ETLA

Working Papers | 59

25.5.2018

Investor Attention and Technology Salience:

DOES PERSONAL DATA RELATED INNOVATION BOOST FIRM VALUE?

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

Juha Luukkonen
The Research Institute of the Finnish Economy
juha.luukkonen@etla.fi

Suggested citation:
Koski, Heli & Luukkonen, Juha (25.5.2018). "Investor Attention and Technology Salience: Does Personal Data Related Innovation Boost Firm Value?".

ETLA Working Papers No 59.
http://pub.etla.fi/ETLA-Working-Papers-59.pdf

Abstract

This paper empirically analyzes how markets value personal data related innovation in four prominent domains, in which firms' potential to exploit value from data is identified to be considerable: finance, health, location-based services and artificial intelligence. We link the innovation economics literature to psychology-grounded financial economics theories of investor attention and salience theory. Our data from 117 large technology companies active in the ICT sector from the years 2007–2014 suggest that firms' personal data related innovations and knowledge stocks in technology domains of location-based services and artificial intelligence contributed substantially to firm value. The premiums gained from personal data related innovation were particularly significant for data giants holding knowledge stocks in the location-based service domain. Our empirical results indicate that a strong positive relationship between personal data related knowledge stocks of the location-based services domain and firm value relates primarily to investor attention intensified during periods of media hype. Our data provide new insights into the market valuation of intangible assets: investors seem to overweight more salient right tails of firms' knowledge stocks of emerging technologies while neglecting salient left tails.

Tiivistelmä

Sijoittajien huomio ja teknologian näkyvyys: Lisäävätkö henkilödataan liittyvät innovaatiot yrityksen arvoa?

Tämä aineistoanalyysiin perustuva tutkimus arvioi sitä, miten markkinat arvottavat henkilödataan liittyviä innovaatiota neljällä teknologia-alueella, joilla potentiaali hyödyntää dataa on tunnistettu poikkeuksellisen suureksi: i) sijaintiin perustuvat palvelut, ii) terveysteknologia, iii) rahoitusalan palvelut ja iv) tekoäly, jota hyödynnetään myös kaikilla edellä mainituilla teknologia-alueilla. Tutkimus yhdistää innovaatiotaloustieteellisen kirjallisuuden ja psykologiaan pohjautuvan rahoitusalan teorioita. 117 suuresta ICT-alan teknologiayrityksestä koostuva aineisto vuosilta 2007–2014 osoittaa, että patenttisalkut ovat nostaneet suurten teknologiayritysten arvoa erityisesti niiden henkilötietoja hyödyntävien teknologioiden osalta, jotka liittyvät käyttäjien paikantamiseen ja tekoälyyn. Kansainväliset datajätit ovat hyötyneet selvästi muita yrityksiä enemmän investoinneistaan käyttäjien paikantamiseen liittyvien ja paikkatietoja hyödyntävien teknologioiden kehittämisessä. Tutkimus viittaa siihen, että sijoittajat kiinnittävät erityistä huomiota yritysten aineettoman omaisuuteen hypetetyillä teknologia-alueilla, joiden odotetaan tavoittavan tulevaisuudessa massamarkkinat. Yrityksen henkilödataan liittyvän patenttisalkun arvoa määrittää sen absoluuttista kokoa enemmän sen suhteellinen koko verrattuna teknologia-alueen keskimääräiseen patenttisalkkuun. Sijoittajat näyttävät huomioivan pelkästään positiiviset poikkeamat patenttisalkuissa: teknologia-alueiden innovaatiojohtajien aineettoman omaisuuden määrä kasvattaa merkittävästi yritysten arvoa.

The authors are grateful for Tekes (now Business Finland) for its financial support. We thank Ari Hyytinen, Mika Pantzar, Antonio Vezzani and the participants of the JRC Industrial Research and Innovation team seminar in Seville for helpful comments.

Kiitämme Tekesiä (nyk. Business Finland) tutkimuksen rahoittamisesta ja Ari Hyytistä, Mika Pantzaria, Antonio Vezzania ja JRC:n Industrial Research and Innovation -ryhmän (Sevilla) seminaarin osallistujia hyödyllisistä kommenteista.

Key words: Firm value, Innovation, Personal data, Investor attention, Saliency theory, Technology salience

Asiasanat: Yrityksen arvo, Datatalous, Henkilödata, Innovaatiot, Sijoittajien huomio, Teknologian näkyvyys

JEL: D22, G41, L2, O3

1. Introduction

Digitalization has increasingly transformed business value creation from the use of material resources such as machinery and raw materials towards the exploitation of intangible resources. Both academic and public discussions have considered data to be a powerful source of economic value. For instance, the World Economic Forum (2011) stated that "personal data represents a post-industrial opportunity", and the Economist (May 17, 2017) noted that "The world's most valuable resource is no longer oil, but data".[4] Individuals generate the majority of the world's data, and companies are the main group controlling the data (see, e.g., Computer Science Corporation, 2012)[5]. Despite privacy concerns and stated preferences for privacy that individuals express in various surveys (e.g., National Cyber Security Alliance (NCSA)[6], U.S. Consumer Privacy Index 2016 and Data Protection 2015 Eurobarometer[7]), research shows that individuals are willing to share their personal data relatively easily when incentivized to do so (see, e.g., Athey et al., 2017).[8]

Inconsistency between consumers' stated privacy preferences and their actual behavior, called the privacy paradox, has ultimately allowed companies to extract and use large quantities of personal data. A lack of strict regulations that protect individual's personal data has further supported the corporate exploitation of data.[9] Some prominent domains through which firms' can extract value from data include financial, location-based and health services (see, e.g., McKinsey Global Institute, 2011). Artificial intelligence (i.e., AI) is exploited in data analysis throughout these and various other technology domains. Particularly large technology companies such as Google and Apple have expanded their operations to a wide range of activities involving the use of personal data. These data giants collect and analyze data covering information, for instance, on individuals' behavioral patterns (e.g., via users'

[4] Source: https://www.economist.com/news/leaders/21721656-data-economy-demands-new-approach-antitrust-rules-worlds-most-valuable-resource, Date of access Dec 29, 2017.
[5] Computer Science Corporation (2012) evaluated that approximately 70 percent of the world's data are generated by individuals, while approximately 80 percent are managed by companies.
[6] For a summary of the 2016 survey results, see https://www.trustarc.com/resources/privacy-research/ncsa-consumer-privacy-index-us/. Date of access, Jan 3, 2018.
[7] See http://ec.europa.eu/justice/data-protection/files/factsheets/factsheet_data_protection_eurobarometer_240615_en.pdf. Date of access, Jan 3, 2018.
[8] McKinsey & Company's (2016) survey of US, German and Chinese consumers further suggests that consumers are more willing to share their personal data via navigation and mobility applications than through various other applications.
[9] The General Data Protection Regulation effective on May 2018 is a (EU-wide) applies stricter rules in the field of data protection. See: http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L_.2016.119.01.0001.01.ENG

web searches and online shopping behaviors), locations (e.g., via mobile applications revealing and recording users' locations) and physiological and health-related characteristics and activities (e.g., via health and fitness applications).

The intention of this paper is to empirically assess how markets value personal data related innovation in four emerging technology areas. We analyze investors' prospects for value generation from personal data-related knowledge stocks in finance, health, and location-based services and more generally and partially overlapping with the first three domains from patented inventions related to artificial intelligence. Investors' expectations on how firms' innovations in different technology domains affect their profits are captured by the relationship between firm value and a firm's knowledge assets (see, e.g., Griliches, 1981; Hall et al., 2005; Blundell et al., 1999). This line of literature has made implicit the assumption that investors are rational decision makers who use objective probabilities of the values of knowledge stocks. We contribute to these studies by proposing that investors' valuations of a firm's knowledge stocks are affected by their scarce cognitive resources, causing investors to pay uneven attention to a firm's knowledge stocks in different technology domains.

We relate the innovation economics literature to two streams of psychology-grounded financial economics studies: i) empirical work exploring the relationship between investor attention and asset prices (see, e.g., Odean, 1999; Barber and Odean, 2008; Da et al., 2011) and ii) the relatively recent literature introducing salience theory (see, e.g., Bordalo et al, 2012, 2013; Cosemans and Frehen, 2017). We note that while intangibles are the main determinant of a firm's shareholder value, the value of a firm's knowledge assets is difficult for investors to evaluate, as accounting rules do not require their declaration in official financial statements. Consequently, investors' assessments necessarily rely on other secondary sources of information. We propose that investor attention focuses particularly on firms' knowledge assets in technology domains receiving extensive attention and on the more salient knowledge stocks of a firm in emerging technology domains.

We use firm-level financial information and USPTO patent data from 117 large technology companies active in the ICT sector from 2007 – 2014 for our empirical analysis. Our data show that larger knowledge stocks in personal data related new technologies in location-based services and artificial intelligence domains have boosted firm value substantially. We further find support for our psychologically oriented hypotheses of investor attention and technology salience explaining how markets value firms' knowledge assets. Our data suggest that a positive relationship between firm value and personal data related knowledge stocks in location-based services domain arises primarily from intensified investor attention observed during periods of media hype. The data further suggest that investors overweight salient right tails of a firm's knowledge stocks in all emerging technology domains considered.

The rest of the paper is organized as follows. Section 2 presents a theoretical background and discusses relevant literature. Section 3.1 illustrates innovation trends in light of ideas patented in the USPTO across four technology domains of interest, and section 3.2 introduces the data and variables used for the empirical part of this work. Section 4 presents the estimation results. Section 5 concludes.

2. Previous studies and theoretical background

2.1 Knowledge assets and firm value

Innovation outputs (e.g., patent counts and citations) and inputs (i.e., typically research and development expenditures) have largely been used as a proxy for measuring technological advancement. Research findings in the empirical literature (see, e.g., Bloom and Van Reenen, 2002; Hall et al., 2005) suggest that the order of magnitude of a firm's stocks of innovation assets or knowledge capital closely relates to its market value and further that the relationship between innovation and firm value can be used to shed light on how the market values new technologies. Specifically, a firm's successful innovation activities in technology areas with high revenue potential and growth expectation may raise a firm's future market prospects and thus increase its market value.

The empirical literature exploring the relationship between a firm's innovation activities and market valuation starts with Griliches (1981) who found that a firm's past R&D expenditures and number of patent applications relate positively to its value. Later research on the topic has applied more specific approaches with varying data and conditions. We cover here a few influential studies using mainly UK panel data for different time periods. Blundell et al. (1999) found that a firm's innovation headcount and patent stocks are positively related to the firm's market value and that this relationship is even stronger for firms with higher market shares. In other words, their data show that markets more heavily value the innovative activities of more dominant firms.

The empirical study of Bloom and Van Reenen (2002) concludes that both the quantity and quality of patents matters for firm valuation: the relationship between a firm's patent stock and citation weighed patent stock in relation to the firm's capital stock are positively and statistically significantly related to firm value. They further found that high levels of market uncertainty weaken this relationship or diminish the effects of innovation on firm value. The empirical analysis of Toivanen et al. (2005) compares the impacts of a firm's R&D expenditures, gross investments and patent counts on the firm's valuation. They found that innovation inputs play a more significant role while patent counts are negatively correlated with firm value. They speculated that this negative correlation may be attributed to the difficult appropriability conditions faced by inventors. Additionally, Greenhalgh and Rogers (2006) observed that R&D expenditures and patenting in the European Patent Office increase the market valuation of UK companies but that domestic patenting does not have a significant effect on a firm's valuation. They further found that the market valuation of R&D is lower for sectors involving more competition.

Hall and her (varying) coauthors have studied the relationship between innovation and firm valuation using both USPTO and EPO data. Hall et al. (2005) find that the market values both R&D inputs and outputs measured by the USPTO patent counts and their quality, and that unexpected citations and self-citations seem to be more highly valued by the market than expected ones. The study of Hall et al. (2007) using European firm-level data shows that financial markets particularly value inventions patented in both European and US jurisdictions. They further found that the market valuation of patents of different technology areas varies. Software patents or software related patents appear to generally be more valuable than their non-software counterparts. Their data suggest that markets do not value the quality of software patents but rather their brute quantity. Hall and MacGarvie (2010) explored whether the valuation of software-related patents differed from the valuation of other patented innovations before and after a legislative change made in 1995 to expand the patentability of software in the United States. They reported that widely cited inventions were positively correlated with market valuation, whereas patented ideas with low value increasing the patent stock but not the citation stock were instead negatively associated with firm market value. However, this negative relationship disappeared for software firms following software patentability changes made in 1995, which is somewhat in line with Hall et al. (2007).

We employ Hall et al.'s (2005, 2007) theoretical framework to explore the relationship between innovation and firm value. We use a linear market value equation to estimate firm i's market value at time t (V_{it}) as a function of the firm's physical (A_{it}) and knowledge assets (K_{it}):

$$V_{it} = q_t(A_{it} + \gamma K_{it})^{\sigma} \tag{1}$$

As Griliches (1981) proposed, the equilibrium value of a firm in relation to the replacement value of its tangible assets may deviate from the expected value due to intangible capital or due to a lack of competition in certain markets encountered by the firm. Parameter σ allows for nonconstant scale effects, but as in previous studies, we assume constant returns to scale (i.e., $\sigma = 1$). We use equation (1) to formulate an estimable equation for our empirical exploration:

$$\log Q_{it} = \log\left(\frac{V_{it}}{A_{it}}\right) = \beta_s logS_{it} + \sum_l \gamma_l X_{it}^l + \epsilon_{it} \tag{2}$$

where Q_{it} represents Tobin's q, as measured by the ratio of a firm's market value to its replacement cost for firm i at time t, S_{it} is a firm's sales controlling for firm size and X_{it}^l comprises various R&D, patent and stock ratios. γ_l captures the marginal value of the ratio of a firm's knowledge assets to its physical assets.

2.2 Investor attention and technology salience

This section integrates a framework for investors' valuations of firms' knowledge stocks with psychology-based theories of choice under risk. Investors have scarce cognitive resources and bounded time limiting attention that they can give to available stock market information (Kahneman, 1973). Odean (1999) suggested that investors deal with this problem by restricting the set of stocks to which they pay attention to those that have recently caught their attention or that have recently experienced abnormally good or bad performance. Given that investor attention is unobservable to the researcher, empirical explorations - following the ground-breaking study of Barber and Odean (2008) - have typically used prominent observable phenomena such as news stories, extreme returns and unusual trading volumes as proxies for measuring investor attention. The use of such indirect proxies for studying the effects of investor attention on stock prices relies on the underlying assumption that investors pay (similar) attention to all published information concerning abnormal or extreme stock returns or news articles mentioning a firm.

Da et al. (2011) acknowledged the shortcomings of indirect investor attention measures (i.e., that prominence does not equal attention) and proposed a more direct attention measure for studying the relationship between investor attention and asset prices. They quantify investor attention by the aggregate Google search frequency of a firm’s stock ticker and company name. Various empirical studies published in the financial economics literature have since measured investor attention by Google searches (see, e.g., Dimpfl and Jank, 2016; Chen, 2017). This line of empirical work has aimed at finding more convincing measures for investor attention to explore how observable attention to firms’ stocks at the firm level (e.g., Da et al., 2011) and among dominant companies’ stocks, e.g., using Dow Jones Index (see Dimpfl and Jank, 2016), impact asset prices or stock market fluctuations more generally.

Our theoretical framework connects the innovation economics literature to financial economics theories. We address that the investors’ limited cognitive resources affect their valuations of firms’ knowledge assets. The financial worth and development of a firm’s knowledge assets are more difficult to measure than stock market prices. Firms’ financial statements typically comprise incomplete information concerning firms’ intangible assets. Intangibles are highly important determinants of shareholder value: investors use them to form expectations for the firm’s future profits. However, accounting principles do not require firms to systematically assess and report the value of their intangible assets. Therefore, investors tend to use information other than company financial statements when evaluating the value of firms’ knowledge assets. Markets particularly focus on emerging technologies and on prospects for future profits that are recognized, e.g., by the Gartner hype cycle published since 1995. The Gartner hype cycle, which is keenly followed by many investors, ranks and outlines the analyst’s perception of major emerging technologies and how far from mainstream adoption over the next 10 years they appear to be.

A firm's creation and securing of intellectual property rights for emerging technologies that are expected to be adopted widely forecast higher returns to the firm's innovation output. We propose that investors consequently pay more attention to a firm's innovation and knowledge assets in hyped technology fields. Most investors tend to focus less on a firm's knowledge stocks in technology domains facing less extensive attention or that have no or weak expectations of abnormal future returns to innovation.[10] In other words, we suggest that investors pay more attention and place a more positive value on a firm's knowledge assets in hyped technology domains; a firm's larger stocks of intangible assets in hyped technology domains are more strongly positively associated with firm value than for other technologies. This leads us to the following hypotheses:

1.1 Firm value relates positively to extensive attention or hype in technology domains in which a firm has knowledge stocks.

1.2 Higher magnitudes and higher quality of firm knowledge stocks in hyped technology domain are more positively related to firm value than firm knowledge stocks in other technology domains.

Salience theory employs the definition of salience used in the psychology literature: "when one's attention is differentially directed to one portion of the environment rather than to others, the information contained in that portion will receive disproportionate weighting in subsequent judgments" (Taylor and Thompson, 1982; Bordalo et al., 2012). Salience theory captures the context-dependence of investor decision making (i.e., that preferences change when the context in which choices are presented changes). In stock markets, investors tend to focus on the most salient payoffs and are consequently attracted to purchasing stocks with salient upsides, creating further excess demand for and the overvaluation of such stocks (see Cosemans and Frehen, 2017). Instead, stocks with salient downsides tend to become undervalued and earn higher returns later on while future returns of overvalued stocks are lower. Furthermore, salience theory proposes that due to differences in the salience of different states, probability weights decision makers attach to different states and use while making choices are not equal to objective probabilities. Decision makers tend to be risk-seeking when a lottery's upsides are salient and risk-averse when downsides are salient.

[10] Relatedly, a theoretical study by de Marzo et al. (2007) addresses the question of investor overinvestment in new, high-risk technologies.

We propose that investors similarly act as “local thinkers” while facing incomplete information concerning the value of a firm’s intangibles and concerning uncertainty related to the future market developments of emerging technologies. In other words, investors use weights that favor the knowledge stocks of emerging technologies with more salient payoffs in their decision making. The context dependence of decision making means that investors compare the expected value of or payoffs from a firm’s knowledge stocks in a certain technology domain to those of other firms’ knowledge stocks in the same technology domain. Investors thus not only use the size and expected profits of a firm’s knowledge stock while assessing firm value, but they also overweight (underweight) more salient (less salient) knowledge stocks of the firm. Here, the salience of knowledge stocks is dependent on a firm’s innovation performance relative to the innovation performance of other firms in certain technology domain. Technology salience is thus a *firm-specific* attribute, while theories concerning investor attention captures *technology-specific* variations in investor attention.

Salience theory suggests that investors use value differences rather than absolute values for their decision making and overweight the tails of value distributions when they are salient. This is the major deviation of salience theory from the prospect theory of Tversky and Kahneman (1992), suggesting that the ranking of payoffs determines distortion probabilities and that tails are always over-weighted. Instead, salience theory emphasizes the importance of choice contexts and the relative magnitude of payoffs. Here we assume that the salience value of a firm’s knowledge stock K_{ijt} in technology domain j (or technology salience) at time t depends on its distance from the average value of firms’ knowledge stocks in the same technology domain, $\overline{K}_{jt}$, relative to the average values of knowledge stocks in technology domain j:

$$\frac{K_{ijt} - \overline{K}_{jt}}{\overline{K}_{jt}}$$

The denominator of the technology salience term captures the assumption of diminishing sensitivity based on the state of technology maturity. When emerging technology j is in the more mature (early) stages of adoption, the average knowledge stock per company in technology domain j tends to be higher (smaller). As technology j achieves more widespread adoption and as revenues from this new technology increase, differences in knowledge stocks matter less and the salience of a firm’s technology stock j decreases. We further assume convexity such that the diminishing sensitivity gets weaker with greater average knowledge stocks in technology domain j.

We consequently propose the following technology salience hypothesis:

2.1 Investors tend to overweight a firm's salient knowledge stock such that the more salient its positive (negative) deviation is from the average magnitude and quality of knowledge stock in technology domain, the more it increases (decreases) firm value.

We further test the alternative prospect theory hypothesis:

2.2 Investors always overweight tails of knowledge stocks such that a positive (negative) deviation of a firm's knowledge stock from the average magnitude and quality of knowledge stock in technology domain relates positively (negatively) to firm value.

3 Data and some descriptive findings

3.1 Data

Our data were drawn from two databases. First, we used the web-based search service of Patent Inspiration (www.patentinspiration.com) to obtain information on patents granted by the USPTO. We extracted information on all patents falling under IPC categories H04 (i.e., electrical communication systems), G06 (i.e., computing, calculating and counting devices) and A61B5/00 (i.e., measuring for diagnostic purposes and the identification of persons). The majority of software-related patents and all AI patents are classified under IPC class G06. Patented ideas related to collecting, transmitting, exchanging and analyzing personal data largely appear under IPC classes G06 and H04. These two IPC classes also dominate the patent portfolios of data giants. For example, Google, Amazon and Facebook, respectively, published 65 (26), 70 (29) and 74 (41) % of their USPTO patents under IPC class G06 (H04) from 2005 -2014 (see Koski and Luukkonen, 2017). Technologies measuring, collecting and transmitting personal health, wellness and fitness data are further patented under IPC category A61B5/00.

We extracted all patents of four technology domains (i.e., location-based services, financial services, health and AI[11]) from January 2001 to December 2014. The IPC codes of these domains are described in Annex 1. As our empirical work focuses on patented ideas related to collecting, transmitting and analyzing personal data, we performed a search using a search criterion stating that the term "personal data" or "personal information" had to appear in the title, abstract or description of a patent application. This search criterion separates patented ideas for new technologies targeted for personal data related tasks from other patents of the four selected domains. Given the importance of U.S. markets in these technology areas (U.S. is the single largest software market in the world) and because software are patentable in the USPTO (unlike in the European Patent Office, EPO), we restricted our analysis to patents granted in the United States. Furthermore, according to the PwC survey (2017), investors continue to view the United States as the country most important to companies' overall growth prospects.

Second, we extracted the financial statements of patentees available from publicly traded companies in the Bureaua van Dijk's Orbis database, which comprises private and listed companies operating around the world. Due to the limited availability of financial statement data, observations were restricted to the years 2007-2014. We complemented these data for some major USPTO patentees lacking information in the Orbis database with financial information obtained from publicly available sources (e.g., financial statements available in each firm's web page). The data covered more than 170 large companies, from which we removed non-ICT companies. Consequently, our estimations cover a sample of 117 large technology companies with 768 observations from SIC classes 357 (i.e., computer and office equipment), 366 (i.e., communications equipment), 367 (i.e., electronics components and accessories), 481 (i.e., telephone communications), 482 (i.e., telegraph and other message communications), 489 (i.e., communication services), 594 (i.e., miscellaneous shopping goods stores; merely Amazon), 599 (i.e., retail stores not classified elsewhere (Ebay)) and 737 (i.e., computer programming, data processing, and other computer related services).

We generated technology-specific investor attention measure by extracting the monthly Google search frequencies of the terms "artificial intelligence", "location-based services", "wearables" and "fintech" -

[11] Annex 1 describes the technology domains and IPC classes used to extract patents.

related to the four technology domains of interest - from the website. Google normalizes monthly search data by 1) dividing the number of users' searches of the term in question by all Google searches made in a given month and 2) setting the term's highest monthly value during the time period considered to 100 and presenting the term's other monthly search frequency values in relation to the maximum. We further divide the normalized search volume index by 100. Thus, our Google search volume index (SVI_{jt}) generate values of between 0 and 1.

To test hypothesis 1.1, we added SVI measures to the basic model for firms that had patented innovations in the four technology domains of interest to empirically assess whether the hype in the technology domains relates to firm value. To test hypothesis 1.2, we used interactions of knowledge stock variables and the SVI variables for each technology domain: SVI_{jt} x K_{ijt}. The idea here is that when the hype for a certain technology is at its peak, the future profit prospects of technology grow, and investors pay more attention to firms' knowledge stocks in that technology domain. Consequently, investors give more weight to firms' patent stocks in hyped technology fields while assessing firm value. SVI-interacted patent stocks to R&D ratio variables thus generate the same value as fully observed patent stock variables when technology hype and investor attention reach a peak (i.e., when the SVI measure is valued at 1). Otherwise, we assume that when hype and investor attention are weaker, investors set weights below 1 – i.e., weights decreasing with declining hype or media attention - to patent stock variables in the technology domain while assessing firm value.

Salience theory suggests that decision makers place differing weight on choices based on actual payoffs and their salience (Bordalo et al., 2012). Cosemans and Frehen compute salience weights from values they calibrated in their prior 2012 paper "to match empirical evidence on long-shot lotteries". We do not make any assumptions on the weights investors attach to different states of technology salience. We allowed our technology salience measure (TS_{jt}) – i.e., the distance of firm knowledge stock j at time t from average knowledge stock $\bar{J}$ divided by the average knowledge stock – to be non-linearly related to firm value and for investors to set different weights to the lowest and highest percentiles of the technology salience measure. We formed dummy variables for the annual i) lowest 10 %, ii) highest 10 % and iii) middle 80 % of technology salience measure in each technology domain. We then multiplied each technology salience variable by the set of dummy variables. The first two interaction variables capture how investors value the magnitude of deviations of the most salient knowledge stocks from the average of a certain technology domain, and the third one covers less salient knowledge stocks. We conduct a robustness test by changing the percentile limits of the most salient knowledge stocks to the minimum and maximum 15 %.

The above empirical framework based on the three interaction variables was used to test hypothesis 2.1. We used three dummy variables for the annual lowest 10 %, highest 10 % and middle 80 % of the technology salience measure in each technology domain to test the alternative prospect theory hypothesis. If prospect theory correctly captures investor behavior (i.e., investors always overweigh tails), we should observe statistically significant coefficients for dummy variables capturing the 10 % tails of the knowledge stock distribution in a certain technology domain. We also undertook robustness tests by using dummy variables for the lowest and highest 15 % (and mid 70 %) of the technology salience measures.

As a dependent variable of the estimated models, we used log Tobin's q excluding intangibles. In addition to our major explanatory variables measuring investor attention and technology salience, we constructed variables used in estimations following the work of Hall et al. (2005). A firm's knowledge assets were measured by R&D intensity (or R&D stocks divided by the total value of assets), patent count stocks generated in relation to R&D stocks and average citations per patent as follows:

$$\sum_{l} \gamma_l X_{ijt}^{l} = \gamma_1 \frac{R\&D\ stock_{it}}{A_{it}} + \gamma_{2j} \frac{Patent\ stock_{ijt}}{R\&D\ stock_{it}} + \gamma_{3j} \frac{Citation\ stock_{ijt}}{Patent\ stock_{ijt}}$$

Following previous studies, knowledge stocks were constructed from a 15 % annual depreciation rate for each stock as follows:

$$K_{ijt} = K_{ijt} + 0{,}85 * \ K_{ijt-1}$$

γ_l are estimated coefficients capturing investors' expectations of future profits that a firm will gain from its knowledge assets. A firm's knowledge stocks were measured by the size (i.e., patent count stock) and quality (i.e., patent citation stock) of stocks of published patents in each technology domain of interest. Firms' knowledge stocks or other patents published in IPC classes G06 and/or H04 were used as control variables.

- TABLE 1 HERE –

To explore whether markets value differently knowledge stocks of technology companies identified as massive collectors and exploiters of personal data, we generated a dummy variable called data_giant. This variable is valued at 1 for Amazon, Apple, Google, IBM, Yahoo, Facebook, and Microsoft and at 0 otherwise. This is a rough proxy for companies intensely exploiting personal data; it is possible that our sample includes some less well known firms that heavily exploit personal data. We further controlled for a firm's home country and year-specific effects with dummy variables. Potential industry-specific variations in a firm's propensity to patent and in patent quality were controlled by 3-digit SIC (i.e., Standard Industrial Classification) dummy variables.

3.2 Technology hypes and innovation in selected domains

This section sheds light on personal data related patented ideas and on technology hype cycles in the four technology domains of interest. The health sector manages highly personal and often sensitive customer data. Various data giants have entered the digital health sector: they have patented ideas collecting personal health data and launched new health technology applications and devices. For instance, in April 2015, IBM launched Watson Health, which collects personal health data, and Watson Health Cloud, which enables the combination, de-identification and sharing of health data for use by, e.g., doctors, insurers and researchers.[12] Furthermore, in March 2016, IBM announced that it would invest USD 150 million[13] over the next several years to Watson Health's first European Center of Excellence in Milan. Data giants' interest in the digital health sector has further been reflected in their hiring decisions. Various technology companies such as Google, Apple, Microsoft and IBM that focus on gathering, analyzing and storing data have hired leading biomedical researchers (see Wilbanks and

[12] See https://www-03.ibm.com/press/us/en/pressrelease/49436.wss. Date of access, Oct 10, 2017.
[13] See, e.g., https://www-03.ibm.com/press/us/en/pressrelease/49436.wss. Date of access Oct 10, 2017.

Topol, 2016). These developments echo large technology companies' expectations on the profits that can be extracted from investments made in innovation in the digital health technology field.

We restricted our analysis concerning health domain to wearable health and fitness devices 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[14]. Currently, three main types of wearable devices and applications may collect personal health data: i) those monitoring physiological attributes associated with certain diseases such as 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). According to Gartner's hype cycle, mobile health monitoring is found among major emerging technologies for 2012 - 2014, and wearables are observed for 2013 to 2015, both reaching mainstream adoption within 5-10 years at the time of analysis.

Markets for location-based services such as geotargeted advertisements and offers, e.g., for local restaurants and shops, have emerged alongside the widespread adoption of smartphones. Gartner's 2006 hype cycle forecasted that "location-aware technologies should hit maturity in less than two years"[15] and that "location-aware applications will hit mainstream adoption in the next two to five years." In 2009, location-based mobile advertising still occupied rather early stages of development, but large technology companies' actions already reflected their intentions to capitalize on personal data use for location-based services and advertising. For instance, in November 2009, Google acquired mobile advertisement network AdMod for $ 750 M[16], and in April 2011, eBay purchased location-based media and advertising company WHERE[17]. In 2016, markets for location-based mobile advertisements

[14] See: http://www.marketsandmarkets.com/PressReleases/wearable-electronics.asp, accessed April 7, 2017.

[15] Location-aware technologies were defined as "the use of GPS (global positioning system), assisted GPS (A-GPS), Enhanced Observed Time Difference (EOTD), enhanced GPS (E-GPS), and other technologies in the cellular network and handset to locate a mobile user." (see https://www.gartner.com/newsroom/id/495475, accessed February 12, 2018).

[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://techcrunch.com/2011/04/20/ebay-acquires-location-based-media-and-advertising-company-where/ (accessed February 12, 2018).

grew to over $12 billion and were expected to reach $32 billion by 2021, covering 45 percent of total mobile advertising revenues.[18]

The finance sector manages and analyzes vast quantities of customer data such as financial records and credit card information. The European Banking Authority (2017) noticed an accelerated use of personal data in the financial services sector: "…the EBA has observed a growing number of financial institutions using consumer data in innovative ways across the EBA's regulatory remit, comprising deposits, mortgages, personal loans, payment accounts, payment services and electronic money." Large technology companies have exhibited an interest in competing with traditional banks at least in certain segments of the financial services sector. For instance, Facebook has patented a technology that can be applied for credit grading that bases the acceptance of individuals' loan applications on the credit ratings of borrowers' social networks[19], and Amazon provides loans to sellers through its online marketplace.

Artificial intelligence or machine learning applied to personal data relates closely to all of 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: algorithms automatize many tasks by, on the one hand, reducing human errors and decreasing the processing times of, e.g., insurance decisions. On the other

---

[18] See, e.g., https://bluedotinnovation.com/location-based-advertising-local-mobile-ads.html (accessed February 12, 2018).

[19]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).

hand, new problems may emerge. AI-based solutions enable ways to base decisions - such as those concerning loan applications – using data other than traditionally used data (e.g., borrowers' loans to income ratios). However, when AI algorithms crunch such data as borrowers' Internet browsing histories and social network and online shopping patterns, credit decisions can not only become flawed by erroneous data (of which use potential borrowers may not even be aware of) but can also spur discriminating or unfair decisions. Overall, the use of AI algorithms and automatizations may offer many benefits such as enhanced levels of efficiency and accuracy, but this also raises new concerns relating to the collection, combination and analysis of different forms of personal data.

- TABLE 2 HERE –

Table 2 presents a set of exemplary titles of personal data related patented ideas for the four selected domains of major data giants (i.e., Google, Apple, Facebook and Amazon). These titles illustrate that emerging technologies in these domains are generated to, e.g., infer a person's current location and predict the subsequent movements and locations based on the person's location history; they are aimed at identifying users via facial-, voice- and motion-based recognition; they exploit information on users' moods and behaviors collected from multiple data sources; and they are aimed at estimating a person's age. These titles illustrate data giants' aspirations to target the market for and gain market power in the exploitation of personal data on various fronts by securing the ownership of patent rights for personal data related innovation.

- FIGURE 1 HERE –

The number of personal data related patent applications in the USPTO for financial services, location-based services, health, and artificial intelligence technology domains has dramatically increased since the early 2000s (Koski and Pantzar, 2018). The USPTO aims at publishing patent applications if not earlier, promptly after the 18 months of an initial filing application date. Figure 2 illustrates the annual number of patents published in four technology domains for the sampled companies for 2007 – 2014. There has been a clearly increasing trend in published personal data related patents in financial and location-based services, whereas published patents involving personal health underwent a growing trend at the very end of the sample period. It seems that the annual number of personal data related AI patent publications has followed a rather flat pattern of growth, temporarily declining during the years 2010 – 2011, and elevated again since 2012.

Figure 2 shows Google search frequency volume indices from January 2007 to December 2017 for terms relating to the four technology domains of interest. The figure implies that "wearables" and "fintech" did not reach attention hype until the end of or after our sample years 2007 - 2014. This descriptive finding concerning wearables is consistent with annual Gartner hype cycles identifying mobile health monitoring and wearables as major emerging technologies from 2012 – 2015. Emerging technologies related to innovation in financial services are also presented relatively recently in the Gartner hype cycle context. For instance, mobile over the air (OTA) payment and biometric authentication methods appeared in Gartner 2013 hype cycle for emerging technologies, whereas cryptocurrencies and digital security were not pinpointed in the hype cycle until 2014 and 2015 and blockchain not until 2016 and 2017.

Instead, location-based services already underwent a peak in terms of Google searches, and artificial intelligence appeared to be relatively popular during the sampled years. Gartner's annual hype cycles for emerging technologies provide a rather similar picture: location-aware applications are listed under Gartner's 2005 hype cycle with the expected time to reaching productivity plateauing at two to five years. Artificial intelligence or machine learning as such are not mentioned as Gartner's emerging technologies until 2015 - 2017. Instead, various AI-related technologies such as autonomous vehicles appeared among major emerging technologies between 2010 and 2014.

4 Estimation results

We estimated the following random effects model for the valuation of firms' personal data related knowledge assets (results reported in Table 3):

$$\log Q_{it} = \alpha_0 + \beta_s logS_{it} + \sum_l \gamma_l X_{it}^l \ + a_i + u_{it} \qquad (3)$$

To test investor attention hypotheses 1.1 and 1.2, we applied Google search volume index (SVI_{jt})

of technology domain j for time t to equation 3 and its interactions with knowledge stock variables SVI_{jt} x K_{ijt} (results reported in Table 5). For testing technology salience hypothesis, we added to equation 3 variables for the annual technology salience of a firm's knowledge stock in technology domain j for the lowest 10 %, middle 80 % and highest 10 % percentiles (results reported in Table 6). Our test of prospect theory involved using dummy variables for the percentiles to which a firm's knowledge stock in technology domain j at time t was part of (results reported in Table 7):

The random effects model was our preferred choice over the fixed effects model as it allows us to include essentially important time invariant explanatory variables (e.g., dummies for data giants) in the model.[20] Prior to estimating different variants of the model, we applied a Breusch and Pagan Lagrange Multiplier test on random effects. The test clearly rejected the null hypothesis that var(u) = 0 (i.e., that there are no random effects and that pooled OLS estimations would be suitable). We further tested the suitability of our model specification based on Arellano's (1993) artificial estimation approach, which re-estimates the random effects model by adding to the set of explanatory variables additional regressors measuring deviations of original regressors from the mean. The Sargan-Hansen test statistics did not reject the appropriateness of the random effects model relative to the fixed effects model.[21]

- TABLE 3 HERE –

[20] We also estimated the reference model using non-linear least squares. This produced rather similar results to those obtained by the random effect model. However, estimation results of the random effects model are easier to interpret than those of the non-linear least squares model.
[21] The test is similar to the Hausman test for the fixed vs. random effects model otherwise, but it also (unlike the Hausman test) extends to heteroskedastic- and cluster-robust models.

The estimation results of model 1 show that knowledge stocks related to personal data related patented ideas in artificial intelligence and location-based services domains have positive and clearly statistically significant coefficients (see Table 3). Instead, the estimated coefficient of the knowledge stocks in the health domain appears to be negative and statistically significant. The knowledge stocks for financial services and the reference group of all other patents in G06 and H04 IPC classes are not statistically significantly related to firm value. Our estimation results for the relationship between the citation stocks and firm value are partly similar to those found by Hall and MacGarvie (2010), which suggests that the citation stocks of software patents tend to rather be negatively related to firm value. In our estimations, the citation stocks of artificial intelligence and the reference groups of patents appear negative and statistically significant. The citations stocks of patents in other technological domains are not strongly statistically significantly related to firm value.

- TABLE 4 HERE –

Table 4 presents the estimation results of the reference model including a dummy variable for the data giants and interactions of the data giant dummy, patent and citation stocks. Our data indicate that the relationship between firm value and knowledge stocks of personal data related new technology in location-based services domain is clearly of stronger than average magnitude for the data giants. The average marginal effect (not shown in Table 4)[22] for variable “L patent stock/RD” was measured as 0.29 for the data giants and as 0.03 for other companies. These empirical findings show that a one percent increase in personal data related published patent stock in the location-based services domain increases the data giants’ firm value more than nine-fold relative to an increase in the value of other companies. The citation stock estimations further suggest that quality- or citation-adjusted knowledge stocks of the personal data related financial services domain are particularly valuable for the data

[22] Tables for average marginal effects are available from the authors.

giants. The average marginal effect of variable "F citation stock/patent" was measured as 0.23 for the data giants and as highly statistically significant, while for other companies, the marginal effect was substantially smaller and not statistically significant. These empirical findings show that markets value the knowledge stocks of dominant companies more consistent with empirical research reported by Blundell et al. (1999).

- TABLE 5 HERE –

The google search volume index (SVI) dummy variable for "fintech" is clearly positively and significantly related to firm value while SVI for "wearables" relates negatively to firm value. Interactions between variable SVI_L and variables "L patent stock/RD" and "L citation stock/patents" appear to be positive and highly statistically significant. Instead, the estimated coefficient for variable "L patent stock/RD" is now not statistically significant (and negative). After the inclusion of the interaction terms into the model, coefficients of the knowledge stock variables express what the impact of the knowledge stock (size and quality) in question would be if the corresponding SVI variable took a value of 0 (i.e., if there were no expressed media attention or hype). In other words, the estimation results show that the impact of the size and quality of personal data related knowledge stocks in location-based services on firm value are strongly associated with media hype or investor attention. Moreover, the negative and statistically significant coefficient for SVI_H and variable "H patent stock/RD" and the non-significance of the coefficient for variable "H patent stock/RD" (that was negative in models 1 and 2) imply that bad publicity may have decreased the firm value of those companies with relatively greater patent stocks of technologies that can be used to collect personal health data.

- TABLE 6 HERE –

Table 6 shows the estimation results of the model for technology salience using 10 and 15 percentile limits for the most salient knowledge stocks. Interestingly, in all technology domains of interest, except health (of which most salient knowledge stocks show a negative coefficient), coefficients for the right

10 and 15 percentile tails of the technology salience measure appear to be positive and statistically significant. The left 10 and 15 percentile tails of deviations of knowledge stocks from the averages do not relate statistically significantly to firm value. For location-based services and AI, the middle 80 percentile reference groups of technology salience relate positively and negatively, respectively, to firm value but when the distribution tails are widened to 15 percentiles, this statistical significance disappears.

- TABLE 7 HERE –

Table 7 presents estimation results for the prospect theory hypothesis. Here we use dummy variables capturing whether a firm's knowledge stock size and value in technology domains of interest appear in the 10 or 15 percentile tails or in the mid of the distribution. Our data, by and large, reject the prospect theory hypothesis suggesting that investors always overweight tails.

5 Conclusions

This paper presents one of the first empirical explorations shedding light on the "black box" of how technology companies extract value from personal data related innovation, or how a firm's market value depends on its personal data related knowledge stocks. It also contributes more generally to the empirical innovation economics literature on the market valuation of intangible assets, emphasizing – unlike previous studies – the role of investors' cognitive capacities in the valuation of firms' assets. Our data for 2007 – 2014 suggest that firms' personal data related innovations and knowledge stocks in technology domains of location-based services and artificial intelligence have contributed substantially to firm value among large technology companies active in the ICT sector.

The estimated coefficients of knowledge stocks comprising personal data related patented ideas in the location-based services domain are highly statistically significant. The importance of personal data related new technologies in the location-based services domain is at least partly explained by the emergence and rapid growth of markets for location-based mobile advertising from the second half of the 2010s. The market prospects of location-based services were reflected, e.g., by data giants' acquisitions of various location-based mobile advertisement companies. Our data show that markets have valued the knowledge stocks of data giants in this technology domain more than those of other companies. Among data giants, a one percentage point increase in the size of personal data related patent stocks relative to R&D in location-based services domain was related to an approximately 28 percent increase in firm value. Among other companies, the corresponding average increase in firm value due to an increase in the size of a knowledge stock in location-based services was 3 percent. The dominant personal data exploiters clearly gained higher premiums from their personal data related innovation, at least in certain influential emerging technology domains, compared with other large technology companies.

Our theoretical framework links innovation economics literature on the relationship between a firm's knowledge stocks and firm value to psychologically grounded finance literature on investor attention and salience theory. We propose that investors' bounded cognitive capacities essentially affect their assessments of major determinants of shareholder value, i.e., a firm's intangible assets. Our research findings provide some support for our proposition that investor attention is drawn to emerging technologies attracting more media attention and that knowledge stocks in hyped technology domains have a stronger impact on firm value. It seems that the strong positive relationship we find between personal data related knowledge stocks in location-based services and firm value relates primarily to investor attention that is intensified during times of media hype. In other words, technology-specific variations in investor attention are relevant to the valuation of firms' intangible assets: firms with significant technology stocks or knowledge assets in hyped emerging technology domains may be overvalued due to attention from investors.

Furthermore, our data suggest that technology salience matters or that investors rather assess how much a firm's knowledge stock in a certain technology domain differs from that of its competitors than the absolute value of the firm's knowledge stock. Furthermore, our empirical work shows that investors do not merely pay attention to the tails of knowledge stock distribution as prospect theory would suggest, but they also assess how much observed values in tails deviate from the reference point. Our data show that knowledge stocks affect firm value primarily via the salient upward deviating knowledge stock sizes of emerging technologies related to personal data (i.e., the knowledge stocks of innovation leaders). The salience of the size of knowledge stocks dominates, but data also provide evidence showing that the salience of knowledge stock quality may impact firm value.

Our empirical findings offer insights into the roles of investor attention and technology salience in the valuation of knowledge stocks in selected domains of personal data related emerging technologies. They strongly indicate that investor attention focuses on forerunners in the emerging technology domains. It seems that the right tail deviations of leading innovator companies are noticed and that these abnormally or exceptionally large knowledge stocks dominate the valuation of firms' knowledge stocks. To our knowledge, previously reported studies have not addressed and empirically approached these psychologically grounded theories of the valuation of a firm's knowledge stocks. An intriguing question that is left for future work concerns whether our empirical conclusions are specific to the technology domains considered or whether they apply to the market valuation of emerging technologies more generally. Furthermore, to what extent investors' overvaluation of salient right tails of knowledge stocks instigate stock market bubbles involving emerging technologies?

Sources

Arellano, M. (1993). On the testing of correlated effects with panel data, *Journal of Econometrics* 59, 87-97.

Athey, S., Catalini, C. and Tucker, C.E. (2017). The digital privacy paradox: Small money, small costs, small talk (September 27, 2017). MIT Sloan Research Paper No. 5196-17; Stanford University Graduate School of Business Research Paper No. 17-14. Available at SSRN: https://ssrn.com/abstract=2916489 or http://dx.doi.org/10.2139/ssrn.2916489

Barber, B. M. and Odean, T. (2008). All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *Review of Financial Studies 21*, 785-818.

Bloom, N. & J. Van Reenen (2002): Patents, real options and firm performance. *Economic Journal 112*, 97-116.

Blundell, R., R. Griffith & J. Van Reenen (1999): Market share, market value and innovation in a panel of British manufacturing firms. *Review of Economic Studies* 66, 529-554.

Bordalo, P., Gennaioli, N. and Shleifer, A. (2012). Salience theory of choice under risk. *Quarterly Journal of Economics* 127, 1243-1285.

Bordalo, P., Gennaioli, N. and Shleifer, A. (2012). Salience and asset prices. *American Economic Review* 103, 623-628.

Chen, T. (2017). Investor attention and global stock returns. *Journal of Behavioral Finance* 18, 358-372.

Computer Science Corporation, CSC (2012). Big Data Universe Beginning to Explode. http://www.csc.com/insights/flxwd/78931-big_data_universe_beginning_to_explode.

Cosemans, M. and Frehen, R. (2017). Salience theory and stock prices: Empirical evidence. 2017 SFS Cavalcade Paper. Available at SSRN: https://ssrn.com/abstract=2887956 or http://dx.doi.org/10.2139/ssrn.2887956

Da, Z., Engelberg, J. and Gao, P. (2011). In search of attention. *Journal of Finance* 66, 1461-1499.

DeMarzo, P., Kaniel, R. and Kremer, I. (2007). Technological innovation and real investment booms and busts, *Journal of Financial Economics* 85, 735-754.

Dimpfl, T. and Jank, S. (2016). Can Internet search queries help to predict stock market volatility? *European Financial Management* 22, 171–192.

European Banking Authority (2017). Report on innovative uses of consumer data by financial institutions. 28 June 2017.

Greenhalgh, C., and M. Rogers (2006): The value of innovation: The interaction of competition, R&D and IP. *Research Policy* 35, 562-580.

Griliches, Zvi (1981). Market value, R&D and patents*. Economics Letters, vol. 2. pp. 183-187.*

Hall, B., A. Jaffe & M. Trajtenberg (2005): Market Value and Patent Citations. *The RAND Journal of Economics* Vol. 36, No. 1, 2005. pp. 16-38

Hall B., G. Thoma, S. Torrisi (2007): The market value of patents and R&D: Evidence from European firms*. NBER Working Paper No. 13426*

Hall, B., M. MacGarvie (2010): The private value of software patents. *Research Policy Volume 39, Issue 7. pp. 994-1009*

Kahneman, D. (1973). *Attention and Effort*. Prentice-Hall, Englewood Cliffs, NJ.

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

Koski, H. and Pantzar, M. (2018). Personal data – market in making. Mimeo.

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

McKinsey&Company (2016). Car data: paving the way to value-creating mobility. Perspectives on a new automotive business model. Advanced Industries. March 2016.

Odean, T. (1999). Do investors trade too much? *American Economic Review* 1279–98.

PwC (2017). Global Investor Survey/February 2017. https://www.pwc.com/gx/en/ceo-survey/2017/industries/pwc-global-investor-survey-feb2017.pdf

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

Taylor, S., and Thompson, S. (1982). Stalking the elusive vividness effect. Psychological Review 89, 155–181.

Toivanen, O., P. Stoneman, and D. Bosworth (2002): Innovation and the market value of UK firms, 1989-1995. *Oxford Bulletin of Economics and Statistics* 64, 39-61.

Tversky, A., and Kahneman, D. (1992) Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty* 5, 297-323.

Wilbanks, J.T. and Topol, E.J. (2016). Stop the privatization of health data. *Nature* 20 July, 2016. https://www.nature.com/news/stop-the-privatization-of-health-data-1.20268. Accessed Jan 8, 2018.

World Economic Forum (2011). Personal data: The emergence of a new asset class.

Variable name	Description	Mean	S.D.	Median
lnTobin1	Tobin's Q without intangible assets	0.359	1.003	0.410
lnGrossSales	Annual gross sales, logarithmic transformation	8.300	2.220	8.526
RD	Research & development expenditures (1000$)	1091	1949	361
Marketcapitalization	Total dollar market value of a company's outstanding shares (1000$)	30323	63747	7672
Totalassets_nointangibles	Total tangible assets (1000$)	20815	36512	5542
RDoverAsset	R&D stock over tangible assets	0.572	1.455	0.385
AIstock_overRD_pd	Artificial intelligence patents: published patent stock over R&D stock	0.000	0.002	0.000
Fstock_overRD_pd	Finance patents: -||-	0.002	0.021	0.000
Hstock_overRD_pd	Health patents: -||-	0.010	0.163	0.000
Lstock_overRD_pd	Location patents: -||-	0.001	0.003	0.000
GHstock_overRD_nopd	All other G06/H04 patents: published patent stock over R&D stock	0.726	3.358	0.197
FwCover_stockAI_pd	Artificial intelligence patents: adjusted citation stock over published patent stock	0.172	0.595	0.000
FwCover_stockF_pd	Finance patents: -||-	0.335	0.977	0.000
FwCover_stockH_pd	Health patents: -||-	0.074	0.303	0.000
FwCover_stockL_pd	Location patents: -||-	0.202	0.552	0.000
FwCover_stockGH_nopd	All other G06/H04 patents: adjusted citation stock over published patent stock	1.169	0.787	0.991
SVI_AI	Google search volume index for artificial intelligence	0.631	0.051	0.623
SVI_F	-||- finance	0.366	0.141	0.301
SVI_H	-||- health	0.277	0.206	0.218
SVI_L	-||- location	0.488	0.136	0.511

n= 768

Table 2. Examples of the titles of patented personal data innovation in selected domains

Company	Finance	Location-based services	Health	Artificial intelligence
Google	Identifying consumers in a transaction via facial recognition, Managing devices associated with a digital wallet account Text message payment, Hands-free transactions with voice recognition.	Inferring a current location based on a user location history, Sharing of profile information with content providers, Providing content based on previously determined device locations.	Physiological measurement using wearable device, Molded electronic structures in body-mountable devices, Contact lenses having two-electrode electrochemical sensors.	Methods, systems, and media for personalizing computerized services based on mood and/or behavior information from multiple data sources, Systems and methods for promoting search results based on personal information, Estimating age using multiple classifiers.
Apple	User interface for payments, Motion-based payment confirmation, Allocation and distribution of payment for podcast services.	Identifying and locating users on a mobile network, Facilitating access to location-specific information using wireless devices, Dynamic content presentation based on proximity and user data.	Seamlessly embedded heart rate monitor, Personal item network and associated methods.	Intelligent automated assistant, content item recommendations based on content attribute sequence
Facebook	Systems and methods for providing subsequent payment options for identified eligible users, Performing risk checks for electronic remittances.	Predicting locations and movements of users based on historical locations for users of an online system, Travel recommendations on online social networks, Personalized location information for mobile devices.		Systems and methods for identifying users in media content based on poselets and neural networks, Systems and methods for estimating user attention, Tag prediction for content based on user metadata, Methods and systems for recommending applications.
Amazon	User-to-user payment service; credit card reader authenticator, Image analysis for user authentication, Real-time mobile wallet server.	Cellular system information sharing, Content display controls based on environmental factors.		User tracking based on client-side browse history, Method for using customer attributes to select a service representative, Determining user interest from non-explicit cues.

Table 3. Estimation results of the Random Effects model for the valuation of firms' personal data related knowledge assets

Dep. variable: log Tobin's q	(1)	(2)
Log sales	-0.0790*	-0.0944**
	(-1.82)	(-2.14)
RD stock/assets	0.0556***	0.0544***
	(16.26)	(15.41)
AI patent stock/RD	21.84***	28.12***
	(5.47)	(5.97)
F patent stock/RD	1.458	1.234
	(1.48)	(1.36)
H patent stock/RD	-0.124*	-0.284***
	(-1.86)	(-2.78)
L patent stock/RD	55.49***	54.20***
	(8.47)	(7.63)
G04&H06 patent stock/RD	-0.00431	-0.00143
	(-0.61)	(-0.20)
innovator	0.118	0.0762
	(0.82)	(0.52)
AI citation stock/patents		-0.0789**
		(-2.05)
F citation stock/patents		0.0210
		(1.07)
H citation stock/patents		0.285*
		(1.70)
L citation stock/patents		0.0241
		(0.57)
G04&H06 citation stock/patents		-0.115**
		(-2.53)
Constant	0.704*	0.948**
	(1.87)	(2.39)
Time dummies	Yes	Yes
SIC dummies	Yes	Yes
Country dummies	Yes	Yes
R2 overall	0.579	0.582
Number of obs.	768	768
Number of firms	117	117

t statistics in parentheses
I) Basic model: patents stocks
II) Patent and forward citation stocks
* p<0.1, ** p<0.05, *** p<0.01

Table 4. Estimation results of the Random Effects model for the valuation of firms' personal data related knowledge assets – the role of data giants

Dep. variable: log Tobin's q	(3)	(4)
Log sales	-0.0876	-0.0872
	(-1.71)	(-1.70)
RD stock/assets	0.0553***	0.0543***
	(16.24)	(15.37)
AI patent stock/RD	21.95***	28.33***
	(5.49)	(5.52)
F patent stock/RD	1.527	1.247
	(1.55)	(1.34)
H patent stock/RD	-0.111	-0.331*
	(-1.56)	(-2.37)
L patent stock/RD	54.90***	54.10***
	(7.64)	(7.27)
G04&H06 patent stock/RD	-0.00528	-0.00104
	(-0.73)	(-0.14)
innovator	0.0994	0.0525
	(0.68)	(0.35)
AI citation stock/patents		-0.0830
		(-1.66)
F citation stock/patents		0.0133
		(0.80)
H citation stock/patents		0.359
		(1.42)
L citation stock/patents		0.00846
		(0.18)
G04&H06 citation stock/patents		-0.132**
		(-2.84)

Table 4 continued from the previous page

Dep. variable: log Tobin's q	(3)	(4)
data giant	-0.0967	-1.656**
	(-0.22)	(-2.26)
data giant X AI patent stock/RD	54.25	265.2
	(0.20)	(1.07)
data giant X F patent stock/RD	-133.9	-259.5*
	(-0.79)	(-1.81)
data giant X L patent stock/RD	529.1***	724.8***
	(2.80)	(4.40)
data giant X H patent stock/RD	54.56	331.2
	(0.31)	(1.54)
data giant X G04&H06 patent stock/RD	0.296	1.297**
	(0.75)	(2.44)
data giant X AI citation stock/patents		0.109
		(1.59)
data giant X F citation stock/patents		0.151***
		(5.58)
data giant X H citation stock/patents		-0.314
		(-1.09)
data giant X L citation stock/patents		-0.0400
		(-0.49)
data giant X G04&H06 citation stock/patents		0.450**
		(2.19)
Constant	0.747*	0.886**
	(1.75)	(1.97)
Time dummies	Yes	Yes
SIC dummies	Yes	Yes
Country dummies	Yes	Yes
R2 overall	0.588	0.593
Number of obs.	768	768
Number of firms	117	117

t statistics in parentheses
I) Basic model: patents stocks
II) Patent and forward citation stocks
* p<0.1, ** p<0.05, *** p<0.01

Table 5. Estimation results of the Random Effects model for the valuation of firms' personal data related knowledge assets: testing investor attention hypotheses

Dep. variable: log Tobin's q	(5)	(6)
Log sales	0.0191	0.000671
	(0.35)	(0.01)
RD stock/assets	0.309**	0.250*
	(2.16)	(1.79)
AI patent stock/RD	99.55*	98.95
	(1.71)	(1.55)
F patent stock/RD	-3.117	-2.444
	(-0.66)	(-0.51)
H patent stock/RD	0.0483	-0.170
	(0.26)	(-0.75)
L patent stock/RD	-30.61	-33.38
	(-0.89)	(-0.80)
G04&H06 patent stock/RD	-0.00290	-0.00265
	(-0.11)	(-0.11)
AI patent stock/RD X SVI_AI	-121.9	-114.1
	(-1.38)	(-1.19)
F patent stock/RD X SVI_F	14.14	11.75
	(0.87)	(0.74)
H patent stock/RD X SVI_H	-0.293***	-0.156
	(-4.81)	(-1.38)
L patent stock/RD X SVI_L	113.9***	109.3**
	(2.74)	(2.29)
SVI_AI	1.000	1.100
	(0.48)	(0.51)
SVI_F	9.850***	9.594***
	(5.19)	(4.90)
SVI_H	-4.806***	-4.652***
	(-5.86)	(-5.68)
SVI_L	2.639*	2.389
	(1.66)	(1.45)
AI citation stock/patents		0.296
		(1.58)
F citation stock/patents		0.0243
		(0.47)
H citation stock/patents		0.230
		(1.60)
L citation stock/patents		-0.220*
		(-1.74)
G04&H06 citation stock/patents		-0.0366
		(-0.66)
AI citation stock/patents X SVI_AI		-0.612*
		(-1.86)
F citation stock/patents X SVI_F		-0.0104
		(-0.08)
H citation stock/patents X SVI_H		-0.00672
		(-0.04)
L citation stock/patents X SVI_L		0.592***
		(2.60)
Constant	-4.851***	-4.504***
	(-4.93)	(-4.64)
Time dummies	Yes	Yes
SIC dummies	Yes	Yes
Country dummies	Yes	Yes
R2 overall	0.600	0.620
Number of obs.	461	461
Number of firms	68	68

t statistics in parentheses
I) Basic model: patents stocks
II) Patent and forward citation stocks
* p<0.1, ** p<0.05, *** p<0.01

Table 6. Estimations results of the Random Effects model: testing the technology salience hypothesis

	Patent stocks			Citation stocks		
	Tech. Salience Percentile			Tech. Salience Percentile		
	≤10%	>10% & <90 %	≥ 90%	≤10%	>10% & <90 %	≥ 90%
Patent stocks / RD				Citation stocks / patents		
AI	0.025	-0.021**	0.004***	0.010	0.010	-0.001
	(0.24)	(-2.38)	(3.67)	(0.06)	(1.13)	(-0.72)
F	-0.150*	-0.078	0.010***	-0.065	-0.003	0.007
	(-1.93)	(-1.01)	(2.58)	(-0.54)	(-0.11)	(1.29)
H	0.344	-0.031**	-0.003**	0.311	-0.015	-0.001
	(1.29)	(-2.00)	(-2.25)	(0.87)	(-0.81)	(-0.13)
L	-0.065	0.098**	0.024***	0.062	0.010	0.011**
	(-0.58)	(2.13)	(15.14)	(0.52)	(0.85)	(2.28)
Constant	0.966*			1.415**		
	(1.77)			(2.27)		
Log sales	-0.081*			-0.097**		
	(-1.76)			(-2.06)		
RD stock/assets	0.055***			0.054***		
	(15.79)			(15.78)		
Innovator	0.153			0.127		
	(1.02)			(0.87)		
Time dummies	Yes			Yes		
SIC dummies	Yes			Yes		
Country dummies	Yes			Yes		
R2 overall	0.598			0.589		
Number of obs.	768			768		
Number of firms	117			117		

t statistics in parentheses
* p<0.1, ** p<0.05, *** p<0.01

	Patent stocks			Citation stocks		
	Tech. Salience Percentile			Tech. Salience Percentile		
	≤15%	>15% & <85 %	≥ 85%	≤15%	>15% & <85 %	≥ 85%
Patent stocks / RD				Citation stocks / patents		
AI	0.004	-0.039	0.004***	0.081	-0.008	-0.002
	(0.04)	(-0.75)	(3.93)	(0.72)	(-0.31)	(-1.24)
F	-0.136	-0.079	0.010**	-0.071	0.009	0.008
	(-1.58)	(-0.97)	(2.52)	(-0.59)	(0.15)	(1.35)
H	0.326	-0.044	-0.003**	0.184	0.023	0.000
	(1.25)	(-0.74)	(-2.24)	(0.63)	(1.61)	(0.08)
L	-0.038	0.132	0.023***	0.023	0.031	0.012**
	(-0.33)	(0.90)	(15.18)	(0.19)	(1.45)	(2.55)
Constant	1.004*			1.301**		
	(1.84)			(2.41)		
Log sales	-0.084*			-0.095**		
	(-1.87)			(-2.01)		
RD stock/assets	0.055***			0.054***		
	(16.25)			(16.12)		
Innovator	0.156			0.135		
	(1.02)			(0.94)		
Time dummies	Yes			Yes		
SIC dummies	Yes			Yes		
Country dummies	Yes			Yes		
R2 overall	0.597			0.591		
Number of obs.	768			768		
Number of firms	117			117		

t statistics in parentheses
* p<0.1, ** p<0.05, *** p<0.01

Table 7. Estimations results of the Random Effects model: testing the prospect theory hypothesis

	Patent stocks			Citation stocks		
	Tech. Salience dummies			Tech. Salience dummies		
	≤10%	>10% & <90 %	≥ 90%	≤10%	>10% & <90 %	≥ 90%
Patent stocks / RD				Citation stocks / patents		
AI	-0.159	-0.197	0.050	0.201	0.249	0.023
	(-0.89)	(-1.36)	(0.47)	(1.06)	(1.31)	(0.22)
F	0.059	0.000	-0.011	0.089	0.000	0.158**
	(0.58)	(.)	(-0.04)	(0.77)	(.)	(2.31)
H	-0.289	0.000	-0.294*	-0.257	0.000	-0.136
	(-1.12)	(.)	(-1.81)	(-0.96)	(.)	(-0.84)
L	0.017	0.000	0.050	-0.128	0.000	0.219
	(0.13)	(.)	(0.60)	(-1.14)	(.)	(1.37)
Constant	1.384**			1.184**		
	(2.47)			(2.09)		
Log sales	-0.080*			-0.095**		
	(-1.69)			(-2.04)		
RD stock/assets	0.055***			0.055***		
	(15.34)			(15.85)		
Innovator	0.183			0.148		
	(1.22)			(1.03)		
Time dummies	Yes			Yes		
SIC dummies	Yes			Yes		
Country dummies	Yes			Yes		
R2 overall	0.574			0.585		
Number of obs.	768			768		
Number of firms	117			117		

t statistics in parentheses
* p<0.1, ** p<0.05, *** p<0.01

	Patent stocks			Citation stocks		
	Tech. Salience dummies			Tech. Salience dummies		
	≤15%	>15% & <85 %	≥ 85%	≤15%	>15% & <85 %	≥ 85%
Patent stocks / RD				Citation stocks / patents		
AI	-0.052	-0.083	0.054	0.142	0.195	0.075
	(-0.40)	(-0.46)	(0.50)	(0.94)	(0.90)	(0.60)
F	0.030	0.000	-0.169	0.085	0.000	0.071
	(0.37)	(.)	(-1.11)	(0.74)	(.)	(0.65)
H	-0.280	0.000	-0.186	-0.432	0.000	-0.388**
	(-1.07)	(.)	(-1.07)	(-1.42)	(.)	(-2.13)
L	0.075	0.000	0.104	-0.204*	0.000	-0.056
	(0.47)	(.)	(0.95)	(-1.94)	(.)	(-0.59)
Constant	1.306**			1.467***		
	(2.34)			(2.64)		
Log sales	-0.090**			-0.090**		
	(-1.99)			(-1.99)		
RD stock/assets	0.055***			0.054***		
	(16.36)			(15.78)		
Innovator	0.212			0.130		
	(1.38)			(0.91)		
Time dummies	Yes			Yes		
SIC dummies	Yes			Yes		
Country dummies	Yes			Yes		
R2 overall	0.577			0.580		
Number of obs.	768			768		
Number of firms	117			117		

t statistics in parentheses
* p<0.1, ** p<0.05, *** p<0.01

Figure 1. Patents published in the USPTO in selected domains among the sampled companies, 2007 – 2014

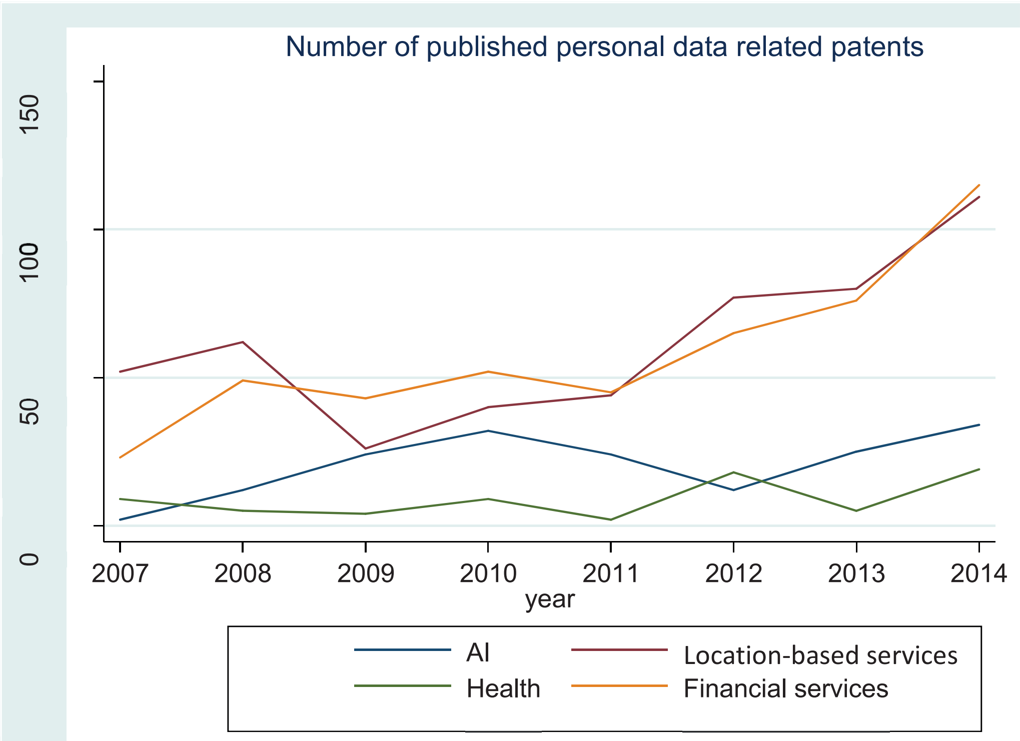
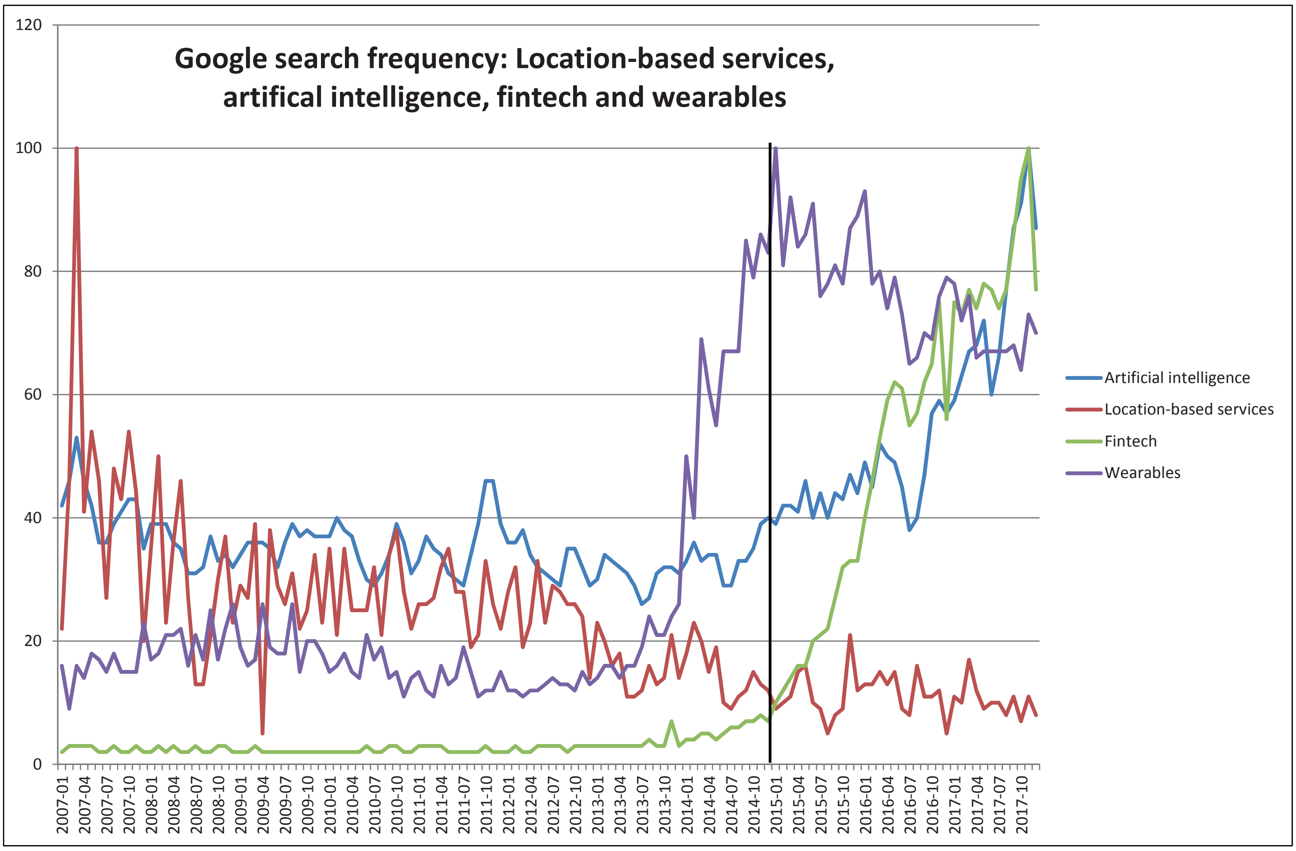


Figure 2. Google search frequency: location-based services, artificial intelligence, fintech and wearables



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