths of the large technology companies exploiting personal data that have led them to the global dominance. Figure 3 gives a picture of the volume of R&D investments undertaken by GAFAMI. Overall, the R&D investments of data giants have grown more than ten-fold from 2010 to 2018 when the total R&D investments of companies amounted

Figure 3. R&D expenditures of data giants



to over USD 95 billion. Among the six companies, Amazon and Google witnessed the most notable increase in their R&D spending during 2010-2018, respectively, from USD 1.7 billion and USD 3.7 billion to close to USD 29 billion and USD 21 billion.

Figure 4. Patent applications of data giants filed with the USPTO, 2005-2016

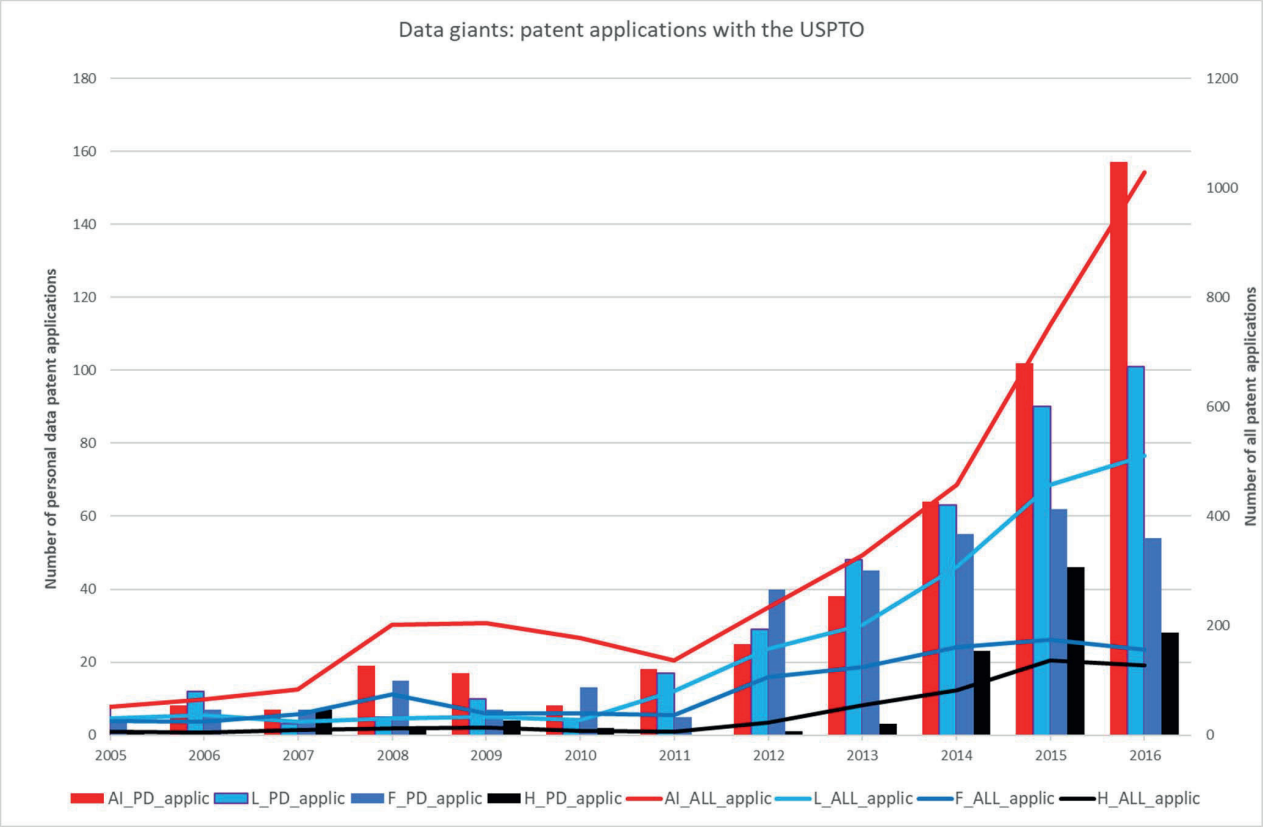


Figure 4 presents GAFAMI's annual number of patent applications filed at the USPTO in the selected domains. The bars plotted on the graph describe the number of *personal data* related patent applications (measured by the primary vertical axis), while the lines describe *all* patent applications of GAFAMI in the selected domains (measured by the secondary vertical axis). First, if we compare the data giants' patenting activities to those of the total filed patent applications in the domains (Figure 2), we observe very similar overall patterns in the growth of filed patent applications. However, among the data giants, the most prominent domain, and that with the most intense growth since 2011, concerned artificial intelligence – not finance, as in the total number of filed patent applications. GAFAMI's over 3,700

AI patent applications during the sample years comprise a wide variety of technologies aimed, e.g., at intelligent security management, medical risk factors evaluation and inventory management. The personal data related AI patent applications of GAFAMI often describe technologies designed for predicting users' online behavior and providing personalized services. Such patent applications include, e.g., those titled with "Predicting user navigation events", "Inferring user mood based on user and group characteristic data"and "Systems and methods for providing personalized content".

The data giants filed close to 1,900 patent applications in the location-based domain during 2005-2016. These include, e.g., technologies aimed at predicting the persons' movements, i.e. where the person will be and at what time (e.g., Facebook's patent application titled "Predicting locations and movements of users based on historical locations for users of an online system"). The following examples of the titles of patent applications in the location-based services domain describe some purposes the patented ideas of GAFAMI are designed for: "Location based recommendations" (by Amazon), "Easy location sharing" and "Identifying and locating users on a mobile network" (by Apple), "Travel recommendations on online social networks" (by Facebook), "Presenting information for a current location or time" and "Inferring user interests" (by Google).

There were clearly fewer patent applications filed in the wearable health and fitness technology and finance domains than in the AI and location-based services domains. However, in 2016, the number of the data giants' patent applications in the health and fitness (finance) domain was over nine-fold (four-fold) compared to that of the year 2010. It is evident also from the product launches of the data giants that massive interest in banking and well-being unites them.[7] Particularly IBM and Apple have heavily invested in the digital heath technologies. In 2015, IBM launched Watson Health Cloud that enabled the combination, de-identification and sharing of health data for use by, e.g., doctors, insurers and researchers.[8] In the following year, IBM told that it would invest USD 150 million[9] in Watson Health's first European Center of Excellence in Milan. Large technology companies' gradual entry to the digital health technology markets was also reflected in their hiring decisions. In the 2010s, Google, Apple, Microsoft and IBM began to hire leading biomedical researchers (Wilbanks and Topol, 2016).

[7] This is well reflected in huge and risky investments made in these two areas (http://www.economist.com/news/business/21636754-new-tech-bubble-seems-be-inflating-when-it-pops-it-should-cause-less-damage?fsrc=scn%2Ffb%2Fwl%2Fpe%2Ffrothycom)

[8] See https://www-03.ibm.com/press/us/en/pressrelease/49436.wss. Accessed Oct 10, 2017.

[9] See, e.g., https://www-03.ibm.com/press/us/en/pressrelease/49436.wss. Accessed Oct 10, 2017.

The number of filed patent applications from 2005 to 2016 mirrors the global data giants' expansion of their operations into the financial services markets. The data giants have developed various fintech solutions such as mobile payment applications. In 2014, Apple filed a patent application titled "Person-to-person payments using mobile devices" and further launched its Apple Pay application that was aimed at transforming the mobile payment industry. Also, Google and Amazon have filed various patent applications related to mobile payment systems and launched their own digital payment services. Another example of the finance-sector related patented ideas is Facebook's patented technology that can be applied to credit grading that bases the acceptance of individuals' loan applications on the credit ratings of borrowers' social networks.[10] Technological advancements in the finance domain were accompanied by the data giants' entry into certain segments of the financial markets. For instance, Amazon began to provide loans to sellers through its online marketplace, and Facebook announced plans to go into new businesses such as banking and payment systems.[11] More recently, Facebook has released its plans to enter the digital currency markets with the launch of its own cryptocurrency, Libra[12].

Figure 5 shows how the share of personal data related patent application filed by the data giants in four domains have evolved between the time periods 2005-2010 and 2011-2016. The six large technology companies covered about the quarter of the all personal data related patents filed in the artificial intelligence domain with the USPTO during 2011-2016, while their corresponding percentage was 16 during 2005-2010. The data giants' patents filed in the location-based service domain has also grown notably, from less than three percent in 2005-2010 to close to ten percent during 2011-2016. Patent shares in the health and finance domain have remained more stable.

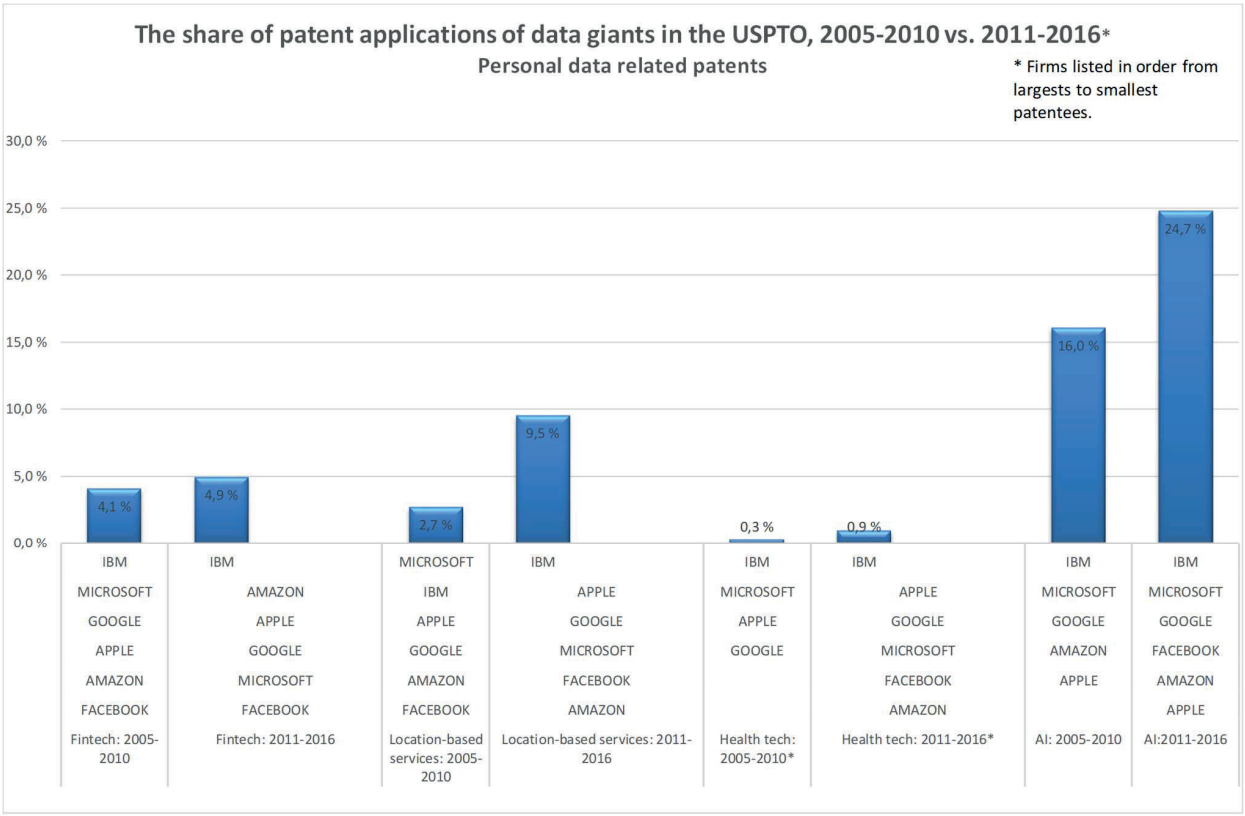
[10]See description of the patent: http://patft.uspto.gov/netacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&u=%2Fnetahtml%2FPTO%2Fsearch-adv.htm&r=4&p=1&f=G&l=50&d=PTXT&S1=((%22facebook%22.ASNM.)+AND+%40PD%3E%3D20150804%3C%3D20151231)&OS=AN/%22facebook%22+AND+ISD/8/4/2015-%3E12/31/2015&RS=(AN/%22facebook%22+AND+ISD/20150804-%3E20151231).

[11] See, e.g., https://time.com/3542410/facebook-money/. Accessed Nov 4, 2019.

[12] Facebook has been active in proposing a new Libra currency based on blockchain technology. Libra Association consisting of many leading firms such as Spotify, Uber and Vodafone manifests optimistic view according to which in future even the unbanked could participate in the financial realm: "…1.7 billion adults globally remain outside of the financial system with no access to a traditional bank, even though one billion have a mobile phone and nearly half a billion have internet access. *An Introduction to Libra. White paper* from the Libra Association Members (July 23rd, 1/12)

During 2011-2016, over 50 percent of the technology giants' 300 personal data related wearable health and fitness tech patent applications were filed by IBM and close to one fifth by Apple. Google's move into health and wellbeing was reflected by its 50 personal data related patent applications filed with USPTO in the domain during 2011-2016 as opposed to merely one patent application between 2005-2010. Amazon and Facebook both filed three personal data related health and fitness tech patent applications with the USPTO during 2011-2016, while Microsoft's total count was 30. Microsoft's move into health tech can also be detected by its launch of a new service platform, Microsoft Health, and the wearable health/fitness monitoring device, Microsoft Band, in 2014.[13]

Figure 5. Share of personal data related patent applications filed by data giants with the USPTO



[13] Microsoft health was described as "…a new service that helps you live healthier by providing actionable insights based on data gathered from the fitness devices and apps that you use every day. It's designed to work for you, no matter what phone you have, device you wear, or services you use. Microsoft Health makes tracking personal fitness easier, more insightful, and more holistic. Source: http://www.microsoftnow.com/2014/10/microsofts-fitness-band-and-microsoft-health.html.

4.2 Acquisitions in data-intensive domains

The six data giants acquired in total about 700 companies during 2005-2017. In other words, data giants' path to the global dominance and expansion to the new markets has involved a rather massive transfer or internalization of external knowledge of other companies. Figure 6 indicates that there was a slightly increasing trend in the total number of annual acquisitions the data giants undertook between 2005 and 2017. The number of buyouts in the broad domain of location-based services is limited to the firms active in the markets for "navigation and mapping" where we observed only six acquisitions during the sample time. Instead, firms expertizing in artificial intelligence/data analytics and advertising were more often the targets of GAFAMI: from 2005 to 2017 the data giants bought over 140 AI/data analytics firms and close to 70 advertising companies. Over 60 percent of the acquired companies active in artificial intelligence category were bought between 2013-2017.

Figure 6. Number of data giants' acquisitions: 2005-2017

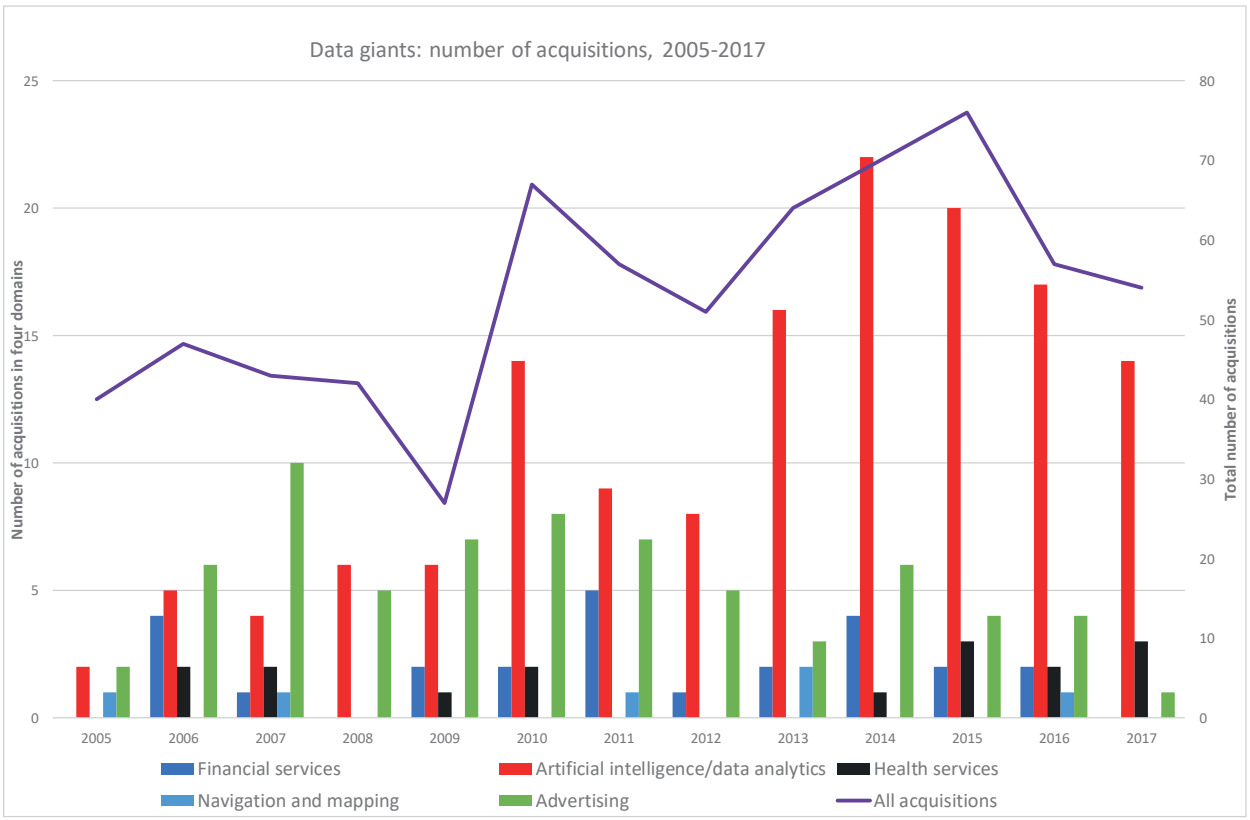


Figure 7. Percentage shares of acquired companies among data giants, 2005-2017

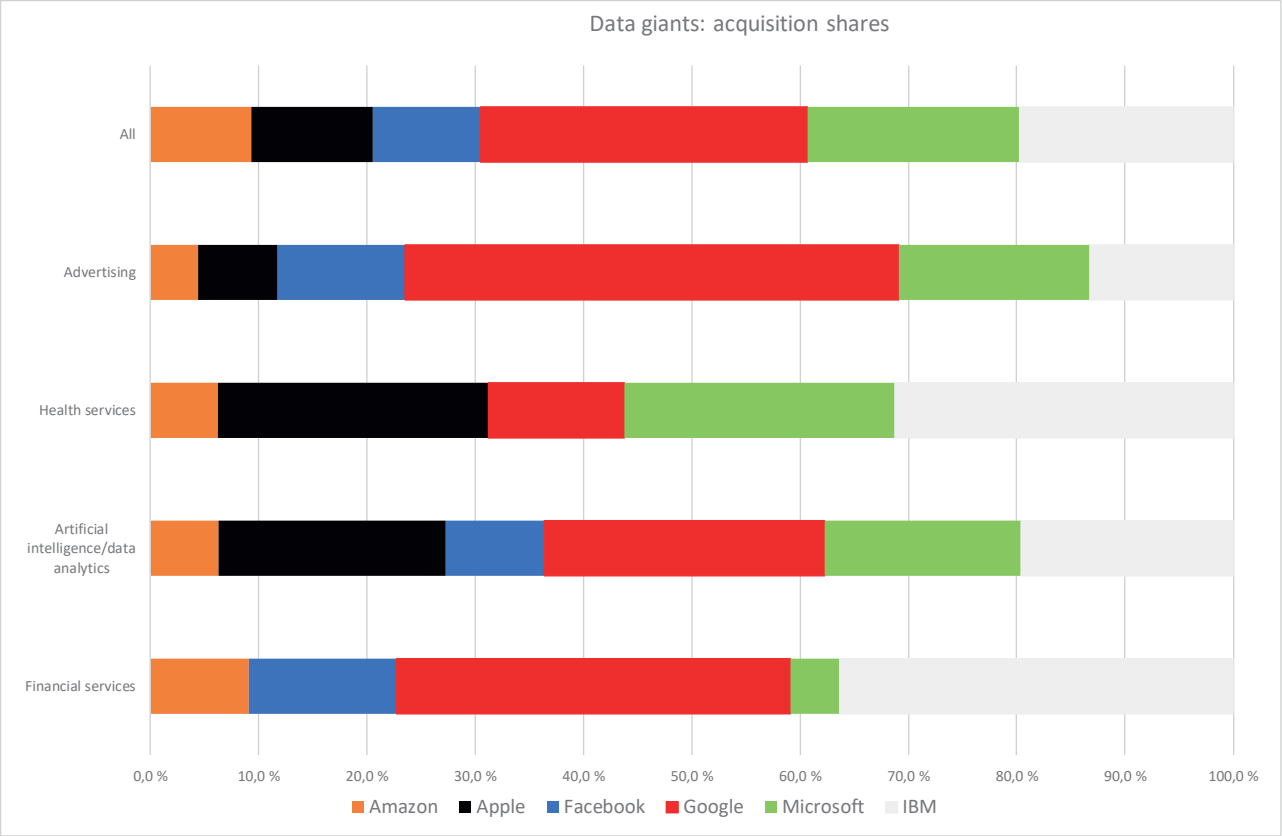


Figure 7 compares the relative shares of GAFAMI of all acquisitions among the sample technology giants from 2005 to 2017 in different domains (excluding location-based services as it had only six acquisitions). Overall, Google was the most aggressive acquirer among the technology giants (i.e., it acquired 30 % of all firms bought by the data giants), while Microsoft and IBM both accounted for about a fifth of all acquisitions. Google bought over 45 percent of advertising companies among GAFAMI. IBM and Apple acquired both a quarter of the health service companies bought by the data giants. One of the largest health sector deals was IBM's acquisitions of Truven Health Analytics[14] for USD 2.6 billion in 2016 and its buyout of Merge Healthcare[15] for USD 1 billion in 2015. In 2015, IBM further acquired Phytel (for an undisclosed price) that provided medical care solutions and developed prescription, lab

[14] Truven Health Analytics delivered information, analytic tools, benchmarks, and services to the healthcare industry.
[15] Merge Healthcare developed enterprise image sharing applications for patients and physicians.

and clinical data for health and disease management. Facebook was the only company among GAFAMI that did not buy any health service firms.

IBM and Google accounted for a combined total of - divided equally between the two firms – almost 73 percent of the acquisitions of the firms active in financial services, while Apple had no buyouts from the financial sector. The acquisitions of artificial intelligence companies were divided more equally between Google (i.e. 26 %), Apple (i.e., 21 %), IBM (i.e. 20 %), and Microsoft (i.e. 18 %). Amazon and Facebook had relatively smaller shares of the buyouts of the AI companies, respectively 6 % and 9 %.

Our data reveal that the data giants acquired close to 70 advertising companies from 2005 to 2017. In 2009, location-based mobile advertising was still in rather early stages of development. The data giants' intentions to capitalize on personal data use for location-based services and advertising were, however, already reflected in their actions. In November 2009, Google acquired mobile advertisement network AdMod for USD 750 M[16]. During the 2010s, the markets for location-based mobile advertisements grew substantially reaching over $12 billion in 2016. These markets were expected to reach $32 billion by 2021, covering 45 percent of total mobile advertising revenues.[17] Over 45 percent of the acquisitions of advertising companies were undertaken by Google, which in 2018 was the single largest digital advertiser in the United States with a close to 40 percent digital ad revenue share.[18]

4.3 Data giants' core vs. non-core acquisitions monetizing data

We further inspected the market entry and expansion patterns of GAFAMI by exploring their acquisitions in their own core market areas and outside the core markets. We define an acquisition as taking place in the firm's core market area if the acquired firm is active in one or more of the same core business categories as the acquirer. In other words, the overlapping core business categories of acquirer and acquired company means that the acquirer and acquired firm are likely providing substitutes and are thus competitors at least in one product market area. Figure 8 shows that IBM and Google have acquired annually the largest number of companies in their core market areas during the sample years.

[16] See, e.g., http://www.businessinsider.com/google-to-acquire-mobile-ad-network-admob-for-750-million-in-stock-2009-11?r=US&IR=T&IR=T Accessed February 12, 2018.
[17] See, e.g., https://bluedotinnovation.com/location-based-advertising-local-mobile-ads.html Accessed February 12, 2018.
[18] Information source: www.statista.com with the original data from EMarketer, MediaPost.

Figure 8. Number of acquired companies in the core market areas of data giants, 2005-2017

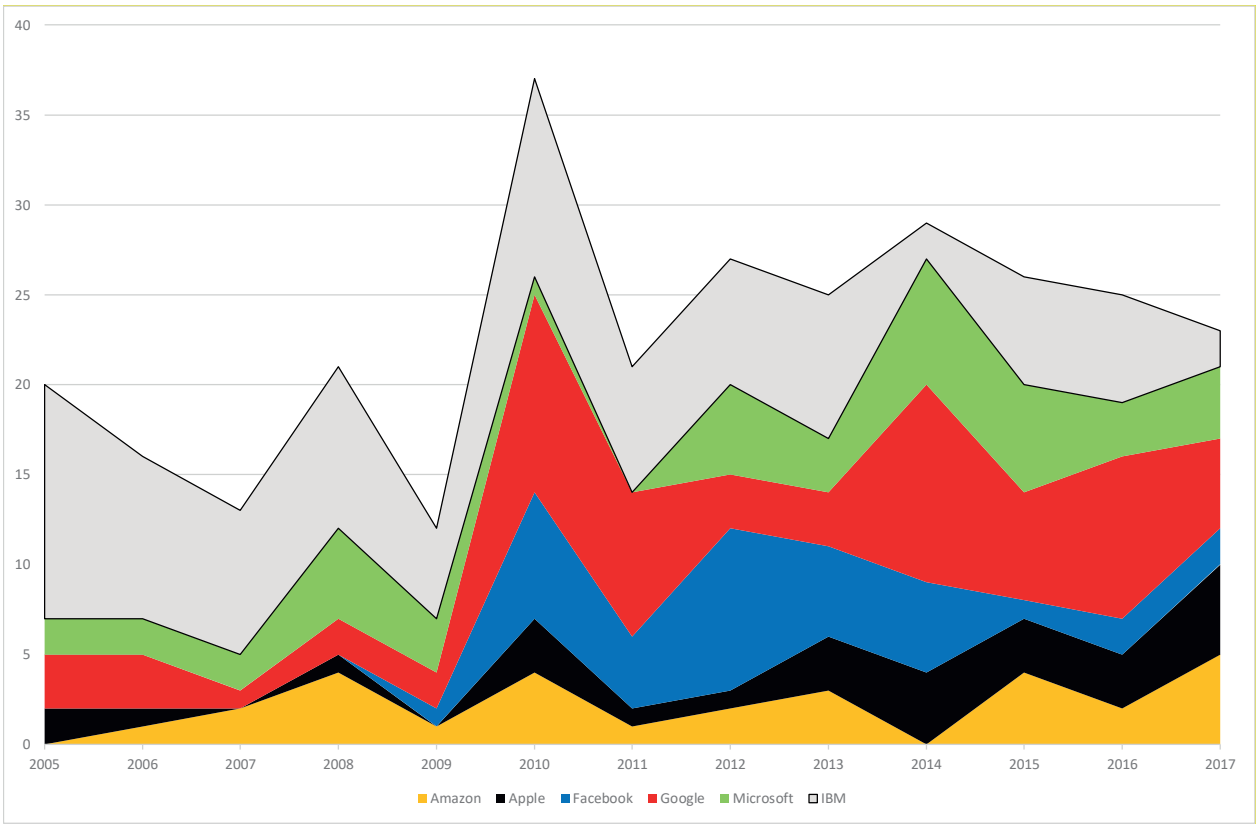
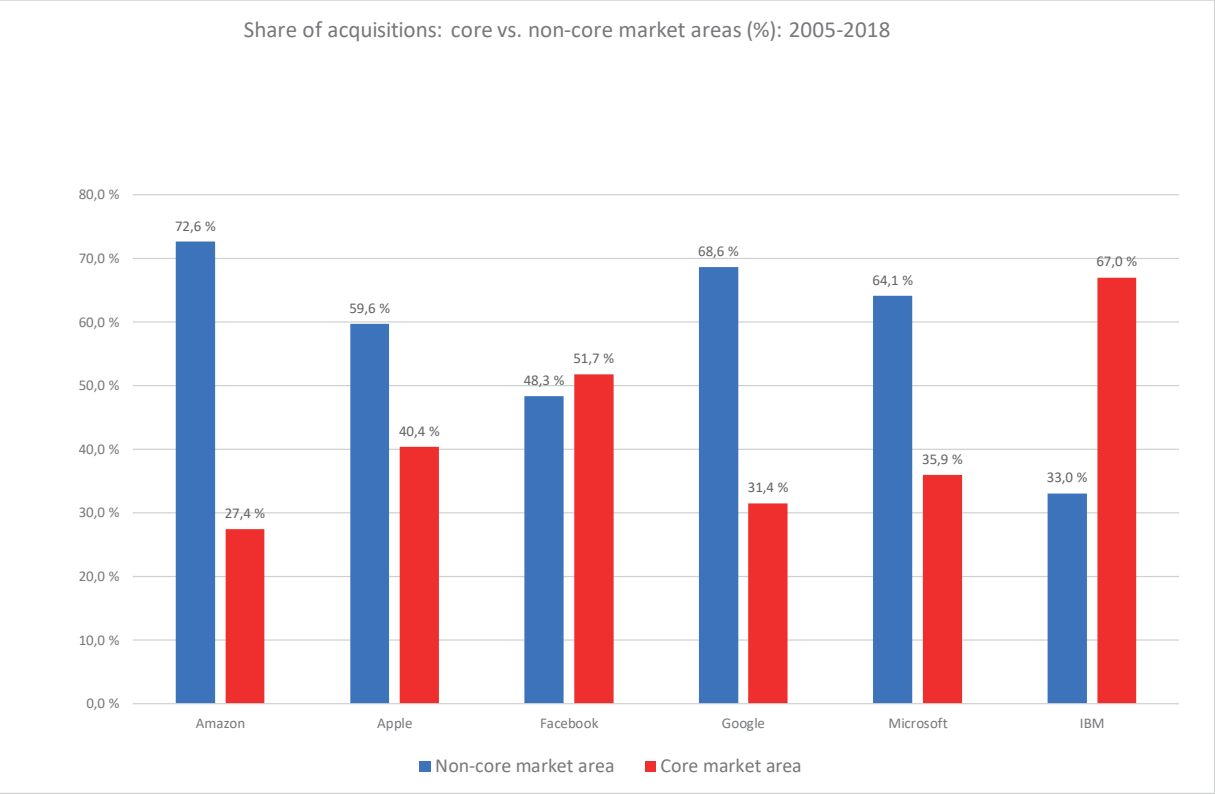


Figure 9 presents the share of the data giants' acquisitions that concerned firms that were functioning in the acquirer's core market areas vs. outside of them. Overall, the data giants - besides IBM - have frequently targeted companies active in the markets outside their core business. Over two thirds of IBM's acquisitions concerned its competitors in the firm's existing markets. The data giants' entry and expansion to the new markets is particularly visible in the acquisition history of Google. Google made over 200 acquisitions from 2005 to 2017. About 69 percent of the acquired companies by Google were active outside their own core market areas. One of Google's most notable and successful moves was its acquisition of video sharing platform YouTube (that has now over 2 billion users) for USD 1.6 billion in 2006.

Microsoft's largest acquisition was its over USD 26 billion deal with LinkedIn in 2016. For Microsoft, this meant a step into the social media markets, making it the world's leader in both organizational and personal professional data and further enlarged its digital advertising business. Amazon's over USD 13

billion acquisition of Whole Foods in 2017 was a flamboyant sign of its expansion from online retail to the brick-and-mortar business. After this acquisition, in 2019, Amazon was announced not only to be the world's largest e-commerce company but also the world's largest brick-and-mortar firm.[19] Data-wise, Amazon's expansion to the brick-and-mortar business meant that it now had, besides access to massive online behavior data, also access to data on physical shopping behavior (e.g., stores visited, products seen and bought or not bought).

Figure 9. Data giants' acquisitions: core vs. non-core market areas, 2005-2018



The saturation of the markets for smart phones has driven Apple to explore and pursue new avenues in the data economy. In 2014, Apple's acquired Beats Electronics, which provided streaming music subscription services in addition to making headphones and speakers, for USD 3 billion. In January 2015, Apple subsequently bought out Semetric offering insight, analytics, discovery, recommendations, and

[19] See, e.g., https://www.forbes.com/sites/laurendebter/2019/05/15/worlds-largest-retailers-2019-amazon-walmart-alibaba/#3fefc294171c.

targeting services in entertainment and tracking billions of points of personal data concerning, e.g. the usage of music, TV, movies and book. These two buyouts, supported by some other ones, provided Apple not only with a stronger competitive position in the markets for music-streaming but also access to new tools and data to analyze music consumption patterns and behavior. Furthermore, given that the Semetric platform covers also a great variety of other entertainment use, this acquisition further opened future possibilities for Apple to potentially conquer data-driven markets beyond music.

About 58 percent of Facebook's buyouts concerned firms that were active in its core market areas. Some prominent deals Facebook has made in its core market areas comprise its buyout of messaging service provider, WhatsApp, in 2014 for USD 19 billion and the social media platform for photo sharing, Instagram, in 2012 for about USD 1 billion. Both WhatsApp and Instagram were companies with the fast-growing user bases further strengthening Facebook's core business of monetizing personal data via advertising.

5. Data giants in consumer markets

During the last twenty years, and most notably during the past decade, data giants have diversified their operations into new market areas. New services, new generation software and data generating devices have been launched at an accelerating pace. Smart watches, wearables, smart speakers (echo), smart televisions, even thermostats, have had an important role in integrating consumers into the data economy. Apple's focus was originally computer devices. It exported the logic of the computer to new domains by creating radically new product classes, such as iPod (2001), iPhone (2007) and iPad (2010). With Apple Watch launched in 2014 the company joined the movement where the human body is invaded in terms of wellness and health. Amazon has its origin in physical objects and book online retailing. It launched its reading device Kindle in 2007 and Fire phone and Fire TV set-top box in 2014. It has diversified from books to fresh food (AmazonFresh 2007, AmazonGo 2018) and many other items. With over 647,000 employees in 2018 Amazon has gone a long way from being a bookstore. Today Amazon is a major provider of cloud computing services and devices; content brokering, data analysis and data storage (Amazon web services), are in the core of Amazon's current business.

In a similar vein Google has moved from being a seemingly trivial search engine to multiple directions. WiFi-enabled, programmable, self-learning, sensor-driven the Nest Learning Thermostat was introduced in 2011. Google acquired Nest Labs for USD 3.2 billion in January 2014, and today it is marketed under the brand GoogleNest. The driverless car is one example of Google's attempt at crowdsourced information collection and finally artificial intelligence based on data gathering (Levy, 2011, 385). In October 2019, Google's parent company bought Fitbit tracker company with 28 million active users. Clearly, one of the underlying reasons was access to vast amounts of health/fitness data that the Fitbit buyout provided them[20]. This USD 2.1 billion investment fits well to *Google Health's* profile, in which *Calico* and *Verily* are working at the intersection of technology, data science and healthcare[21]. In a quite similar vein, Apple and IBM have jointly for several years constructed a health ecosystem consisting of devices, e.g. smart watches, cloud services and artificial intelligence[22]. The latest data giant in the health field is Amazon. Amazon Care was launched in September 2019. Currently it is a virtual medical clinic only for Amazon employees[23]. It is estimated that there were 325,000 mobile health apps available in 2017[24]. According to the IMS Institute for Healthcare Informatics there are roughly 170,000 health apps available on the Apple and Android app stores."[25]. No matter how accurate these figures are, the increase since the introduction of smart phones has been tremendous.

Indeed, all the data giants share an interest in personal data and data integrations. The patent and acquisition data furthermore show that all GAFAMI have a strong interest in artificial intelligence. IBM Watson is probably the most famous manifestation of these developments. Today IBM Watson Customer Engagement business unit supports marketing and commerce with AI analytics. *Data for Customer Experience analytics*, *Watson Marketing Insights* and *Watson Real-Time Personalization* use consumer behavior as the input of machine learning and B2B marketing. Google, Microsoft and Amazon in

[20] https://www.techradar.com/news/5-likely-reasons-why-google-just-bought-fitbit
[21] https://health.google
[22] https://www.cnet.com/news/apple-partners-with-ibm-on-new-health-data-analysis/
[23] Quite obviously this a test bed exactly like Amazon Go, a store that uses cameras and sensors to eliminate the need for checkout lines. The two-year test with employees ended 2018 when the concept was launched to the general public. https://www.cnbc.com/2019/09/24/amazon-launches-employee-health-clinic-amazon-care.html
[24] https://research2guidance.com/325000-mobile-health-apps-available-in-2017/; see also Nussbaum, Ryan, et al. "Systematic review of mobile health applications in rehabilitation." Archives of physical medicine and rehabilitation 100.1 (2019): 115-127. (https://www.sciencedirect.com/science/article/pii/S0003999318311754)
[25] Stephen Armstrong pointed out recently in British Medical Journal that most health apps provide little or no protection for the data they collect "despite such data being classified as "sensitive" under data protection law." British Medical Journal (2016);353:i3406;, doi: 10.1136/bmj.i3406

their turn have developed AI applications for the consumer market and for consumer daily use. Most notably, during the last few years each company have introduced their voice-activated digital assistants requiring advanced machine-learning technologies to function. Two-way conversation with a natural language, voice recognition and voice-controlled technology and cloud-based nature of personal digital assistants means that each linguistic nuance gradually accumulates a statistical signature in data giants' database. In other words, assistants' adaptive learning curves converge, "as if thousands of millions of dialect speaker are all passively educating the same virtual toddler in correct pronunciation" (Finn 2017, 63). At the same time assistants encourage people to machine-legible ways of communicating.

The first digital assistant, Apple's Siri, was launched in 2011 at the same time with Apple's new generation of Iphone.[26]. Microsoft's Cortana was first introduced in 2014. The development was started five years earlier and the goal was to make a future operating system for Windows phone and Windows. Currently Cortana is integrated into the Bing search engine and Microsoft ecosystem. Google launched its artificial intelligence-powered digital Google assistant in 2016. Already in 2017, the assistant was installed on more than 400 million devices. In addition to two-way conversations (e.g. with Internet content or home technology), Google has announced that in the future the assistant will be able to use its camera to recognize visual information, purchasing products and sending money[27]. The final goal for Google seems to be anticipating the needs of people even before they recognize what to do next (Finn,2017, 173). 'And the vision is obvious: "Since its earliest days, Brin and Page have been consistent in framing Google as an artificial intelligence company – one that gathers massive amounts of data and processes that information with learning algorithms to create a machinelike intelligence that augments the collective brain of humanity."(Levy, 2011, 385).

Amazon started work on Amazon's Echo assistant (later known as Alexa) in 2011 and it was released in March 2014 mainly for Prime customers. In market terms Amazon has been maybe the most successful company in introducing its assistant Alexa because of its device-specific nature. Originally Amazon Echo

[26] It was a spin-off of a project originally developed by the SRI International Artificial Center. Interestingly, Apple's notion of a digital personal assistant was already a concept video in 1987 several years before the introduction of the Internet or smart phones. It was called The Knowledge Navigator.
[27] https://www.minterest.com/google-products-services-you-probably-dont-know/, accessed Nov 4, 2019.

was supposed to become a smart speaker which is able to control home automation with voice interaction. With Alexa streaming podcasts, providing weather, sport, traffic information and news has become normalized in many American homes. In January 2019 over 100 million Alexa-enabled devices has been sold[28].

Practices from Silicon Valley are also spilling over into traditional business. Hendrikse et al. (2018), for instance, coined the term ”Appleization of finance” (cf. “Googleization”) to characterize the ways financial incumbents are nowadays mimicking the data giants. These developments have primarily actualized in the period following the financial crisis. This fits well to our finding that the finance domain has been the most prominent one in terms of personal data patenting[29]. As Henrikse et al. (2018) note, Apple has not been a typical platform firm because of its seemingly old-fashioned way of keeping both physical products and digital services under its own control. In a closer look, Apple’s data-driven platform logic fits well for established finance exactly because of facilitating status quo, keeping both customers and developers under full control. A large variety of hardware and related software, Apple’s Software Development Kit and an Application Program Interface, are essential building blocks of Appleization. Mechanisms that Apple’s success comes from, e.g. network effect, technological lock-ins (i.e. Apple content can only be played on Apple hardware devices) and high switching costs are lucrative from an established finance point of view.

6. Conclusions

The aim of our analysis was to shed light on the creation and development of markets for personal data. We primarily used the patent and acquisition datasets for this purpose. Our analysis focused on the activities of six technology giants, GAFAMI, in the generation of the markets for personal data. Our analysis draws a picture of large technology companies that have increasingly and heavily invested in research and development and, as a result, generated new patentable ideas relating to personal data particularly in the artificial intelligence domain. During 2011-2016, over 50 percent of about technolo-

[28]Source: https://techcrunch.com/2019/01/04/more-than-100-million-alexa-devices-have-been-sold/, accessed Nov 4, 2019.
[29] So far, it has been mainly the GAFA’s digital payment systems that have disrupted established channels in which banks are the middlemen. It depends on the regulation whether the data giants will have access also to deposits and basic bank accounts in the future.

gy giants' 300 personal data related health tech patent applications were filed by IBM and close to one fifth by Apple. The data giant's entry to financial services markets is reflected, e.g., by their patented ideas related to mobile payment applications and the launch of digital payment platforms and services.

The data giants have complemented their internal R&D efforts to boost their capabilities for advanced data analytics and value creation from data by the large number of acquisitions of companies. Our data show that Google's emergence as the digital advertising giant - extracting value from personal data with over USD 96 billion digital ad revenues in 2018 – has occurred alongside with its acquisitions of potential competitors in the markets for advertising and user generated content. Google was the most aggressive acquirer measured by the total number of acquisitions among the data giants. Overall, the data giants - with the exception of IBM - have heavily relied on the acquisitions of companies active in the markets outside of their core business. Particularly Amazon has built its retail dominance complemented with a variety of different business activities by targeting its acquisitions forcefully to companies in the new market areas: over 70 percent of Amazon's acquisitions from 2005 to 2017 concerned companies that were active outside of its core markets.
IBM has instead focused more on buying firms active in its core business or at least partly in the same markets with it; it was the most hostile acquirer of the competing firms. IBM was, however, also the company that bought the highest number of firms in the two data-intensive sectors: financial services and health. Furthermore, IBM filed clearly the highest number of patent applications both overall and in data-intensive domains, building up its technological capabilities and patents stocks. Though Facebook absorbed about one fifth of the digital advertising revenues in the Unites States in 2018 and most of its revenues were generated by advertisements, it acquired clearly fewer advertising companies than Google, IBM and Microsoft. However, Facebook made several major acquisitions in its core business areas, social media platforms and messaging that guaranteed lock-in effects related to network externalities. Facebook's global dominance in the most popular social media and messaging service platforms provided it with the personal data of a massive user base of over 2 billion individuals from which it could extract value via advertising.

Our investigation reflects the paths the data giants have taken to conquer their share of global markets for data. Amazon has based its data empire on the extraction of data concerning online and physical

shopping behavior. Facebook has access to a massive variety of personal data via its social media and messaging platforms, while Microsoft holds the world's greatest stocks of organizational and personal professional data. Google's gigantic data assets include billions of data points of online search patterns and people's physical whereabouts. IBM's past investments have generated vast health databases for it. Apple's own devices track and monitor fitness and health data, and it has further expanded its activities into the collection and analysis of the usage of music and entertainment as well as. As the consequence of all this, we currently witness the markets for data that covers each segment of our everyday lives – by and large, the digital trail of our every action is recorded and controlled by the six data giants.

References

Arner, D. W., Barberis, J., & Buckley, R. P. (2015). The evolution of Fintech: A new post-crisis paradigm. Georgetown Journal of International Law, 47, 1271.

Christi S. and Barberis J. (2016) The FinTech Book. The financial technology handbook for investors, entrepreneurs and visionaries. Chichester: Wiley.

Finn, E. 2017. *What algorithms want. Imagination in the age of computing.* Cambridge, MA: MIT Press.

Galasso, A. and Schankerman, M. (2015). Patents and Cumulative Innovation: Causal Evidence from the Courts. *Quarterly Journal of Economics* 130, 317-369.

Harhoff, D., Narin, F, and Vopel, K. (1999), Citation Frequency and the Value of Patented Inventions. *Review of Economics and Statistics* 81, 511-15.

Hendrikse, R., Bassens, D., and Van Meeteren, M. (2018). The Appleization of finance: Charting incumbent finance's embrace of FinTech. Finance and Society 4(2): 159-80.

Huang, H., Gartner, G., Krisp, J. M., Raubal, M., & Van de Weghe, N. (2018). Location based services: ongoing evolution and research agenda. *Journal of Location Based Services*, *12*(2), 63-93.

Koski, H. and Luukkonen, J. (2017). Technology giants in patent wars: Competition, litigations and innovation. Mimeo.

Koski, H. ja Luukkonen, J. (2018). Technology salience and investor attention: Does personal data related innovation boost firm value? Etla Working Papers No 59.

Levy, S. 2011. In the Plex: How Google Thinks, Works, and Shapes Our Lives. New York: Simon & Schuster.

Mayer-Schönberger, V. and Cukier, K. (2013). Big Data - A Revolution That Will Transform How We Live, Work and Think. John Murray.

McKinsey Global Institute (2011), Big data. The next frontier for innovation, competition and productivity. http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation.

Olshansky, S.J., Carnes, B.A., Yang, Y.C., Miller, N., Anderson, J., Beltran-Sanchhez, H. and Ricaneck, K. Jr. (2017). The future of smart health. *Computer*, November 2016, pp. 14-21.

Philippon, T. (2016). *The fintech opportunity* (No. w22476). National Bureau of Economic Research.

Sironi P. (2016). FinTech Innovation. From robo-advisors to goal based investing and gamification.

Swan M. (2015). Blockchain. Blueprint for a new economy. Beijing: O'Reilly.

Topol E. (2012). The creative destruction of medicine. How the digital revolution will create better health care. New York: Basic books.

Trajtenberg, M. (2019). Artificial intelligence as the next GPT: A political economy perspective. In *The economics of Artificial Intelligence: An Agenda.* Editors: Agrawal, A., Gans, J and Goldfarb, A. University of Chicago Press-

Wachter R. (2015) The digital doctor. Hope, Hype, and Harm at the Dawn of Medicine's Computer Age. New York: McGrawHill.

Annex 1. Technology domains and IPC classes

Artificial intelligence

IPC	Technology group description
G06N3/00	Biological model Computer systems based on biological models
G06N3/02	Biological model using neural network models
G06N3/04	Biological model Architectures
G06N3/06	Biological model Physical realization
G06N3/063	Biological model using electronic means
G06N3/067	Biological model using optical means
G06N3/08	Biological model Learning methods
G06N3/10	Biological model Simulation on general-purpose computers
G06N3/12	Biological model using genetic models
G06N5/00	Knowledge-based model Computer systems utilizing knowledge-based models
G06N5/02	Knowledge-based model Knowledge representation
G06N5/04	Knowledge-based model Inference methods or devices
G06N7/00	Specific mathematical model. Computer systems based on specific mathematical models
G06N7/02	Specific mathematical model using fuzzy logic
G06N7/04	Specific mathematical model. Physical realization
G06N7/06	Specific mathematical model Simulation on general-purpose computers
G06N7/08	Specific mathematical model using chaos models or non-linear system models
G06N99/00	Other AI technology subject matter not provided for in other groups of this subclass

Health

IPC	Technology group description
A61B5/00	Measuring for diagnostic purposes. Identification of persons. “Measuring" covers also detecting or recording.

Financial services

IPC	Technology group description
G06Q20/00	Payment architectures, schemes or protocols
G06Q20/02	Payment architectures, schemes or protocols involving a neutral third party, e.g. certification authority, notary or trusted third party
G06Q20/04	Payment circuits

G06Q20/08	Payment architectures
G06Q20/22	Payment schemes or models
G06Q20/30	Payment schemes or models characterized by the use of specific devices
G06Q20/34	Payment schemes or models using cards, e.g. integrated circuit cards or magnetic cards
G06Q20/38	Payment protocols
G06K19/10	Record carriers for use with machines and with at least a part designed to carry digital markings - at least one kind of marking being used for authentication, e.g. of credit or identity cards

Location based services

IPC	Technology group description
H04W4/02	Services making use of location information
H04W4/06	Selective distribution of broadcast services, e.g. multimedia broadcast multicast service; Services to user groups; One-way selective calling service
H04W8/02	Processing of mobility data, e.g. registration information at HLR [Home Location Register] or VLR [Visitor Location Register]; Transfer of mobility data
H04W8/18	Processing of user or subscriber data, e.g. subscribed services, user preferences or user profiles; Transfer of user or subscriber data
H04W40/20	Communication routing or communication path finding based on geographic position or location
H04W48/04	Access restriction; Network selection; Access point selection based on user or terminal location or mobility data, e.g. moving direction or speed
H04W64/00	Locating users or terminals for network management purposes
H04H60/49	Arrangements for broadcast applications with a direct linkage to broadcast information or to broadcast space-time; Broadcast-related systems for identifying locations

Sources:

USPTO Class 706 Data processing: Artificial intelligence.

Report on FY2014 Trend survey of patent application technology: Artificial intelligence (2016) https://www.jpo.go.jp/shiryou/pdf/gidou-houkoku/26_21.pdf.

Elinkeinoelämän tutkimuslaitos

The Research Institute of the Finnish Economy

ISSN-L 2323-2420
ISSN 2323-2420 (print)
ISSN 2323-2439 (pdf)

Tel. +358-9-609 900
www.etla.fi
firstname.lastname@etla.fi

Arkadiankatu 23 B
FIN-00100 Helsinki