INSTITUTE
OF ECONOMICS

Scuola Superiore
Sant'Anna

LEM | Laboratory of Economics and Management

Institute of Economics
Scuola Superiore Sant'Anna

Piazza Martiri della Libertà, 33 - 56127 Pisa, Italy
ph. +39 050 88.33.43
institute.economics@sssup.it

LEM
WORKING PAPER SERIES

The direction of technical change in AI and the trajectory effects of government funding

Martina Iori [a]
Arianna Martinelli [a]
Andrea Mina [a,b]

[a] Institute of Economics & EMbeDS, Scuola Superiore Sant'Anna, Pisa, Italy.
[b] Centre for Business Research, University of Cambridge, UK.

2021/41 November 2021
ISSN(ONLINE) 2284-0400

The direction of technical change in AI and the trajectory effects of government funding

Martina Iori[a], Arianna Martinelli[a], and Andrea Mina[a,b]

[a]Scuola Superiore Sant'Anna & EMbeDS, Institute of Economics Piazza Martiri della Libertà, 33, 56127 Pisa, Italy
[b]Centre for Business Research, University of Cambridge, Trumpington Street, Cambridge, CB2 1AG, UK

Abstract

Government funding of innovation can have a significant impact not only on the rate of technical change, but also on its direction. In this paper, we examine the role that government grants and government departments played in the development of artificial intelligence (AI), an emergent general purpose technology with the potential to revolutionize many aspects of the economy and society. We analyze all AI patents filed at the US Patent and Trademark Office and develop network measures that capture each patent's influence on all possible sequences of follow-on innovation. By identifying the effect of patents on technological trajectories, we are able to account for the long-term cumulative impact of new knowledge that is not captured by standard patent citation measures. We show that patents funded by government grants, but above all patents filed by federal agencies and state departments, profoundly influenced the development of AI. These long-term effects were especially significant in early phases, and weakened over time as private incentives took over. These results are robust to alternative specifications and controlling for endogeneity.

Keywords: R&D; Technical change; Government subsidies; Technology policy; General purpose technology
JEL codes: O31; O33; O38; D85

1 Introduction

Innovation is a fundamental driver of economic growth (Aghion & Howitt, 1992; Grossman & Helpman, 1991; Nelson & Winter, 1982; Romer, 1990). Because of market failures in the production of knowledge that underpins technical change (Arrow, 1962; Nelson, 1959), governments have played an important role in designing appropriate incentives and in supporting R&D activities in the economy (Bloom et al., 2019). Yet, as argued by Azoulay et al. (2019), in the literature more contributions have focused on firm R&D investments and their spillover effects, than those that have addressed the impact of public funding. Interest in this topic has grown considerably over the last few years. The need to address complex societal challenges, for which uncoordinated private investments in new technologies might be insufficient, has been among the causes of this recent scholarly interest (Mazzucato, 2015; Van Reenen, 2020).

Studies that have focused on the role of government include analyses of the rate of returns of R&D investments (Hall et al., 2010), and policy evaluations of the effects of R&D subsidies (Akcigit et al., 2018; Bloom et al., 2002; Dechezlepretre et al., 2016; Wilson, 2009) and of government grants (Bronzini & Iachini, 2014; Howell, 2017; Santoleri et al., 2022) on private innovation outcomes. Despite the heterogeneity of results found in this literature, these studies try to quantify the impact of public funding on the rate of technical change in the economy. Systematic assessments of the role of government funding on the direction of technical change have proved more difficult. We have historical evidence of the deep influence that governments had in shaping science and technology efforts during times of war and crisis (Gross & Sampat, 2020; Mowery, 2010; Ruttan, 2006). We also have a stream of contributions on the specific role of government in funding breakthrough biomedical research, recently reviewed in Azoulay et al. (2019). All in all, however, quantitative studies of the impact of government funding on the *direction* of technical change are very rare.[1] Behind the scarcity of systematic evidence on this issue there is also the tendency of government to intervene early in the R&D process: public funding plays a role in the development of fundamental knowledge that is typically quite far from having immediate applications, and we can neither detect its effect through short-term market outcomes, nor gauge its impact in the long run, when there may be market outcomes, but these cannot be easily connected with early public investments (Griliches, 1992). It is therefore especially difficult to assess the impact of government investments in technologies with very long lead development times.

One salient characteristic of technical change is its *cumulativeness* (Dosi, 1982; 1988; Green & Scotchmer, 1995; Henderson & Cockburn, 1996; Sampat & Williams, 2019; Scotchmer, 1991). New knowledge builds on prior knowledge, often in a recombinatory way (Kaplan & Vakili, 2015; Weitzman, 1998; Wuchty et al., 2007), to generate new solutions to problems, which in turn open up opportunities for further development. Knowledge accumulation in science and technology is a process that involves different individuals, organizations and institutions. In its essence it is an evolutionary process that over time should select in more useful and valuable knowledge, on which further knowledge will be built, and select out less valuable or obsolete knowledge. Dosi (1988) conceptualized the broad patterns of cumulative change as *technological trajectories* that emerge over time and can be viewed in retrospect as the path-dependent outcome of dispersed research efforts converging into particular ways of solving problems. In this paper, we develop the idea that the government can play a fundamental role in directing technical change and influencing the patterns of knowledge accumulation.

We focus on the long-term development of Artificial Intelligence (AI). AI research encompasses knowledge and techniques that are designed to make machines 'intelligent', in the sense that they can function in the environment where they are applied also through foresight (Nilsson, 2010). The idea that human intelligence can be 'mechanized' is not so recent, but it is over the last few decades, with the development of modern AI, that computing technologies and machine learning have allowed to achieve unprecedented results and have opened up multiple prospects of commercial application. Even though AI includes many different research areas, it is possible to identify among its core components machine learning, deep learning, NLP (natural language processing) platforms, predictive APIs (application programming interface), image recognition and speech recognition.

Following a well-established tradition, we use patents as indicators of innovation activities (Hall et al., 2001). However, we depart from the literature using patent citation counts as measures of impact (Hall et al., 2005; Trajtenberg, 1990), and also from a standard 'spillover' framework of analysis (Bloom et al., 2013; Griliches, 1992; Jaffe, 1986; Jaffe et al., 1993). Based on network theory, in this paper we measure the long-term effect that discrete inventions have on the main technological trajectory of AI development. We download from the US Patent and Trademark Office (USPTO) 114,670 AI patents. We identify patents

[1] By 'direction' here we do not refer to biases in technical change that favor the use of one particular factor of production over another (Acemoglu, 2002), but rather the long-term orientation of technology development in a knowledge search space (Dosi, 1988).

that recognize receipt of government finance in their funding acknowledgements, as well as patents filed by government agencies and state departments. We then put to empirical test the conjecture that government-funded patents have an effect on the technological trajectory. We find that these patents had profound effects on the cumulative development of AI. Patents filed by Federal and state departments and agencies had the strongest impact. Moreover, the effects of government funding were especially significant in early phases of technology development, and weakened over time as private incentives took over. While our empirical settings and variables ensure a low risk of reverse causality and citation bias issues, our sample might bias as public investments may target research areas with the most potential for follow-on innovation (Azoulay et al., 2019). We control for this possibility using a quasi-experimental design based on both propensity-score matching and instrumental variable, and our general results hold.

The paper aims to make three contributions. Firstly, we provide novel and original evidence on the influence of government funding on the direction of technical change. Secondly, we contribute to the development and application of a novel way to measure the effect of innovation on follow-on technological developments. Thirdly, we contribute to the emergent literature on the economics of artificial intelligence by providing novel quantitative evidence of key financing patterns that have supported the development of these technologies over the last thirty years.

The paper is organized as follows. In section 2 we briefly discuss the recent economic literature on AI. Section 3 presents the data we use in this study. Section 4 details the methodology we apply to identify the technological trajectory and measure the long-term cumulative patterns of technological development in the field. Section 5 shows resulting network and indicators. Section 6 presents the empirical strategy we use to examine the effect of government funding on the AI trajectory, then our results and finally a series of robustness checks. The final section summarizes the findings, discusses the limitations of our work, and draws the contribution to a close.

2 AI as general purpose technology

AI involves "[the automation of] activities that we associate with human thinking, activities such as decision-making, problem-solving, learning…".[2] It has the potential to generate broad spillovers that can go way beyond the boundaries of information and communication technologies, and open up further scientific, technological and economic opportunities in several domains. AI is a likely candidate as the dominant general purpose technology of the coming era (Cockburn et al., 2018). General purpose technologies (GPT) are groups of techniques and applications associated with deep transformations in economic systems (Bresnahan & Trajtenberg, 1995; Helpman, 1998; Jovanovic & Rousseau, 2005). Their distinctive characteristics are pervasiveness, high dynamism, and strong complementarities. AI is beginning to display these characteristics, and as a GPT, AI could indeed generate waves of radical innovations leading to widespread economic disruption (Trajtenberg, 2019). AI, especially through the evolution of machine learning, could affect the production of most goods, and the organization and provision of non-routine tasks and services. When subjected to the same empirical tests Moser and Nicholas (2004) used for electricity, artificial intelligence is indeed displaying the emergent characteristics of a general purpose technology (Martinelli et al., 2021).

Because of this transformative potential, scholars have recently developed a strong interest in AI and the effects of its diffusion (Agrawal et al., 2019; Felten et al., 2021; Jacobides et al., 2021; Krakowski et al., 2022). The largest share of research has focused on the effect of automation on productivity growth and

[2]From: Bellman, R. (1978). *An introduction to artificial intelligence: can computers think?*. Thomson Course Technology.

employment. While AI will probably increase productivity in the long run (Furman & Seamans, 2018), there is no consensus on its impact on labor. Acemoglu and Restrepo (2018) propose a conceptual framework to evaluate AI's implications for employment. They suggest that, while in the short-run AI will replace a large number of tasks, the creation of new tasks will balance out this effect in the long run. However, this process will be slow, and the pace of change will be constrained by skill mismatches. In discussing the implications for the division of income between labor and capital, Aghion et al. (2017) emphasize that, despite the AI work displacement effect, the labor share might remain substantial if in the future AI-adopting sectors will contribute gradually less to aggregate growth, which is hard to improve (Baumol, 1967).

Despite these caveats, there is widespread concern about the increase in poverty and inequality that could be due to the rapid diffusion of human-replacing innovations (Furman & Seamans, 2018), at least in the short run. Developing countries could also suffer from the rise of capital share of GDP due to AI adoption (Korinek & Stiglitz, 2021). To mitigate the negative consequences of AI diffusion, Korinek and Stiglitz (2019) propose a rise in capital taxation and intellectual property rights reforms. Trajtenberg (2019), instead, highlights the government's role in designing innovative strategies to reform education, support personal services, and direct technical change towards human-enhancing innovations.

Beyond the direct effect on growth and labor, AI may affect businesses in several other ways. Firstly, AI may alter the sources of competitive advantage by changing the balance between substitution and complementation in managerial tasks (Krakowski et al., 2022). Secondly, it might change the innovation process itself (Cockburn et al., 2018). On the one hand, the introduction of AI in technology production can foster the growth of new ideas, enhancing innovation (Aghion et al., 2017). On the other hand, machine learning is likely to be "an invention of a method of inventing", as Griliches (1957) observed in the case of hybrid corn (p. 502). It is also worth mentioning that successful application of AI requires the use of a large amount of data, especially for predictions and decision-making. This feature raises security and privacy concerns, and might have a profound impact on industrial structure (Varian, 2018).

All the factors and dimensions that have been highlighted in the emergent field of the economics of AI are a clear indication of the potential for *paradigmatic* change (Dosi, 1982) of this technology, precisely the kind of *paradigmatic* change associated with GPTs. Against this backdrop, AI is an ideal context of analysis to explore the role played by government funding for technologies that require very long lead-development times. It can be argued that the role of government is especially important for modern GPTs because they are technically and commercially very risky — or more precisely characterised by fundamental uncertainty in 'Knightian' terms — and are therefore either very costly or simply impossible to finance by means of private funds, given the uncertainty and time-horizons of returns. In order to explore this issue, we take a long-term evolutionary perspective on the problem of knowledge and innovation. We build on the theoretical notion of a *technological trajectory* (Dosi, 1988) to capture the impact of patents on the longitudinal network of follow-on inventions, and investigate the effect of government funding on the direction of technical change.

3 Data

In this section, firstly, we describe our data sources and the criteria we used to identify AI patents. Secondly, we discuss how we detected patents related to US government funding, and we provide some descriptive statistics.

3.1 Sample construction

Patent data are widely accepted and used as proxies for innovation activities (Griliches, 1990; Hall et al., 2001). Over the years scholars have been very active in developing patent indicators to highlight different characteristics of the disclosed invention, such as patent value (Lanjouw et al., 1998; Trajtenberg, 1990), patent technological breadth (Lerner, 1994), legal scope (Kuhn & Thompson, 2019), and patent generality and originality (Trajtenberg et al., 1997).

Our analysis uses patents granted by the USPTO from 1976 to 2019 and related to AI. We retrieved these data from the EPO-PATSTAT database (Autumn 2019 version). We selected AI inventions by combining the procedures suggested by the World Intellectual Property Organization (WIPO) report on artificial intelligence (WIPO, 2019) and the United Kingdom Intellectual Property Office (UKIPO) report on great technologies (UKIPO, 2014). These criteria combine the selection of specific technological classes with a text-based search of technical keywords on patent titles and abstracts. Due to the focus on an evolving domain, the integration of keywords, based on an extensive review of the literature, is important to refine searches and capture trends that do not fit perfectly in a consolidated classification system of technological domains.[3] We rely on the recent and highly detailed Cooperative Patent Classification (CPC) system to select AI-related technology classes, and we include, among others, group Y10S 706 (i.e. *Data processing: artificial intelligence*) and several subclasses of the class G06 (i.e. *Computing; Calculating; Counting*).[4] Regarding keywords, our list includes the expected *artificial intelligence* and a variety of machine learning methodologies and tools for big-data management.[5]

This patent selection process results in a total of 118,949 patents. We use backward citations to generate a network of patents linked to one another. By excluding 1) components of the network that are not connected to the core largest component, thus removing marginal or irrelevant patents (i.e. patents that are never cited or patents accidentally included in the selection process), and 2) citations that violate the time constraints (e.g. references whose earliest publication date *follows* the earliest publication date of a citing patent), we retain 96.42% of the selected patents and obtain a final sample of 114,670 AI inventions.[6]

These patents span the entire period of analysis and mainly belong to computer technologies, including audiovisual technologies and digital communication, and control systems engineering with applications in transport and medical technologies. In particular, most patents concern pattern recognition in images and texts, image analysis, speech recognition, position control (e.g. automatic pilot), and data processing in general.

The leading assignees are well-known information and communication technology companies. The field was initially characterized by a race between the largest American and Japanese technology companies. The list includes, on the one hand, International Business Machines Corporation (IBM), AT&T Corporation (Bell Laboratories), Boeing Company, United Technologies Corporation, and Texas Instruments Incorporated, and on the other, Hitachi, Sharp, and Toshiba. Among these companies, AT&T and especially IBM are important players throughout the entire period of analysis. In the 90s, we also observe the emergence of new companies, such as Xerox and Microsoft among leading assignees. In particular, Microsoft becomes the dominant player, together with IBM, at the beginning of the new century. The last decade is instead characterized by the

[3] See Baruffaldi et al. (2020) for a detailed discussion on this.
[4] See Appendix A for the detailed list of CPC subclasses.
[5] See Appendix A for the detailed list of keywords used for the sample selection.
[6] In directed graphs, a weakly connected component is the maximal sub-graph in which each pair of nodes is connected when one ignores the edge direction. By considering only the largest community of connected inventions from our sample, we also ensure the elimination of noise generated by the inclusion of patents that, although captured through the keyword searches, are not part of the core field.

increasing relevance of the following generation of world-leading IT companies, including Google, Amazon, Apple, Samsung, and Intel.

Appendix B provides an overview of our data.

3.2 Government funded patents

A focus on the US is justified by the strong interests in AI developed within the US innovation system, the active role played by the US government in this space, and the availability of information on US government funding in the data.

Following the literature (Fleming et al., 2019), we exploit two kinds of information to detect patents directly supported by US government funding.[7] Combining the disambiguation efforts of the EPO-PATSTAT database and the USPTO[8] on assignee and applicant categories, we identify patents assigned to federal agencies, national laboratories, and state departments. Among AI patents, we find 929 patents assigned to one of those organizations. As shown in Table 1, the Department of Defense – with its Navy, Army, and Air Force divisions – supported the large majority of these inventions. The other important player is the National Aeronautics and Space Administration (NASA).

Assignee	Number of patents	%
Secretary of the Navy	370	39.83
National Aeronautics and Space Administration	153	16.47
Secretary of the Army	109	11.73
Secretary of the Air Force	106	11.41
Department of Energy	33	3.55
National Security Agency	29	3.12
Department of Health and Human Services	29	3.12
United States Postal Service	22	2.37
Lawrence Livermore National Security	19	2.05
Department of Commerce	10	1.08

Table 1: Most frequent US federal agencies, national laboratories, and state departments as assignees of AI patents.

The second source of information on government funding is the Government Interest Statement in patent texts as reported by the USPTO. Since 1980, the Bayh-Dole Act has allowed contractors to retain ownership of inventions developed with federal funding. In return, it obligates applicants to disclose a government interest in their patents. As reported in Table B6, the main federal contractors include both companies that are leading assignee in the field (such as IBM) and universities.

In our sample, 3597 patents acknowledge government funding through a Government Interest Statement. Interestingly, some of them were granted before 1980, although the statement was not yet mandatory at the time. Even in this case, as highlighted in Table 2, the Department of Defense is, by far, the primary supporter of AI research. Besides, a significant fraction of patents does not correctly specify the funding agency, but

[7]Fleming et al. (2019) identified US patents relying on federal support in three ways: patents owned by the US government, patents acknowledging support from the US government, and patents that directly cite a patent or scientific paper that meets one of the first two criteria. To better identify the effect of government funding, we do not include the latter category in our definition of patents relying on government funding. However, we tested the robustness of our results to the inclusion of indirect government funding (see Section 6.4).

[8]To retrieve data on government funding, we combine EPO-PATSTAT database with the Patentsview database from the USPTO.

refer instead to the United States Government in general. Other important sponsors are the Department of Health and Human Services, the National Science Foundation, and the Department of Energy.

Federal agency	Number of patents	%
Department of Defense	1670	46.43
United States Government	703	19.54
Department of Health and Human Services	627	17.43
National Science Foundation	478	13.29
Department of Energy	462	12.84
National Aeronautics and Space Administration	166	4.61
Small Business Administration	42	1.17
Department of Transportation	36	1.00
Department of Commerce	36	1.00
Department of Homeland Security	35	0.97

Table 2: Most relevant federal agencies that provide funding for supporting the development of AI patents by federal contractors (private companies and universities).

4 Measuring trajectory effects

We exploit the connections through citations between patents to track the development of technological trajectories, and to examine the role of specific inventions in shaping these trajectories. Patent citation networks are a meaningful analytical tool to identify technological trajectories. Each series of patents linked through citations identifies chains of local, cumulative, and irreversible technological developments, consistently with the definition of technological trajectories provided by the literature (Dosi, 1982; 1988). Why do we need a network approach to identify those patents that shaped the direction of technical change? Let us take as an example the introduction of systems based on probabilistic learning, specifically Hidden Markov Models, in speech recognition research in the 80s. As a major advance in statistical modelling, this is unquestionably a milestone in the development of AI because, among other things, it laid the foundations for the autonomous speech recognition applied in current virtual assistants.[9]

We can identify two patents corresponding to this breakthrough: patent US4587670A filed by AT&T Bell Laboratories and patent US4718094A by IBM. Even though these two inventions were crucial for the future development of the field and for modern AI applications, if we looked at the number of citations to identify their importance, with less than 200 in thirty years, they would not appear to be particularly important. Indeed, citation counts do not capture cumulative effects on follow-on innovations and do not provide strong signals on the direction of technical change. What we need is an indicator able to capture the influence of each patent on the long-term development of the field, i.e. a measure of relevance in a technological trajectory.

Below, we describe how we compute trajectory indicators from a patent citations network.

The citation network built on AI inventions and their references is a large graph with 514,599 nodes and 2,661,528 edges.[10] Since citations respect the time flow, there are no loops, and the network is a Directed

[9]For a history of speech recognition, see Juang and Rabiner (2005), *Automatic Speech Recognition – A Brief History of the Technology Development.*

[10]As the AI patents might be connected through citations to patents not related to AI, to better track field development, we also include these "non AI patents". These "non AI patents" are 399,929, and they are only included in the computation of technological trajectories but not in the econometric exercise.

Acyclic Graph (DAG). In this kind of graph, we can sort nodes in topological order, and it is possible to clearly define paths from sources to sinks without encountering each node more than once. To ease the interpretation of this citation network, we define edge direction following the knowledge flow. In this configuration, sources include early patents in AI or prior art that do not belong to the domain, while sinks are the most recent patents in our sample.

More formally, we can interpret the citation network as a graph $\mathcal{N} = (\mathcal{P}, \mathcal{R})$, where $\mathcal{P}$ is our set of patents and $\mathcal{R} \subseteq \mathcal{P} \times \mathcal{P}$ represents the following citing relation: $u\mathcal{R}v \equiv v$ *cites* u. $\mathcal{R}$ is irreflexive and acyclic, and the same applies to the inverse relation $\mathcal{R}^{inv}$, defined as $u\mathcal{R}^{inv}v \equiv u$ *cites* v. In this second case, the direction of edges is from citing to cited patents. Let us, also, define a function $R(p)$ that maps each patent p with its set of successors in the graph based on the relation $\mathcal{R}$: $R(p) = \{u \in P : p\mathcal{R}u\}$.

The properties of $\mathcal{R}$ (and $\mathcal{R}^{inv}$) make $\mathcal{N}$ a DAG with the following special features.

- Nodes can be sorted by topological order, meaning that a map between nodes and cardinal numbers (the node order) $f : \mathcal{P} \longrightarrow 1 \ldots |\mathcal{P}|$, such that $u\mathcal{R}v \Longrightarrow f(u) < f(v)$, exists.
- It is possible to define sets of minimal and maximal elements as, respectively, $Min\,R = \left\{p \in \mathcal{P} : R^{inv}(u) = \emptyset\right\}$ and $Max\,R = \{p \in \mathcal{P} : R(u) = \emptyset\}$. They represent the list of sources and sinks of the network.
- By definition, every node $p \in \mathcal{P}$ and every edge $(u, v) \in \mathcal{R}$ belong to at least one path between $Min\,R$ and $Max\,R$.

An easier representation of the network is its standard form $\mathcal{N}' = (\mathcal{P}', \mathcal{R}')$, where all patents in $Min\,R$ cites a single source s and all nodes in $Max\,R$ are cited by a single sink t. In this case, the set of patents is $\mathcal{P}' := \mathcal{P} \cup \{s, t\}$, and the citation relation is $\mathcal{R}' := \mathcal{R} \cup \{s\} \times Min\,R \cup Max\,R \times \{t\} \cup \{t, s\}$.

In a graph that has these characteristics, we can measure the significance of each edge in the network based on a connectivity indicator, such as the traversal count (Hummon & Dereian, 1989). Among the several possible definitions of traversal counts, in our analysis we follow Batagelj (2003) and use the Search Path Count (SPC). The SPC assigns to each edge (u, v) a weight w_{uv} equal to the number of paths from each s (source) to each t (target) through (u, v). In other words, it measures the number of paths in the network through a given edge. The higher the weight, the more important the edge is for network connectivity and the development of the entire technological domain.

Given an edge (u, v), the computation of w_{uv} proceeds in three steps. Firstly, we compute the number of paths w_u^- between the source s and the cited patent u. Secondly, we assign the number of paths between the citing node v and the sink t to w_v^+. The SPC weight w_{uv} is then the product between the two quantities:

$$w_{uv} = w_u^- * w_v^+. \tag{1}$$

In a DAG, where a topological order of nodes exists, we can compute the partial weights w_u^- and w_u^+ with a recursive procedure:

$$w_u^- = \begin{cases} 1 & u = s \\ \sum_{v:v\mathcal{R}u} w_v^- & \text{otherwise} \end{cases}, \qquad w_u^+ = \begin{cases} 1 & u = t \\ \sum_{v:u\mathcal{R}v} w_v^+ & \text{otherwise} \end{cases}. \tag{2}$$

Early explorations of this methodology (see for instance Martinelli, 2012; Mina et al., 2007) used traversal counts associated with each edge of the citation network to identify the most relevant trajectories in small technological domains. These trajectories are the paths across the network (from s to t) with the highest total weight W_M, where M is the set of the edges in the path and $W_M = \sum_{(u,v)\in M} w_{uv}$.[11] The edge sequence M with the maximum total weight W_M, is therefore the longest path when considering the weighting w_{uv} of each edge. To better describe the evolution of a domain, we can also consider paths with slightly lower weighted path length. To provide an idea of the computation-intensive nature of this exercise, it is worth noting that the AI patent citation network has $3.2 * 10^{19}$ possible paths, and the longest path is equal to $1.7 * 10^{20}$. We will use this approach in Section 5 to identify the most relevant technological trajectories in AI.

While SPC weights are commonly used to trace technical change dynamics, they can also measure the relevance of single inventions from a trajectory perspective. Following Batagelj (2003), we extend the standard definition of SPC weights – which usually apply to edges – to the nodes of our citation network:

$$\widetilde{w}_p = w_p^- * w_p^+. \tag{3}$$

This measure indicates the number of paths from s to t through the patent p. A patent with a high weight is a patent that "cumulates" a large knowledge flow within the network. This indicator has considerable advantages over simple citation count. The citation count, which in this framework would correspond to nodes' outdegree,[12] would be local in nature. On the contrary, the trajectory indicator summarizes complex citation chains and captures the invention's influence on the evolution of an entire field, rather than on close patents only. Measuring the trajectory effects provides valuable information on which inventions have the strongest influence on the direction of technical change as a whole. Therefore, we will use this measure as a primary indicator of patent impact in our econometric analysis (see Section 6). The correlation matrix reported in Table B5 confirms the difference between the trajectory indicator and the number of citations.

Another relevant indicator connected to the network structure is the node position in the graph. The patent position in the citation network is a more precise indicator of timing than the patent application year because it marks time in terms of the patent citation network and therefore in terms of the overall evolution of the field. In particular, the $timing_p$ measures the patent p distance from network sources and is defined in a recursive way:

$$\text{timing}_p = \begin{cases} 0 & p \in Min\,R \\ 1 + \max_{n:n\mathcal{R}p}(\text{timing}_n) & \text{otherwise} \end{cases}. \tag{4}$$

In other words, the timing takes value 0 for network sources and, for all the other patents, it is equal to 1 plus the maximum timing of their cited patents. Intuitively, the timing's low values refer to the early stages of the technology (i.e. closer to sources), while high values indicate innovations in a mature phase (i.e. closer to sinks).

5 Technological trajectories: the evolution of artificial intelligence

To provide an overview on AI inventions and invention chains, first of all we present the most relevant technological trajectories in the field. We are interested in capturing the evolution of the entire field, and for illustrative purposes we include also AI patents granted before 1976.[13]

[11] Alternative definitions of main paths – from a local perspective – produce overall similar results.

[12] The centrality degree is defined as the number of links incident upon a node. As this network is directed, we can distinguish between two types of degree centrality measures. Indegree is a count of the number of ties directed to the node, and outdegree is the number of ties that the node directs to others. As the directionality of our network follows the potential "knowledge flow", if a patent receives three citations it will direct three ties to three nodes.

[13] We obtain a citation network with 555,454 nodes – among which 122,052 AI patents – and 2,754,878 edges.

We weigh each edge by its SPC weight, as defined in Equation 1. We then compute the total SPC weight associated to each path. Figure 1 shows the main paths for AI inventions extracted from the patent citation network described in the previous section. We include nodes belonging to the longest path (i.e. with the maximum total SPC weight), which are the red nodes, and nodes belonging to paths with a total SPC weight that is, respectively, up to 1.5% and 3% lower than the longest path. The latter are the orange and yellow nodes, respectively. Detailed information on patents are in Appendix C. There are several ways to validate this methodology, and, in this work, we rely mainly on two. The first one is by looking at the technologies disclosed in the patents on the trajectories to check how they cover technological milestones in the domain. The second one is by looking at the correspondence between known major firms, institutions, and inventors and patent assignees and inventors on the trajectories.[14]

The densest part of the AI citation network captures the evolution of speech recognition from the mid-1970s to this day. Indeed, speech recognition is one of the main fields in the AI patent sample (WIPO, 2019). It has a 70-year-long history, and its development mirrors the one of the entire AI domain. Early techniques relied on knowledge-based systems, until, from the mid 1980s onwards, probabilistic learning and the rediscovery of neural networks revolutionized the field. After an exploratory phase characterized by the development of different statistical techniques, research moved on the one hand toward multimodal and integrated systems, and on the other toward the use of big data and hardware enhancement. As far as applications are concerned, in recent years we observe an increase in virtual assistants, search engines, and social networks, whereas a more theoretical focus has clearly been on deep learning applications.

These different phases of research emerge when we inspect the patents included in the main path. Speech recognition research started in the early 1950s at Bell Laboratories and is an evolution of studies on optical character recognition. In the following twenty years, templates and keyword spotting methods were the dominant approaches, and the leading players were Bell Laboratories and Japanese companies, especially the Nippon Electric Corporation (NEC). The main path effectively captures all these developments. The earliest inventions concerned word recognition and dictionaries, and involved the leading figures of the time in speech recognition research: Hiroaki Sakoe – inventor of the continuous speech recognition at NEC –, Stephen Levinson – head of the linguistic research at Bell Labs –, and Lawrence Richard Rabiner – also at Bell Labs, and holder of several IEEE awards for outstanding achievements in signal processing and speech/audio recognition –.

The subsequent phase started with a change in the underlying technique and logic, with a shift toward probabilistic learning and more rigorous statistical models. This shift is also detected in the main path, which includes in the mid 1980s the two breakthrough patents on Hidden Markov Models by Bell Labs (patent US4587670A, number 49 in Figure 1) and IBM (patent US4718094A, number 66 in Figure 1), to which we have already referred in Section 4.

In those years, IBM, led by Lalit Bahl and Fred Jelinek – awarded with the IEEE James L. Flanagan Speech and Audio Processing Award –, becomes Bell Labs' main competitor on speech recognition research in the US. Other companies, such as Dragon Systems (among which patents number 88, 94, and 95 in Figure 1)– founded by James and Janet M. Baker –, helped the commercial diffusion of the first speech recognition software programs, which, at that time, were mainly aimed to call centers. In the following years, the

[14]While the definition proposed in this paper tackle the main path analysis from a global perspective, one can define technological trajectory from a local point of view. In the literature on main path analysis, several procedures exist: (1) starting from sources and moving forward by following the edges with the highest SPC weight, (2) starting from sinks and moving backward by following the links with the highest SPC weight, and (3) starting from edges with the highest SPC weight and moving backward and forward by following the same criterion. In our case, all approaches lead to the same main technological trajectory, which lend robustness to our results.

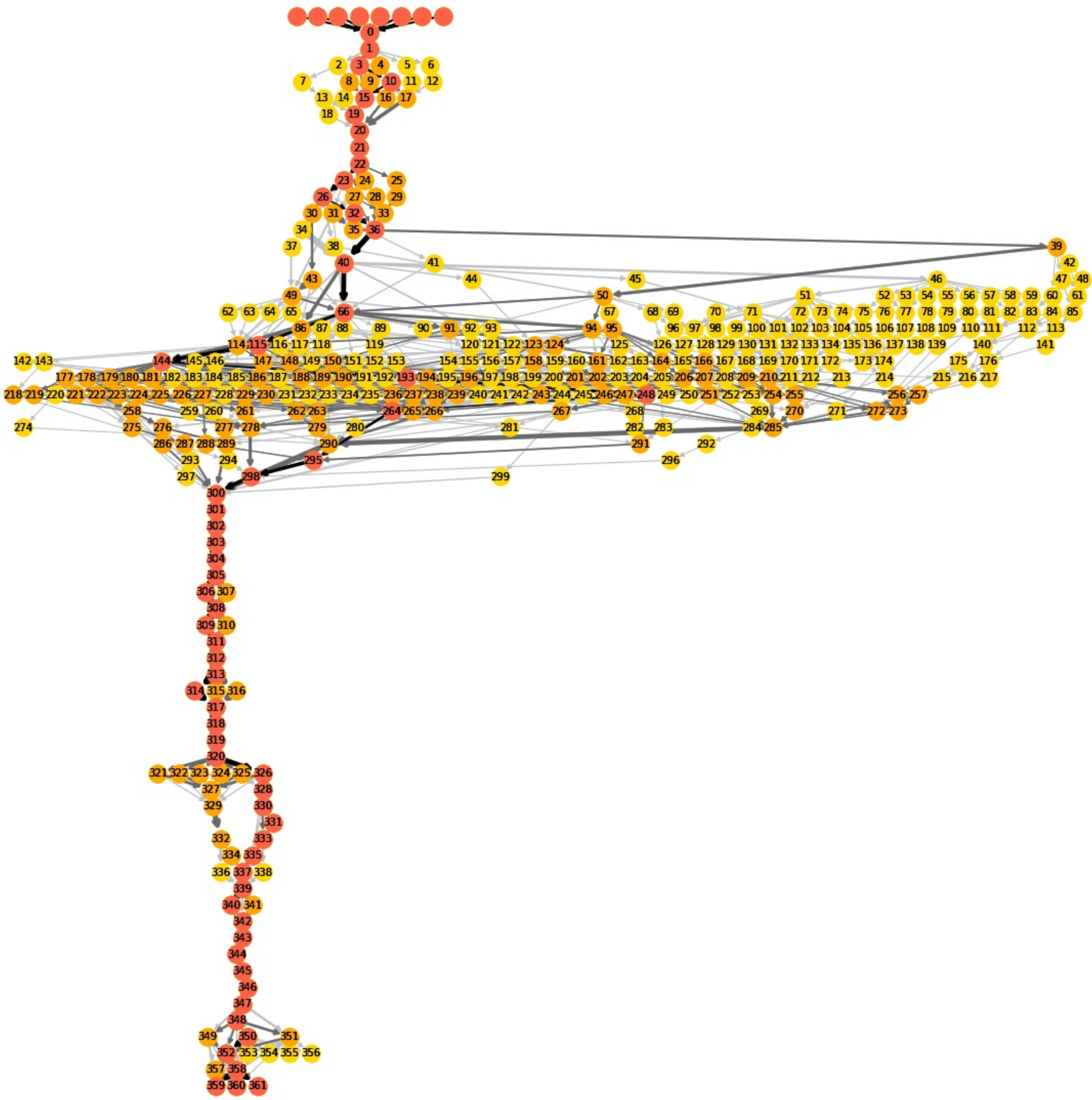


Figure 1: Main trajectories in AI patents. Red nodes belong to the longest path (with the maximum total SPC weight). Orange and yellow nodes are part of paths with a total SPC weight that is, respectively, up to 1.5% and 3% lower than the longest path. The same holds for black, gray, and light-gray edges. The width of each edge (u, v) is proportional to its SPC weight w_{uv}. Detailed information on patents are in Appendix C.

research focused on developing incremental improvements of statistical methodologies and the use of large vocabulary. At the same time, we see a growing interest in grammar, semantics, and translation that pushed the field in the direction of natural language processing.

From the year 2000, we observe the convergence of the main trajectories, and the concentration of leading technology companies such Microsoft, Amazon, Google, Apple, and Facebook, on a single path. Nuance Communications, a major voice recognition developer that acquired IBM and Xerox speech recognition divisions and patents, is the only outsider. This convergence started with Microsoft's patents on Microsoft Speech Server (patent numbers 300 – 307, among which US8229753B2), a milestone towards web-based speech-recognition applications that integrate phones in the standard IT architecture. The R&D efforts that followed focused on multimodal applications and integration in search engines, and were mainly carried out by Nuance Communications, Amazon (by Alexa's 'father', Igor Roditis Jablokov), and Google (Google Voice Search app: patent US8626511B2, number 337). These inventions were the precursors of a clear breakthrough in AI research: the development of intelligent automatic assistants. The first patents covering this development on the main path are Apple's patents related to Siri (numbers 339 and 343 in Figure 1, among which the patent US9117447B2 is the continuation, concerning speech recognition, of the patent *Intelligent automatic assistant* – US9318108B2 –) and developed by the former Stanford Research Institute (SRI) International team. The following patent also covers the well-known speech recognition application that is Amazon Echo (patent US9548066B2, number 344 in Figure 1). In the final phase of the speech recognition trajectory, companies' – primarily Facebook's – efforts focused on applications in multimedia language contexts and predictions of future translations with the support of deep learning techniques.

After sketching the main technological trajectory as revealed by this main path analysis, in what follows we use our entire sample of patents – i.e. independently of whether they appear in the illustrative trajectory of Figure 1 – to subject to econometric tests the relationship between our patent relevance indicator and government funding.

6 Empirical strategy and results

First of all, we examine the trajectory effect of government funding. Secondly, we explore heterogeneity in the timing of government-backed patents to test whether this might affect differently the technological trajectory depending on whether inventions are made in the early phase vs. more mature phases of development.

Thus, with p referring to patents and i to indexing fields, we estimate:

$$\begin{aligned} Ln(\text{trajectory}_{pi}) = & \beta_0 + \beta_1 \,\text{government funding}_p \\ & + \beta_2 \,\text{government funding}_p \times \text{timing}_p \\ & + \beta_3 \,\text{timing}_p + \gamma_p + \delta_i + \epsilon_{pi}, \end{aligned} \tag{5}$$

where $trajectory_{pi}$ is the patent relevance indicator $\widetilde{w}_p$ (SPC weight associated to graph nodes) defined in Equation 3, $government\ funding_p$ is a dummy variable indicating the presence of government funding, $timing_p$ indicates the position of the node in the network (see Equation 4) and defines the time evolution of the graph, the γ_p's are a set of controls at the patent level, and the δ_i's capture subfield fixed effects.

Following the literature, the controls in γ_p account for different patent characteristics. First, we include the number of claims as an ex-ante indicator of patent quality. Second, we include the inventors' team size as an indicator of the disclosed invention's complexity. Third, since a non-negligible share of government

Table 3: Influence of government funding on the trajectory. Estimates follow the semi-logarithmic model presented in Equation 5.

	Dependent variable:		
	log(Trajectory)		
	(1)	(2)	(3)
Government funding	1.184***	1.096***	1.959***
	(0.132)	(0.147)	(0.263)
Government funding*Timing			-0.064***
			(0.011)
US university		0.272	0.282*
		(0.166)	(0.166)
Timing	0.503***	0.503***	0.505***
	(0.002)	(0.002)	(0.002)
Number of claims	0.043***	0.043***	0.043***
	(0.002)	(0.002)	(0.002)
Number of inventors	-0.106***	-0.106***	-0.106***
	(0.011)	(0.011)	(0.011)
Intercept	8.594***	8.592***	8.562***
	(0.078)	(0.078)	(0.078)
3-digit CPC	Yes	Yes	Yes
Observations	114,670	114,670	114,670
R^2	0.435	0.435	0.435
Adjusted R^2	0.435	0.435	0.435
Residual Std. Error	7.292	7.292	7.291
F Statistic	3078.115***	3008.006***	2951.426***

Note: All the models are estimated using OLS.
Robust standard errors are reported in parenthesis.
Legend: *p<0.1; **p<0.05; ***p<0.01

funding goes to universities, we also consider a dummy variable that indicates patents with US universities as assignees to assess whether the effect is driven by universities rather than government funding. Finally, the δ_i's capture subfield fixed effects and control for diverse citation behavior in different fields. We measure subfields through the CPC classification at the 3-digit level, excluding the most common subfield (G06) and marginal codes (those that occur in less than 0.2% of patents). All these control variables and patent information have been retrieved from the EPO-PATSTAT Database (Version Autumn 2019), except for the number of claims, which is drawn from the USPTO database (Patentsview).

All models are estimated using Ordinary Least Squared (OLS) with robust standard errors.

6.1 The role of government funding in the AI technological trajectory

Table 3 presents the estimates of Equation 5. Overall, we find that government exerts a positive and significant effect on the trajectory, that is to say that government funding is associated with inventions that have a long-term impact on future developments of the overall field. To correctly interpret the estimates of dummy variable coefficients in semi-logarithmic equations, we follow Kennedy (1981) and we compute the percentage impact of the dummy variable as:

$$g^* = 100 \cdot \left[\exp\left(\hat{c} - \frac{1}{2} V\left(\hat{c}\right) \right) - 1 \right], \tag{6}$$

where $\hat{c}$ is the estimated coefficient and $V\left(\hat{c}\right)$ is its variance.

It follows that patents receiving government funding have, on average, a trajectory indicator 223.9%

higher than the other patents (specification (1) in Table 3). This effect remains positive and significant even when we control for the presence of US universities as assignees (specification (2)). Although the inclusion of US universities in the picture slightly reduces the effect of government funding, the percentile impact of this funding on the trajectory is still substantial (196%).

In specification (3) of Table 3, we also explore the role of timing effects by including an interaction term between timing and the government funding dummy. The negative sign of the interaction term indicates that the variable timing is less relevant for government-funded patents. For each unitary increase of the timing, the impact of government funding on the trajectory is 6.4% less than patents without government funds. Thus, as shown in Figure D1 in Appendix D, government-funded patents have a higher impact on the trajectory at the early stages of technology.[15]

6.2 Government grants vs. government inventions

We now test whether there is any difference between the two sources of government funding and, if there is any, which type of funding plays a more prominent role. We distinguish inventions made and patents filed by federal agencies and state departments from patents developed by federal contractors (see Section 3.2 for the difference). As we have already done in the previous section, we also explore the relevance of timing of government-backed patents on technology evolution.

We estimate the following specification:

$$\begin{aligned} Ln(\text{trajectory}_{pi}) = & \beta_0 \\ & + \beta_1 \,\text{government interest}_p + \beta_2 \,\text{government interest}_p \times \text{timing}_p \\ & + \beta_3 \,\text{government assignee}_p + \beta_4 \,\text{government assignee}_p \times \text{timing}_p \\ & + \beta_5 \,\text{timing}_p + \gamma_p + \delta_i + \epsilon_{ip}, \end{aligned} \tag{7}$$

where $\textit{trajectory}_{pi}$ is the patent relevance indicator, $\textit{government interest}_p$ is a dummy variable equal to 1 for patents that acknowledged a government interest, $\textit{government assignee}_p$ is a dummy variable that is equal to 1 for patents with a federal agency or a state department as assignee, $\textit{timing}_p$ indicates the node position in the network, the γ_p's and δ_i's are, respectively, the set of controls at the patent level and subfield fixed effects.

Table 4 shows the results. We first notice that both types of government intervention have a relevant impact on the trajectory (specifications (1) and (2)). More precisely, patents with a government interest statement (i.e. assigned to federal contractors) have, on average, a trajectory effect that is 164.9% stronger than other inventions. Federal agencies or state department patents have an even stronger effect since the percentage impact is equal to 868.4%. These effects persist even when we consider the two indicators together, and we include a dummy for US university patents as controls (specification (3)). In this case, the percentage impact of patents by federal contractors is reduced to 56.5%, while the impact of patents by federal agencies or state departments is still 633.7%. We can conclude that, although all types of government funding play a role, patents directly assigned to federal agencies or state departments have a stronger influence on the evolution of AI over time.

Concerning the timing of government-backed patents, we observe the higher importance of both types of

[15] It is worth noting that the timing has a low value also for patents in short sequences of inventions that join the strongest trajectories at different stages. This constitutes further evidence that government funding drives the direction of technological change also at sub-trajectory levels of the evolution of the field.

Table 4: Influence of patents with a government interest statement and patents assigned to government assignees on the trajectory. Estimates follow the semi-logarithmic model presented in Equation 7.

	Dependent variable:					
	log(Trajectory)					
	(1)	(2)	(3)	(4)	(5)	(6)
Government interest	0.983***		0.460***	0.999***	0.481***	0.622**
	(0.134)		(0.157)	(0.281)	(0.156)	(0.288)
Government interest*Timing				-0.037***		-0.010
				(0.012)		(0.012)
Government assignee		2.322***	2.050***	1.959***	4.323***	4.233***
		(0.321)	(0.338)	(0.340)	(0.537)	(0.562)
Government assignee*Timing					-0.230***	-0.224***
					(0.030)	(0.031)
US university			0.551***	0.541***	0.551***	0.548***
			(0.168)	(0.168)	(0.167)	(0.168)
Timing	0.503***	0.503***	0.504***	0.505***	0.505***	0.505***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Number of claims	0.043***	0.043***	0.043***	0.043***	0.043***	0.043***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Number of inventors	-0.106***	-0.102***	-0.104***	-0.104***	-0.105***	-0.105***
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Intercept	8.610***	8.597***	8.574***	8.560***	8.555***	8.551***
	(0.078)	(0.078)	(0.078)	(0.078)	(0.078)	(0.078)
3-digit CPC	Yes	Yes	Yes	Yes	Yes	Yes
Observations	114,670	114,670	114,670	114,670	114,670	114,670
R^2	0.435	0.435	0.435	0.435	0.435	0.435
Adjusted R^2	0.435	0.435	0.435	0.435	0.435	0.435
Residual Std. Error	7.294	7.293	7.291	7.291	7.289	7.289
F Statistic	3074.472***	3078.966***	2944.038***	2886.062***	2891.010***	2831.363***

Note: All the models are estimated using OLS.
Robust standard errors are reported in parenthesis.
Legend: *p<0.1; **p<0.05; ***p<0.01

government intervention (via grants or via direct R&D performance by federal agencies or state departments) for patents with a low timing (specifications (4) and (5)), as we already found for the aggregate indicator in Section 6.1. Once again, the effect is more marked for patents with government assignees. In this second case, the positive impact of timing on the trajectory is 23% lower than for other patents. In other words, although the timing has a positive influence on the dependent variable, the presence of a federal agency or state department as assignee strongly mitigates this effect. As shown also in Figure D2 in Appendix D, government-backed inventions are especially influential at an early stage of technology development, primarily when the government is the assignee.

Overall, these results confirm a very important role of government funding in the long-term development of AI, and this role is especially important during early phases of the development of the field. Moreover, the government as assignee exerts an even stronger influence than government as the sponsor of a grant, highlighting the fundamental importance of research and development carried out in federal agencies and state departments towards the inception phase of the AI technological trajectory.

6.3 Robustness checks: potential sources of endogeneity

In this section, we propose two quasi-experimental designs to address the possible sources of endogenity. As Azoulay et al. (2019) observe, it is possible that public investments target research areas that have the strongest potential for follow-on innovation because of increasing opportunities, and it is therefore important to control for this.

Matching The first quasi-experimental design is based on propensity-score matching (Rosenbaum & Rubin, 1983). We identify treated and control groups by comparing differences in pre-existing patents' characteristics and estimating a probability of receiving different sources of government funding (our treatments). The resulting sub-samples will be, therefore, balanced in the observed covariates. Moreover, patents in treated and control groups will have comparable distributions of the probability of being treated. Then, we replicate estimations in Table 3 and 4 on these balanced sub-samples.

We estimate the probability of receiving the treatment (i.e., the propensity score) through a logistic regression on pre-treatment confounding covariates. Following previous studies (Jaffe et al., 1993; Trajtenberg et al., 1997), the confounding covariates used in this exercise are technology classes (3-digit CPC classes) and time (the variable timing, in our case).[16] The resulting propensity score is used as input for the 1-1 matching without replacement (based on nearest neighbor matching) of treated and control patents (see Figure D3 for covariate balance before and after the matching).

Table 5 summarizes the estimates of the impact of government funding on the trajectory in three different sub-samples of patents. Each sub-sample refers to and is used to test for the impact of a different source of government funding: patents that received any kind of government funding (aggregated category) – specifications (1) and (2) –, patents that acknowledge government interests – specifications (3) and (4) –, and patents with a government assignee – specifications (5) and (6) –. These estimations corroborate the results presented in Sections 6.1 and 6.2: government funding positively affects the trajectory of AI patents, especially at the early stage of the technology. Federal agencies and government departments, even more than government contractors, have a crucial role in the development of this technology.[17] Moreover, once

[16] In our framework, timing is a more appropriate and consistent measure of time than application year or grant year. Timing, indeed, captures the specific time evolution of AI's technological trajectory.

[17] Estimations based on exact matching among patents or the use of propensity score as a control in the regression lead to comparable results.

Table 5: Influence of government funding on the trajectory – 1-1 matching without replacement (propensity score)

	Dependent variable:					
	log(Trajectory)					
	(1)	(2)	(3)	(4)	(5)	(6)
Government funding	1.372***	4.849***				
	(0.191)	(0.317)				
Government funding*Timing		−0.259***				
		(0.014)				
Government interest			1.072***	4.433***		
			(0.198)	(0.332)		
Government interest*Timing				−0.242***		
				(0.015)		
Government assignee					2.977***	7.897***
					(0.384)	(0.605)
Government assignee*Timing						−0.493***
						(0.037)
US university	−1.485***	−1.424***	−1.258***	−1.260***	0.852	1.407
	(0.255)	(0.254)	(0.258)	(0.257)	(1.115)	(1.124)
Timing	0.573***	0.702***	0.578***	0.698***	0.535***	0.783***
	(0.008)	(0.009)	(0.008)	(0.009)	(0.022)	(0.023)
Number of claims	0.039***	0.041***	0.045***	0.046***	0.001	0.009
	(0.006)	(0.006)	(0.007)	(0.006)	(0.014)	(0.013)
Number of inventors	−0.162***	−0.157***	−0.129***	−0.126***	−0.231**	−0.223**
	(0.043)	(0.043)	(0.044)	(0.044)	(0.106)	(0.101)
Constant	8.083***	6.305***	7.846***	6.143***	8.567***	5.909***
	(0.245)	(0.248)	(0.251)	(0.256)	(0.525)	(0.506)
3-digit CPC	Yes	Yes	Yes	Yes	Yes	Yes
Observations	7,864	7,864	7,194	7,194	1,858	1,858
R^2	0.449	0.468	0.470	0.487	0.301	0.348
Adjusted R^2	0.446	0.465	0.467	0.484	0.287	0.335
Residual Std. Error	7.244	7.118	7.089	6.975	8.216	7.936
F Statistic	155.571***	164.013***	154.672***	161.664***	21.185***	25.575***

Note: All the models are estimated using OLS on data matched through propensity score matching (1-1) without replacement) Robust standard errors are reported in parenthesis.
Legend: *p<0.1; **p<0.05; ***p<0.01

we control for selection bias, the estimated impact of government funding not only does not disappear, but is even stronger than in the previous OLS estimations.

Instrumental variable We also consider a quasi-experimental design based on the introduction of an instrumental variable. Following Moretti et al. (2020) and adapting their approach to our setting, we use as instrumental variable the predicted number of patents related to defense R&D in the different technological classes (4-digits CPC) that are associated to each patent. While military R&D is one of the most relevant sources of government R&D funding in the US, it is likely driven by geopolitical reasons rather than economic ones (Mowery, 2010). The exogeneity of defense R&D to the long-term evolution of AI makes the number of patents related to defense R&D a very good candidate to instrument the government funding indicators of our empirical analyses. We therefore use the predicted number of patents associated to defense R&D, i.e. the number of defense R&D patents in the year before the patent's year of application, to rule out endogenous components and address residual concerns of endogeneity. More details on the construction of this instrumental variable are in Appendix D.3.

Concerning the relevance of the variable as instrument for government funding, a positive variation of defense R&D funding might have, in principle, a positive or negative effect on the total variation of government funding in a given technological class since defense R&D may drive or substitute for other sources of government R&D funding. The first stage results, presented in Table D1, show that variations in predicted defense R&D drive general government funding, as found also in Moretti et al. (2020). Indeed, the impact of the predicted defense-related patents on government funding, government interest, and government assignee is positive and significant. Moreover, the F-tests performed on the first-stage regressions reject the null hypothesis that the instruments are weak and the instruments have good statistical power.

Table 6: Influence of government funding on the trajectory - Instrumental variable

	Dependent variable:					
	log(Trajectory)					
	(1)	(2)	(3)	(4)	(5)	(6)
Government funding	45.915***	70.948***				
	(3.446)	(6.403)				
Government funding*Timing		−1.899***				
		(0.211)				
Government interest			53.566***	92.888***		
			(4.318)	(9.173)		
Government interest*Timing				−2.625***		
				(0.298)		
Government assignee					102.234***	109.005***
					(11.167)	(13.467)
Government assignee*Timing						−1.190
						(0.866)
US university	−18.951***	−18.412***	−22.288***	−23.599***	1.072***	1.103***
	(1.550)	(1.744)	(1.932)	(2.431)	(0.213)	(0.206)
Timing	0.521***	0.577***	0.517***	0.591***	0.539***	0.543***
	(0.003)	(0.007)	(0.003)	(0.009)	(0.004)	(0.005)
Number of claims	0.046***	0.048***	0.044***	0.046***	0.060***	0.060***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)
Number of inventors	−0.156***	−0.145***	−0.174***	−0.166***	−0.038**	−0.042**
	(0.017)	(0.017)	(0.018)	(0.020)	(0.018)	(0.018)
Intercept	7.062***	6.179***	7.207***	6.000***	6.484***	6.490***
	(0.144)	(0.227)	(0.144)	(0.263)	(0.238)	(0.240)
3-digit CPC	Yes	Yes	Yes	Yes	Yes	Yes
Observations	114,670	114,670	114,670	114,670	114,670	114,670
F-test	181.6***	114.83***	154.4***	106.61***	88.16***	40.98***
F-test (interaction)		91.43***		88.04***		24.65***

Note: All the models are estimated using 2SLS.
Robust standard errors are reported in parenthesis.
Legend: *p<0.1; **p<0.05; ***p<0.01

In Table 6, we report 2SLS estimates. Our results are broadly confirmed: government funding positively affects the long-run trajectory of patents. This result holds both for patents funded through grants and for patents with the government as assignee. As far as the timing of funding is concerned, in these estimates the interaction term between government assignee and timing is not significant, possibly due to the low share of patents directly assigned to federal agencies and government departments, but the effect of government funding at early stages of technology development remains significant, and is consistent with the OLS regressions results (with and without matching).

6.4 Additional robustness checks

Results are robust to a series of variations in the definition of trajectory indicators, government funding, sample composition, and controls. In what follows, we present the key insights, while Appendix D presents these sensitivity analyses in detail.

Trajectory indicators Firstly, we introduce a different measure of inventions' relevance in the trajectory. Instead of considering an indicator of traversal count that detects nodes with the highest knowledge throughput, we assign to each patent the length of the longest weighted path that goes through it.[18] Since patents with the longest weighted path length are those on the main path, this indicator approximates the probability that the patent is on the main trajectory. Differently from the trajectory indicator defined in Section 4, the longest path length summarizes the complex knowledge chain along the entire path and is less reliant on the node. Even if a direct comparison of coefficients is not possible due to the different magnitude of the indicators (see Table B4), results presented in Tables D2 and D3 are fully consistent with the ones discussed in the previous sections. The relative proportion between different effects is also preserved, and patents with a government assignee are, by far, those with the strongest impact on the trajectory.

Government funding Previous work (see, for instance, Fleming et al., 2019) on the role of state investments in fostering innovations use a broader definition of government funding by including in the analysis also patents that cite government-funded inventions. Although we believe that considering only direct investments leads to a more accurate assessment, we replicate our regression analysis by including patents citing government-funded inventions. For the sake of consistency, we also replace the control variable that detects patents with US universities as assignees with an indicator for patents that cite US universities' inventions. The share of patents that is indirectly connected to government funding (27.8%) is significantly higher than the one of patents that directly received this funding (3.1%). However, the estimations presented in Tables D4 and D5 corroborate our main results, also in terms of coefficient magnitude. Specification (1) of the first table shows that citing government funding increases the trajectory effect by 233.0%. Moreover, the impact of citing government-backed patents is stronger for patents with a low timing value (specification (3)). Unsurprisingly, the interaction term's magnitude is slightly lower than the one in Table 3. Patents that cite government-funded inventions have indeed high chances of following, in terms of time and trajectory, the ones that directly received government funding. Similar considerations also apply to Table D5, where we observe the distinct effects of citing patents that acknowledged government interest or citing patents with government assignees. Once again, the latter indicator is the one with the strongest trajectory effects (492.3% versus 208.6% of patents citing inventions with a government interest statement). As expected, it is common to cite patents by federal contractors, government assignees, and US universities simultaneously. This might lead to a reduction of the (citing) government assignee coefficient compared to the one observed in Table 4 and to a loss of significance of the interaction term between timing and citing government interest (the sign is always negative).

Patent relevance A widespread measure of the relevance of a patent is the number of citations it receives. Contrary to our trajectory measure, this indicator does not take into account the patent's indirect effects on sequences of follow-on innovations and is silent on the direction of technical change. In our setting, the risk of reverse causality is quite low: it is implausible that chains of future technology development impact the funding of innovation. However, it is possible that patent applicants and examiners, exerting control over the sources of knowledge they cite in a patent, might favor 'signals' related to government funding over other signals of quality or relevance of the prior art. Whereas the long-term nature of our indicators mitigates this risk because it is difficult to believe that citing sources make decisions across long chains of citations (e.g. patents citing older patents that in turn cite other patents, etc.), it is still possible that there are citation

[18]Given a node p, we consider all the paths through p, and we select the one with the highest W_M, where M is the set of edges of the path. Then, we assign the (weighted) longest path length W_M to p.

biases within each round of citation decisions. It is therefore important to run a specific robustness test to rule out the existence of this particular bias. In Tables D6, D7, D8, D9, D10, and D11, we present results of matched-sample and IV estimations of the effect of government funding on the number of patent citations. By considering a set of different citation indicators (only on patents in the network, all patents, all citations, citations up to or in 5 years), we observe that government backing has, in general, a negative effect on the number of citations. Analogous results are obtained when we consider separately the effect of patents with federal contractors and the government as assignee. On the basis of these results we can also argue that when we measure the relevance of patents through citation counts, we miss or considerably underestimate their importance in the long run, and we fail to capture their full impact on the entire knowledge domain. This confirms that, at least as far as long-term cumulative impact is concerned, the number of citations is not an appropriate indicator of the direction of technical change.

Time effects Tables 3 and 4 show that the effect of government-backed patents is especially relevant at the early stages of the trajectory. To corroborate our findings, we associate to each patent an indicator of the number of paths originating from the invention, namely the forward trajectory indicator (w_p^+ defined in Equation 2). By excluding previous paths, this indicator ranks patents according to their influence on the following inventions, and older patents will have, on average, a higher value of the measure. Tables D12 and D13 confirm our core results, both in terms of sign and magnitude. Overall, government funding has a forward trajectory effect of 243.6%, while government interest and government assignee alone have impacts of, respectively, 177.1% and 1107.5%. Even if we control for the timing (which negatively affects the forward trajectory indicator, as expected), the interaction term between government funding and timing is negative and significant. These results confirm that early government backing of AI technology was particularly important for future developments.

Sample composition To test the robustness of results to sampling choices, we narrow our definition of inventions in artificial intelligence. In particular, we follow the domain definition suggested by WIPO (2019), without adding any other patents. This mainly excludes big-data analytics patents. Estimations on the 111,525 patents belonging to the weakly connected component of this sample are presented in Tables D14 and D15. Results are fully consistent with those we discussed in the previous sections. We also used the Baruffaldi et al. (2020)'s classification of AI patents, and results do not change.[19] An alternative change in the sample composition can be made by selecting only patents granted after 1980 (113,835 patents). The rationale for this test depends on the introduction in that year of the Bayh-Dole Act, which obligated federal contracts to disclose government interests in their patents. Even though we observe government-funded patents (also through grants) also before 1981, there could be misreporting or under-reporting of government interest statements in patents granted between 1976 and 1980. Tables D16 and D17 show that there are no substantial changes in the impact of government funding on the trajectory. In this sample, government funding is associated with an increase of 228.1% in the trajectory effect. This impact is 168.6% and 879.1% respectively for patents with a government interest statement and patents with a government assignee. The timing effect persists and is in line with results discussed in Tables 3 and 4.

Further controls Finally, we implement our models with different controls. In Tables D18 and D19, we replace the US university control with a variable that takes value 1 when the patent has any university (i.e.

[19]These additional results are available from the authors upon request

from anywhere in the world) as assignee and 0 otherwise. Although the university control becomes negative and significant, our results are not affected by this change in regression controls. We propose an additional robustness check where we add the number of backward citations. In this way, we prove that the trajectory effect is not substantially affected by node indegrees (the number of cited patents) but captures the more complex citation structure of the data. Moreover, since we do not observe any change in our core results, we show that they are not driven by the presence of patents that heavily cite previous inventions (Tables D20 and D21). We also control for the weighted average of lagged growth rates of 3-digits CPC classes that are assigned to patents. This variable captures the potential expansion of patents' technological subdomains. Tables D22 and D23 show that, although this control variable has a prominent impact on the trajectory, this does not affect estimations of the effect of government funding.

7 Conclusions

Governments have several instruments at their disposal to address market failures and influence the development of innovation (Bloom et al., 2019; Steinmueller, 2010). Extant literature has focused overwhelmingly on the rate of technical change and the returns to publicly funded R&D. In this paper, we have addressed the problem of the direction of technical change and investigated the role that governments can play in influencing long-term technology development. We focused on AI because this is likely to become a major source of technological spillovers. Even though its potential is arguably far from full realization, AI is a prime candidate to becoming a new general purpose technology, and this makes its choice as field of study highly relevant. By taking a 'big data' approach to the construction of large longitudinal networks of citations, we have been able to quantify the impact of each patent on long-term cumulative patterns of development in the field that cannot be captured by standard indicators such as the number of citations. We have then demonstrated that patents backed by government grants and patents filed by federal agencies and state departments had profound effects on AI innovation, and that their impact appears to be stronger in early phases, while it weakened over time to leave room to privately funded research. This is especially relevant when we consider market failures in high-risk research areas that are in their infancy, but could generate valuable solutions for societal challenges. Naturally, further research can corroborate the external validity of our results by exploring the long-term evolution of technologies in other contexts, or deepen the analysis of specific patterns and effects of public vs. private funding of innovation.

References

Acemoglu, D. (2002). Technical change, inequality, and the labor market. *Journal of Economic Literature*, *40*(1), 7–72.

Acemoglu, D., & Restrepo, P. (2018). *Artificial intelligence, automation and work* (Working Paper No. 24196). National Bureau of Economic Research.

Aghion, P., & Howitt, P. (1992). A model of growth through creative destruction. *Econometrica*, *60*(2), 323–351.

Aghion, P., Jones, B. F., & Jones, C. I. (2017). *Artificial intelligence and economic growth* (Working Paper No. 23928). National Bureau of Economic Research.

Agrawal, A., Gans, J., & Goldfarb, A. (2019). *The economics of artificial intelligence: An agenda.* University of Chicago Press.

Akcigit, U., Grigsby, J., Nicholas, T., & Stantcheva, S. (2018). *Taxation and innovation in the 20th century* (Working Paper No. 24982). National Bureau of Economic Research.

Arrow, K. (1962). Economic welfare and the allocation of resources for invention. *The rate and direction of inventive activity: Economic and social factors* (pp. 609–626). Princeton University Press.

Azoulay, P., Zivin, J., Li, D., & Sampat, B. (2019). Public R&D investment and private sector patenting: Evidence from NIH funding rules. *Review of Economic Studies*, *86*(1), 117–152.

Baruffaldi, S., van Beuzekom, B., Dernis, H., Harhoff, D., Rao, N., Rosenfeld, D., & Squicciarini, M. (2020). Identifying and measuring developments in artificial intelligence: Making the impossible possible. *OECD Science, Technology and Industry Working Papers 2020/05*.

Batagelj, V. (2003). Efficient algorithms for citation network analysis. *arXiv preprint cs/0309023*.

Baumol, W. J. (1967). Macroeconomics of unbalanced growth: The anatomy of urban crisis. *American Economic Review*, *57*(3), 415–426.

Bloom, N., Van Reenen, J., & Williams, H. (2019). Policies to promote innovation. *Journal of Economic Perspectives*, *33*(3), 163–184.

Bloom, N., Griffith, R., & Van Reenen, J. (2002). Do R&D tax credits work? Evidence from a panel of countries 1979–1997. *Journal of Public Economics*, *85*(1), 1–31.

Bloom, N., Schankerman, M., & Van Reenen, J. (2013). Technology spillovers and product market rivalry. *Econometrica*, *81*(4), 1347–1393.

Bresnahan, T. F., & Trajtenberg, M. (1995). General purpose technologies 'engines of growth'? *Journal of Econometrics*, *65*(1), 83–108.

Bronzini, R., & Iachini, E. (2014). Are incentives for R&D effective? Evidence from a regression discontinuity approach. *American Economic Journal: Economic Policy*, *64*(4), 100–134.

Cockburn, I. M., Henderson, R., & Stern, S. (2018). *The impact of artificial intelligence on innovation* (Working Paper No. 24449). National bureau of economic research.

Dechezlepretre, A., Einio, E., Martin, R., Nguyen, K.-T., & Van Reenen, J. (2016). *Do fiscal incentives increase innovation? an rd design for R&D* (Discussion Paper No. 1413). Centre for Economic Performance.

Dosi, G. (1982). Technological paradigms and technological trajectories: A suggested interpretation of the determinants and directions of technical change. *Research Policy*, *11*(3), 147–162.

Dosi, G. (1988). Sources, procedures and microeconomic effects of innovation. *Journal of Economic Literature*, *26*, 1120–1171.

Felten, E., Raj, M., & Seamans, R. (2021). Occupational, industry, and geographic exposure to artificial intelligence: A novel dataset and its potential uses. *Strategic Management Journal*, *42*(12), 2195–2217.

Fleming, L., Greene, H., Li, G., Marx, M., & Yao, D. (2019). Government-funded research increasingly fuels innovation. *Science*, *364*(6446), 1139–1141.

Furman, J., & Seamans, R. (2018). *Ai and the economy* (Working Paper No. 24689). National Bureau of Economic Research.

Green, J. R., & Scotchmer, S. (1995). On the division of profit in sequential innovation. *RAND Journal of Economics*, *26*(1), 20–33.

Griliches, Z. (1957). Hybrid corn: An exploration in the economics of technological change. *Econometrica*, *25*(4), 501–522.

Griliches, Z. (1990). Patent statistics as economic indicators: A survey. *Journal of Economic Literature*, *28*(4), 1661–1707.

Griliches, Z. (1992). The search for R&D spillovers. *Scandinavian Journal of Economics*, *94*, 29–47.

Gross, D. P., & Sampat, B. N. (2020). *Organizing crisis innovation: Lessons from world war II* (Working Paper No. 27909). National Bureau of Economic Research.

Grossman, G. M., & Helpman, E. (1991). *Innovation and growth in the global economy*. MIT press.

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

Hall, B. H., Jaffe, A. B., & Trajtenberg, M. (2001). *The nber patent citation data file: Lessons, insights and methodological tools* (Working Paper No. 8498). National Bureau of Economic Research.

Hall, B. H., Mairesse, J., & Mohnen, P. (2010). Measuring the returns to R&D. In B. H. Hall & N. Rosenberg (Eds.), *Handbook of the economics of innovation* (pp. 1033–1082). Elsevier.

Helpman, E. (1998). *General purpose technologies and economic growth*. MIT press.

Henderson, R., & Cockburn, I. (1996). Scale, scope, and spillovers: The determinants of research productivity in drug discovery. *RAND Journal of Economics*, *27*, 32–59.

Howell, S. T. (2017). Financing innovation: Evidence from R&D grants. *American Economic Review*, *107*(4), 1136–1164.

Hummon, N. P., & Dereian, P. (1989). Connectivity in a citation network: The development of dna theory. *Social Networks*, *11*(1), 39–63.

Jacobides, M. G., Brusoni, S., & Candelon, F. (2021). The evolutionary dynamics of the artificial intelligence ecosystem. *Strategy Science*, *6*(4), 412–435.

Jaffe, A. B. (1986). Technological opportunity and spillovers of R&D: Evidence from firms, patents, profits, and market value. *American Economic Review*, *76*(5), 985–1001.

Jaffe, A. B., Trajtenberg, M., & Henderson, R. (1993). Geographic localization of knowledge spillovers as evidenced by patent citations. *The Quarterly Journal of Economics*, *108*(3), 577–598.

Jovanovic, B., & Rousseau, P. L. (2005). General purpose technologies. In P. Aghion & S. Durlauf (Eds.), *Handbook of economic growth* (pp. 1181–1224). North Holland.

Kaplan, S., & Vakili, K. (2015). The double-edged sword of recombination in breakthrough innovation. *Strategic Management Journal*, *36*(10), 1435–1457.

Kennedy, P. E. (1981). Estimation with correctly interpreted dummy variables in semilogarithmic equations. *American Economic Review*, *71*(4), 801–801.

Korinek, A., & Stiglitz, J. E. (2019). Artificial intelligence and its implications for income distribution and unemployment. In A. Agrawal, J. Gans, & A. Goldfarb (Eds.), *The economics of artificial intelligence* (pp. 349–390). University of Chicago Press.

Korinek, A., & Stiglitz, J. E. (2021). *Artificial intelligence, globalization, and strategies for economic development* (Working Paper No. 28453). National Bureau of Economic Research.

Krakowski, S., Luger, J., & Raisch, S. (2022). Artificial intelligence and the changing sources of competitive advantage. *Strategic Management Journal*.

Kuhn, J. M., & Thompson, N. C. (2019). How to measure and draw causal inferences with patent scope. *International Journal of the Economics of Business*, *26*(1), 5–38.

Lanjouw, J. O., Pakes, A., & Putnam, J. (1998). How to count patents and value intellectual property: The uses of patent renewal and application data. *The Journal of Industrial Economics*, *46*(4), 405–432.

Lerner, J. (1994). The importance of patent scope: An empirical analysis. *RAND Journal of Economics*, *25*(2), 319–333.

Martinelli, A. (2012). An emerging paradigm or just another trajectory? understanding the nature of technological changes using engineering heuristics in the telecommunications switching industry. *Research Policy*, *41*(2), 414–429.

Martinelli, A., Mina, A., & Moggi, M. (2021). The enabling technologies of industry 4.0: Examining the seeds of the fourth industrial revolution. *Industrial and Corporate Change*, 1–28.

Mazzucato, M. (2015). *The entrepreneurial state: Debunking public vs. private sector myths* (Vol. 1). Anthem Press.

Mina, A., Ramlogan, R., Tampubolon, G., & Metcalfe, J. S. (2007). Mapping evolutionary trajectories: Applications to the growth and transformation of medical knowledge. *Research policy*, *36*(5), 789–806.

Moretti, E., Steinwender, C., Van Reenen, J., & Warren, P. (2020). *The intellectual spoils of war? Defense R&D, productivity and international technology spillovers* (Working Paper No. 26483). National Bureau of Economic Research.

Moser, P., & Nicholas, T. (2004). Was electricity a general purpose technology? Evidence from historical patent citations. *American Economic Review*, *94*(2), 388–394.

Mowery, D. (2010). Military R&D and innovation. In B. Hall & N. Rosenberg (Eds.), *The handbook of the economics of innovation* (pp. 1219–1256). Elsevier.

Nelson, R. R. (1959). The simple economics of basic scientific research. *Journal of Political Economy*, *67*(3), 297–306.

Nelson, R. R., & Winter, S. (1982). *An evolutionary theory of economic change*. Harvard University Press.

Nilsson, N. J. (2010). *The quest for artificial intelligence: A history of ideas and achievements*. Cambridge University Press.

Romer, P. (1990). Endogenous technological change. *Journal of Political Economy*, *98*(5), 72–102.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*(1), 41–55.

Ruttan, V. W. (2006). *Is war necessary for economic growth? military procurement and technology development*. Oxford University Press.

Sampat, B., & Williams, H. L. (2019). How do patents affect follow-on innovation? Evidence from the human genome. *American Economic Review*, *109*(1), 203–236.

Santoleri, P., Mina, A., Di Minin, A., & Martelli, I. (2022). The causal effects of R&D grants: Evidence from a regression discontinuity [Available at SSRN: https://ssrn.com/abstract=3637867]. *Review of Economics and Statistics, accepted and forthcoming*.

Scotchmer, S. (1991). Standing on the shoulders of giants: Cumulative research and the patent law. *Journal of Economic Perspectives*, *5*(1), 29–41.

Steinmueller, W. E. (2010). Economics of technology policy. In B. H. Hall & N. Rosenberg (Eds.), *Handbook of the economics of innovation* (pp. 1181–1218). Elsevier.

Trajtenberg, M. (1990). A penny for your quotes: Patent citations and the value of innovations. *The RAND Journal of Economics*, *21*(1), 172–187.

Trajtenberg, M. (2019). Artificial intelligence as the next gpt: A political-economy perspective. In A. Agrawal, J. Gans, & A. Goldfarb (Eds.), *The economics of artificial intelligence* (pp. 175–186). University of Chicago Press.

Trajtenberg, M., Henderson, R., & Jaffe, A. (1997). University versus corporate patents: A window on the basicness of invention. *Economics of Innovation and New Technology*, *5*(1), 19–50.

UKIPO. (2014). *Eight great technologies: Big data. a patent overview*. UK Intellectual Property Office.

Van Reenen, J. (2020). *Innovation policies to boost productivity* (Policy proposal No. 2020–13). The Hamilton Project, Brookings.

Varian, H. (2018). *Artificial intelligence, economics, and industrial organization* (Working Paper No. 24839). National Bureau of Economic Research.

Weitzman, M. L. (1998). Recombinant growth. *Quarterly Journal of Economics*, *113*(2), 331–360.

Wilson, D. J. (2009). Beggar thy neighbor? the in-state, out-of-state, and aggregate effects of R&D tax credits. *Review of Economics and Statistics*, *91*(2), 431–436.

WIPO. (2019). *Wipo technology trends 2019: Artificial intelligence*. World Intellectual Property Organization.

Wuchty, S., Jones, B. F., & Uzzi, B. (2007). The increasing dominance of teams in the production of knowledge. *Science*, *316*, 1036–1039.

A Patents in artificial intelligence: selection procedure

To identify patents in artificial intelligence, we combine selection procedures suggested by the WIPO report on technology trends in artificial intelligence (WIPO, 2019) and the UKIPO report on great technologies (UKIPO, 2014).

A.1 WIPO selection procedure

WIPO (2019) defines three, non-mutually exclusive, blocks of patents, corresponding to different kinds of criteria. The first group is selected through Cooperative Patent Classification (CPC) codes that clearly identify AI-related inventions (Block 1). The second group is identified through specific keywords (Block 2). We search for keywords in patents' titles and abstracts. Finally, the third group combines more generic CPC and International Patent Classification (IPC) codes and keywords (Block 3). To be part of this final set, patents must belong to one of the CPC classes and, at the same time, have one of the keywords in their title or abstract. Therefore, the final query is: (Block 1) OR (Block 2) OR (Block 3), where blocks are defined through the following regular expressions (search patterns in strings).

Block 1 We search for patents whose CPC codes match the following regular expression:

```
^Y10S706 | ^G06N3 | ^G06N5/003$ | ^G06N5/006$ | ^G06N5/02$ | ^G06N5/022$ | ^G06N5/025$
| ^G06N5/027$ | ^G06N7/005$ | ^G06N7/02$ | ^G06N7/023$ | ^G06N7/026$ | ^G06N7/04$
| ^G06N7/043$ | ^G06N7/046$ | ^G06N7/06$ | ^G06N99/005$ | ^G06T2207/20081$
| ^G06T2207/20084$ | ^G06T3/4046$ | ^G06T9/002$ | ^G06F17/16$ |^G05B13/027$
| ^G05B13/0275$ | ^G05B13/028$ | ^G05B13/0285$ | ^G05B13/029$ | ^G05B13/0295$
| ^G05B2219/33002$ | ^G05D1/0088$ | ^G06K9 | ^G10L15 | ^G10L17 |^G06F17/27$
| ^G06F17/2705$ | ^G06F17/271$ | ^G06F17/2715$ | ^G06F17/272$ | ^G06F17/2725$
| ^G06F17/273$ | ^G06F17/2735$ | ^G06F17/274$ | ^G06F17/2745$ |^G06F17/275$
| ^G06F17/2755$ | ^G06F17/276$ | ^G06F17/2765$ | ^G06F17/277$ | ^G06F17/2775$
| ^G06F17/278$ | ^G06F17/2785$ | ^G06F17/279$ | ^G06F17/2795$ | ^G06F17/28$
| ^G06F17/2809$ | ^G06F17/2818$ | ^G06F17/2827$ | ^G06F17/2836$ | ^G06F17/2845$
| ^G06F17/2854$ | ^G06F17/2863$ | ^G06F17/2872$ | ^G06F17/2881$ | ^G06F17/289$
| ^G06F17/30029$ | ^G06F17/30247$ | ^G06F17/3025$ | ^G06F17/30256$ | ^G06F17/30262$
| ^A61B5/7264$ | ^A61B5/7267$ | ^B29C66/965$ | ^B25J9/161$ | ^Y10S128/924$
| ^Y10S128/925$ | ^F02D41/1405$ | ^F03D7/046$ | ^F05B2270/707$ | ^F05B2270/709$
| ^F16H2061/0081$ | ^F16H2061/0084$ | ^B60W30/06$ | ^B60W30/10$ | ^B60W30/12$
| ^B60W30/14$ | ^B60W30/143$ | ^B60W30/146$ | ^B60W30/16$ | ^B60W30/162$ | ^B60W30/165$
| ^B60W30/17$ | ^G06T2207/30248$ | ^G06T2207/30252$ | ^G06T2207/30256$ | ^G06T2207/30261$
| ^G06T2207/30264$ | ^G06T2207/30268$ | ^B62D15/0285$ | ^G06T2207/30236$ | ^A61B5/7267$
| ^F05D2270/709$ | ^G06T2207/20084$ | ^G10K2210/3038$ | ^G10L25/30$ | ^H04N21/4666$
| ^A63F13/67$ | ^G06F17/2282$ | ^G05D1
```

Block 2 We search for patents whose titles and abstracts match the following regular expression:

```
(\bartific\w*\W+(?:\w+\W+){0,1}?intelligen\w*\b) | (\bcomputation\w*\W+(?:\w+\W+){0,1}?intelligen\w
*\b) | (\bneural[\W_]+(?:\w+\W+){0,1}?network\w*\b) | (\bbayesian[\W_]+
(?:\w+\W+){0,1}?network\w*\b) | (\bchatbot\w*\b) | (\bdata\W+(?:\w+\W+){0,1}?mining\w*\b)
| (\bdecision\W+(?:\w+\W+){0,1}?model\w*\b) | (\bdeep[\W_]+(?:\w+\W+){0,1}?learn\w*\b)
```

```
| (\bgenetic\W+(?:\w+\W+){0,1}?algorithm\w*\b) | (\binductive\W+(?:\w+\W+){0,1}?logic
\W+(?:\d+\W+){0,1}?programm\w*\b) | (\bmachine[\W_]+(?:\w+\W+){0,1}?learn\w*\b)
| (\bnatural\W+(?:\d+\W+){0,1}?language\W+(?:\w+\W+){0,1}?generation\w*\b) | (\bnatural
\W+(?:\d+\W+){0,1}?language\W+(?:\w+\W+){0,1}?process\w*\b)|(\breinforcement\W+(?:\w+\W+){0,1}?
learn\w*\b) | (\b\w*supervised[\W_]+(?:\w+\W+){0,1}?learn\w*\b) | (\b\w*supervised[\W_]+(?:\w+\W+)
{0,1}?train\w*\b) | (\bswarm[\W_]+(?:\w+\W+){0,1}?intelligen\w*\b)
| (\bconnectionis\w*\b) | (\bexpert[\W_]+(?:\w+\W+){0,1}?system\w*\b) | (\bfuzzy\W+
(?:\w+\W+){0,1}?logic\w*\b) | (\btransfer[\W_]+(?:\w+\W+){0,1}?learn\w*\b) | (\blearning
\W+(?:\w+\W+){0,3}?algorithm\w*\b) | (\blearing\W+(?:\w+\W+){0,1}?model\w*\b)
| (\bsupport[\W_]+vector[\W_]machine\w*\b) | (\brandom[\W_]forest\w*\b) | (\bdecision
[\W_]tree\w*\b) | (\bgradient[\W_]model[\W_]boosting\b) | (\bxgboost\b) | (\badaboost\b) | (\
brankboost\b) | (\blogistic[\W_]regression\w*\b) | (\bstochastic[\W_]gradient[\W_]descent\b) | (\
bmultilayer[\W_]perceptron\b) | (\blatent[\W_]semantic[\W_]analysis\b)
| (\blatent[\W_]dirichelet[\W_]allocation\b) | (\bmulti[\W_]agent[\ W_]system\w*\b)
| (\bhidden[\W_]markov[\W_]model\w*\b)
```

Block 3 We search for patents whose titles and abstracts match the following keywords and, at the same time, belong to the following CPC or ICP codes.

Keywords

```
(\bclustering | comput\w*[\W_]creativity\b) | (\bdescriptive\Wmodel\w*\b) | (\binductive
\Wreasoning\b) | (\boverfitting\b) | (\bpredictive\W+(?:\w+\W+){0,1}?analytics\b)
| (\bpredictive\W+(?:\w+\W+){0,1}?model\w*\b) | (\btarget\W+(?:\w+\W+){0,1}?function\w*
\b) | (\btest\W+(?:\d+\W+){0,1}?data\b) | (\btraining\W+(?:\d+\W+){0,1}?data\b)
| (\bvalidation\W+(?:\d+\W+){0,1}?data\b) | (\btest\W+(?:\d+\W+){0,1}?set\w*\b)
| (\btraining\W+(?:\d+\W+){0,1}?set\w*\b) | (\bvalidation\W+(?:\d+\W+){0,1}?set\w*\b)
| (\bbackpropagation\w*\b) | (\bself[\W_]learning\b) | (\bobjective\Wfunction\w*\b)
| (\bfeature\w*\Wselection\b) | (\bembedding\w*\b) | (\bactive\Wlearning\b)
| (\bregression\Wmodel\w*\b) | (\bstochastic\W+(?:\d+\W+){0,2}?approach\w*\b)
| (\bprobabilist\w*\W+(?:\d+\W+){0,2}?approach\w*\b) | (\bstochastic\W+(?:\d+\W+){0,2}?technique\w
*\b) | (\bprobabilist\w*\W+(?:\d+\W+){0,2}?technique\w*\b) | (\bstochastic
\W+(?:\d+\W+){0,2}?method\w*\b) | (\bprobabilist\w*\W+(?:\d+\W+){0,2}?method\w*\b)
| (\bstochastic\W+(?:\d+\W+){0,2}?algorithm\w*\b) | (\bprobabilist\w*\W+(?:\d+\W+){0,2}?algorithm\w
*\b) | (\brecommend\w*\Wsystem\w*\b) | (\btext\W+(?:\d+\W+){0,1}analysis\b)
| (\btext\W+(?:\d+\W+){0,1}analytic\w*\b) | (\btext\W+(?:\d+\W+){0,1}recognition\b)
| (\bspeech\W+(?:\d+\W+){0,1}analysis\b) | (\bspeech\W+(?:\d+\W+){0,1}analytic\w*\b)
| (\bspeech\W+(?:\d+\W+){0,1}recognition\b) | (\bhand_writing\W+(?:\d+\W+){0,1}analysis
\b) | (\bhand_writing\W+(?:\d+\W+){0,1}analytic\w*\b) | (\bhand_writing\W+(?:\d+\W+){0,1}
recognition\b) | (\bfacial\W+(?:\d+\W+){0,1}analysis\b) | (\bfacial\W+(?:\d+\W+){0,1}analytic\w*\b)
 | (\bfacial\W+(?:\d+\W+){0,1}recognition\b) | (\bface\w*\W+(?:\d+\W+){0,1}analysis\b) | (\bface\w
*\W+(?:\d+\W+){0,1}analytic\w*\b) | (\bface\w*\W+(?:\d+\W+){0,1}recognition\b) | (\bcharacter\w*\W
+(?:\d+\W+){0,1}analysis\b) | (\bcharacter\w*\W+(?:\d+
\W+){0,1}analytic\w*\b) | (\bcharacter\w*\W+(?:\d+\W+){0,1}recognition\b)
```

CPC

```
^G06F17/14$ | ^G06F17/141$ | ^G06F17/142$ | ^G06F17/144$ | ^G06F17/145$ | ^G06F17/147$
| ^G06F17/148$ | ^G10H2250/005$ | ^G10H2250/011$ | ^G10H2250/015$ | ^G10H2250/021$
```

```
| ^G06Q30/02$ | ^G06Q30/0201$ | ^G06Q30/0202$ | ^G06Q30/0203$ | ^G06Q30/0204$
| ^G06Q30/0205$ | ^G06Q30/0206$ | ^G06Q30/0207$ | ^G06Q30/0208$ | ^G06Q30/0209$
| ^G06Q30/0211$ | ^G06Q30/0212$ | ^G06Q30/0213$ | ^G06Q30/0214$ | ^G06Q30/0215$
| ^G06Q30/0216$ | ^G06Q30/0217$ | ^G06Q30/0218$ | ^G06Q30/0219$ | ^G06Q30/0221$
| ^G06Q30/0222$ | ^G06Q30/0223$ | ^G06Q30/0224$ | ^G06Q30/0225$ | ^G06Q30/0226$
| ^G06Q30/0227$ | ^G06Q30/0228$ | ^G06Q30/0229$ | ^G06Q30/0231$ | ^G06Q30/0232$
| ^G06Q30/0233$ | ^G06Q30/0234$ | ^G06Q30/0235$ | ^G06Q30/0236$ | ^G06Q30/0237$
| ^G06Q30/0238$ | ^G06Q30/0239$ | ^G06Q30/0241$ | ^G06Q30/0242$ | ^G06Q30/0243$
| ^G06Q30/0244$ | ^G06Q30/0245$ | ^G06Q30/0246$ | ^G06Q30/0247$ | ^G06Q30/0248$
| ^G06Q30/0249$ | ^G06Q30/0251$ | ^G06Q30/0252$ | ^G06Q30/0253$ | ^G06Q30/0254$
| ^G06Q30/0255$ | ^G06Q30/0256$ | ^G06Q30/0257$ | ^G06Q30/0258$ | ^G06Q30/0259$
| ^G06Q30/0261$ | ^G06Q30/0262$ | ^G06Q30/0263$ | ^G06Q30/0264$ | ^G06Q30/0265$
| ^G06Q30/0266$ | ^G06Q30/0267$ | ^G06Q30/0268$ | ^G06Q30/0269$ | ^G06Q30/0271$
| ^G06Q30/0272$ | ^G06Q30/0273$ | ^G06Q30/0274$ | ^G06Q30/0275$ | ^G06Q30/0276$
| ^G06Q30/0277$ | ^G06Q30/0278$ | ^G06Q30/0279$ | ^G06Q30/0281$ | ^G06Q30/0282$
| ^G06Q30/0283$ | ^G06Q30/0284$ | ^G06T1/20$ | ^G06F17/153$ | ^G06F17/50$ | ^G06T7
| ^G10L13 | ^G10L25 | ^G10L99 | ^G07C9 | ^G06F21
```

IPC

```
^B25J9/16$ | ^B25J9/18$ | ^B25J9/20$ | ^A63F13/67$ | ^B60W30/06$ | ^A61B5 | ^B23K31
| ^B29C65 | ^B60W30/10$ | ^B60W30/12$ | ^B60W30/14$ | ^B60W30/165$ | ^B60W30/17$
| ^B62D15/02$ | ^B64G1/24$ | ^B64G1/26$ | ^B64G1/28$ | ^B64G1/32$ | ^B64G1/34$
| ^B64G1/36$ | ^B64G1/38$ | ^E21B41$ | ^F02D41/14$ | ^F02D41/16$ | ^F03D7/04$
| ^F16H61 | ^G01N29/44$ | ^G01N29/46$ | ^G01N29/48$ | ^G01N29/50$ | ^G01N29/52$
| ^G01N33 | ^G01R31/28$ | ^G01R31/30$ | ^G01R31/302$ | ^G01R31/303$ | ^G01R31/304$
| ^G01R31/305$ | ^G01R31/306$ | ^G01R31/307$ | ^G01R31/308$ | ^G01R31/309$ | ^G01R31/311$
| ^G01R31/312$ | ^G01R31/315$ | ^G01R31/316$ | ^G01R31/3161$ | ^G01R31/3163$
| ^B60W30/16$ | ^G01R31/3167$ | ^G01R31/317$ | ^G01R31/3173$ | ^G01R31/3177$
| ^G01R31/3181$ | ^G01R31/3183$ | ^G01R31/3185$ | ^G01R31/3187$ | ^G01R31/319$
| ^G01R31/3193$ | ^G01R31/36$ | ^G01R31/364$ | ^G01R31/367$ | ^G01S7/41$ | ^G05B13/02$
| ^G05B13/04$ | ^G05D1 | ^G06F9/44$ | ^G06F9/4401$ | ^G06F9/445$ | ^G06F9/448$
| ^G06F11/14$ | ^G06F11/22$ | ^G06F11/24$ | ^G06F11/25$ | ^G06F11/26$ | ^G06F11/263$
| ^G06F11/267$ | ^G06F11/27$ | ^G06F11/273$ | ^G06F11/277$ | ^G06F15/18$ | ^G06F17/14$
| ^G06F17/15$ | ^G06F17/16$ | ^G06F17/20$ | ^G06F17/27$ | ^G06F17/28$ | ^G06F19/24$
| ^G06K7/14$ | ^G06K9 | ^G06N3 | ^G06N5 | ^G06N7 | ^G06N99 | ^G06T1/20$ | ^G06T1/40$
| ^G06T3/40$ | ^G06T7 | ^G06T9 | ^G08B29/18$ | ^G08B29/20$ | ^G08B29/22$ | ^G08B29/24$
| ^G08B29/26$ | ^G08B29/28$ | ^G10L13 | ^G10L15 | ^G10L17 | ^G10L25 | ^G10L99
| ^G11B20/10$ | ^G11B20/12$ | ^G11B20/14$ | ^G11B20/16$ | ^G11B20/18$ | ^G16H50/20$
| ^H01M8/04992$ | ^H02H1 | ^H02P21 | ^H02P23 | ^H03H17/02$ | ^H03H17/04$ | ^H03H17/06$
| ^H04L12/24$ | ^H04L12/70$ | ^H04L12/701$ | ^H04L12/703$ | ^H04L12/705$ | ^H04L12/707$
| ^H04L12/709$ | ^H04L12/751$ | ^H04L25/02$ | ^H04L25/03$ | ^H04L25/04$ | ^H04L25/05$
| ^H04L25/06$ | ^H04L25/08$ | ^H04L25/10$ | ^H04L25/12$ | ^H04L25/14$ | ^H04L25/17$
| ^H04L25/18$ | ^H04L25/20$ | ^H04L25/22$ | ^H04L25/24$ | ^H04L25/26$ | ^H04L25/03$
| ^H04N21/466$ | ^H04R25 | ^G07C9 | ^G06F21
```

A.2 UKIPO selection procedure

The UKIPO (2014) procedure is based on keyword searches in patents belonging to specific CPC/IPC classes connected to data management and computation. Keywords include generic references to big data – such as *big data*, *open data*, and *business intelligence* – and names of software connected to big data management. Since the report has been published in 2014, we updated the list of software names. We also removed keywords already included in the WIPO search procedure.

We slightly modify this selection procedure by selecting CPC codes (Block 4) and keywords (Block 5) specific to big-data management. For this two groups we do not require the joint presence in patents. Specific CPC codes have been identified by searching these keywords in CPC code titles. A third group of criteria, instead, requires the joint presence of keywords and IPC/CPC codes (Block 6).

Block 4 We search for patents whose CPC codes match the following regular expression:

```
^G06F16/2465$ | ^G06F16/283$ | ^G06F2216/03$
```

Block 5 We search for patents whose titles and abstracts match the following regular expression:

```
((\b|^)big[\W_]+dat\w*(\b|$)) | ((\b|^)open[\W_]+data(\b|$))
| ((\b|^)data[\W_]+mining(\b|$)) | ((\b|^)data[\W_]+fusion(\b|$))
```

Block 6 We search for patents whose titles and abstracts match the following keywords and, at the same time, belong to the following CPC or ICP codes.

Keywords

```
((\b|^)data[\W_]+warehouse\w*(\b|$)) | ((\b|^)hadoop(\b|$)) | ((\b|^)datameer(\b|$)) | ((\b|^)fico
[\W_]+blaze(\b|$)) | ((\b|^)vertica(\b|$)) | ((\b|^)platfora(\b|$)) | ((\b|^)splunk(\b|$)) | ((\b
|^)mapreduce(\b|$)) | ((\b|^)crowdsourcing(\b|$)) | ((\b|^)cluster
[\W_]+computation(\b|$)) | ((\b|^)distributed[\W_]+file[\W_]+system\w*(\b|$)) | ((\b|^)spark(\b|$))
 | ((\b|^)biometrics(\b|$)) | ((\b|^)cassandra(\b|$)) | ((\b|^)nosql(\b|$))
| ((\b|^)behaviow{0,1}ral[\W_]+analytics(\b|$)) | ((\b|^)business[\W_]+intelligence
(\b|$)) | ((\b|^)hanab) | ((\b|^)hive(\b|$)) | ((\b|^)flume (\b|$)) | ((\b|^)kafka(\b|$))
| ((\b|^)elasticsearch(\b|$))
```

CPC

```
^G06F17/3 | ^G06F19/7 | ^G06F19/3 | ^G06F19/1 | ^G06Q10/063 | ^G06Q30/02 | ^G06F17/5
| ^G06N | ^G06F16/ | ^G16Z99/ | ^G16B40/ | ^G16B50/ | ^G16H50/ | ^G16C20/70$ | ^G06F30/
| ^G06F2216/03$
```

IPC

```
^G06F17/3 | ^G06F19/1 | ^G06Q30/02 | ^G06F17/5 | ^G06N
```

B Patents in artificial intelligence: descriptive statistics

Figure B1 shows the evolution of the number of patents in AI over time. We plot the number of patents based both on the application year and grant year. While the application year is closer to the time of the invention and is usually employed in regression analysis, the grant year represents one of the criteria used for the sample selection (we focus on USPTO granted patents (we focus on USPTO patents granted from 1976). For some early patents the difference between the two years is more than ten years.

Table B1 reports the ten most common technologies in the AI patent sample. Technology fields group International Patent Classification (IPC) codes associated with each patent into 35 broad categories. For the sake of simplicity, each patent has been assigned to the prevalent technology. For patents with more than one prevalent technology, we consider a fractional count. Table B2, instead, shows the ten most common CPC codes at the 7-digit level. Compared to technology fields, CPC codes provide a more detailed classification of technological domains. CPC codes are not mutually exclusive, and each patent may occur in more than one class.

Finally, Table B3 reports the ten most common assignees in the AI patent sample, as disambiguated by the USPTO, during the different decades of analysis.

Table B4 summarizes the descriptive statistics of variables used in the econometric analysis, including those used in the robustness checks. The top panel reports statistical information of continuous variables, while the bottom panel shows the number and share of patents with certain characteristics (dummy variables). Finally, Table B5 reports the correlation matrix of these variables and Table B6 summarizes the main federal contractors.

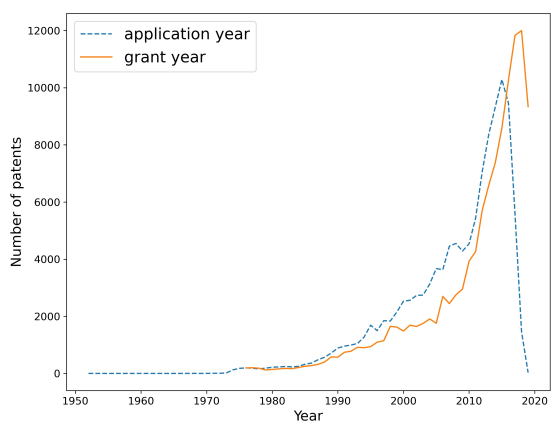


Figure B1: Number of patents in AI per year. The blue dashed line represents the number of patents per application year, while the orange solid line indicates the number of patents per grant year. The grant-year series stops in July 2019 due to data availability.

Technology name	Number of patents	%
Electrical engineering - Computer technology	74192	64.72
Instruments - Control	8513	7.43
Mechanical engineering - Transport	5378	4.69
Instruments - Measurement	4346	3.79
Electrical engineering - Audio-visual technology	4306	3.76
Instruments - Medical technology	3587	3.13
Electrical engineering - Digital communication	2804	2.45
Electrical engineering - Telecommunications	2110	1.84
Electrical engineering - IT methods for management	1808	1.58
Mechanical engineering - Mechanical elements	1094	0.95

Table B1: Main technologies in AI patents. Each patent has been assigned to the prevalent technology. Patents with more than one main technology have been considered as fractional.

CPC class symbol	CPC title	Number of patents	%
G06K 9/00	Methods or arrangements for reading or recognising printed or written characters or for recognising patterns, e.g. fingerprints	54355	47.39
G06T 7/00	Image analysis	15248	13.30
G06F 16/00	Information retrieval; Database structures therefor; File system structures therefor	15000	13.08
G06F 17/00	Digital computing or data processing equipment or methods, specially adapted for specific functions	14252	12.43
G06T2207/00	Indexing scheme for image analysis or image enhancement	13299	11.60
G10L 15/00	Speech recognition	12304	10.73
G05D 1/00	Control of position, course or altitude of land, water, air, or space vehicles, e.g. automatic pilot	11104	9.68
G06F 3/00	Input arrangements for transferring data to be processed into a form capable of being handled by the computer; Output arrangements for transferring data from processing unit to output unit, e.g. interface arrangements	10051	8.76
H04N 5/00	Details of television systems	6268	5.47
G06K2209/00	Indexing scheme relating to methods or arrangements for reading or recognising printed or written characters or for recognising patterns, e.g. fingerprints	6198	5.40

Table B2: Most common CPC classes at 7-digits level in AI patents. CPC codes are not mutually exclusive, and each patent may occur in more than one class.

Assignee	Nb patents	%
Hitachi	187	4.00
IBM	166	3.55
Boeing	114	2.44
Sharp	97	2.07
United Technologies	74	1.58
Toshiba	70	1.50
AT&T	60	1.28
Texas Instruments	58	1.24
Recognition Equipment	49	1.05
NEC	47	1.00

(a) Before 1990

Assignee	Nb patents	%
IBM	812	5.73
Canon	369	2.60
Matsushita Electric Industrial	284	2.00
Hitachi	269	1.90
Ricoh	265	1.87
Xerox	253	1.78
Microsoft	249	1.76
Toshiba	223	1.57
Fujitsu	210	1.48
NEC	194	1.37

(b) 1990–2000

Assignee	Nb patents	%
Microsoft	2008	5.86
IBM	1831	5.34
Sony	678	1.98
Canon	622	1.81
AT&T	516	1.51
HP Development Company	459	1.34
Samsung Electronics Ltd.	445	1.30
Silverbrook Research	374	1.09
Toshiba	373	1.09
Xerox	372	1.09

(c) 2000–2010

Assignee	Nb patents	%
IBM	3902	6.34
Google	2796	4.54
Microsoft	1671	2.71
Samsung Electronics	1133	1.84
Amazon Technologies	1107	1.80
Canon	818	1.33
Sony	785	1.28
Apple	623	1.01
AT&T	621	1.01
Intel	612	0.99

(d) After 2010

Table B3: Leading assignees in AI patents

Variable name	Min	Mean	Max	Std
Trajectory	1	$8.25 \cdot 10^{15}$	$1.87 \cdot 10^{19}$	$2.20 \cdot 10^{17}$
Longest path length	1	$2.27 \cdot 10^{19}$	$1.73 \cdot 10^{20}$	$3.74 \cdot 10^{19}$
Timing	0	16.17	55	11.27
Number of claims	1	18.92	522	12.35
Number of inventors	1	2.74	27	18.80
Application year	1952	2007.75	2019	8.49
Grant year	1976	2010.80	2019	8.41
Number of references	0	31.07	3951	90.61
Number of citations (network)	0	10.92	805	24.08
Number of citations (all)	0	25.07	2288	53.32
Number of citations up to 5 years (network)	0	4.55	240	8.63
Number of citations up to 5 years (all)	0	10.21	1156	18.28
Number of citations in 5 years (network)	0	5.94	240	9.50
Number of citations in 5 years (all)	0	14.16	1156	20.74

Variable name	Number of patents	%
Government funding	3932	3.43
Government interest	3597	3.14
Government assignee	929	0.81
US university	2947	2.57
University	4588	4.00
Citing government funding	34692	30.25
Citing government interest	31837	27.76
Citing government assignee	14075	12.27
Citing US university	31491	27.46

Table B4: Descriptive statistics of continuous (top) and dummy (bottom) variables.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
1: Trajectory	1.00																					
2: Longest path length	0.14	1.00																				
3: Timing	0.03	0.32	1.00																			
4: Nb of claims	0.01	0.01	0.03	1.00																		
5: Nb of inventors	0.00	0.01	0.11	0.05	1.00																	
6: Application year	-0.03	-0.23	0.50	0.00	0.13	1.00																
7: Nb of references	0.11	0.09	0.19	0.09	0.06	0.10	1.00															
8: Nb of cit (network)	0.09	0.27	-0.07	0.14	0.01	-0.31	0.03	1.00														
9: Nb of cit (all)	0.05	0.18	-0.16	0.18	0.02	-0.35	0.03	0.78	1.00													
10: Nb of cit up to 5 years (network)	0.12	0.16	0.08	0.14	0.06	-0.09	0.10	0.68	0.50	1.00												
11: Nb of cit up to 5 years (all)	0.07	0.06	-0.03	0.20	0.06	-0.14	0.10	0.55	0.70	0.75	1.00											
12: Nb of cit 5 years (network)	0.12	0.21	0.26	0.15	0.10	0.10	0.16	0.68	0.47	1.00	0.71	1.00										
13: Nb of cit 5 years (all)	0.06	0.08	0.12	0.21	0.12	0.09	0.20	0.51	0.69	0.71	1.00	0.71	1.00									
14: Government funding	-0.01	-0.02	-0.05	0.01	0.02	-0.05	-0.02	0.02	0.02	0.01	0.01	0.01	-0.00	1.00								
15: Government interest	-0.00	-0.02	-0.04	0.01	0.02	-0.04	-0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.96	1.00							
16: Government assignee	-0.00	-0.00	-0.05	-0.03	-0.02	-0.08	-0.02	-0.00	0.00	-0.02	-0.02	-0.03	-0.04	0.48	0.32	1.00						
17: US university	-0.00	-0.01	-0.04	0.05	0.03	-0.02	-0.02	0.02	0.03	0.02	0.03	0.02	0.02	0.38	0.39	-0.00	1.00					
18: University	-0.01	-0.03	-0.03	0.02	0.05	0.01	-0.03	0.00	0.01	0.01	0.01	0.01	0.01	0.29	0.31	-0.01	0.80	1.00				
19: Citing government funding	0.01	0.01	0.15	0.11	0.06	0.05	0.22	0.05	0.05	0.10	0.10	0.10	0.10	0.14	0.13	0.07	0.10	0.08	1.00			
20: Citing government interest	0.01	0.01	0.16	0.10	0.06	0.07	0.22	0.05	0.04	0.10	0.09	0.10	0.10	0.14	0.14	0.05	0.09	0.08	0.94	1.00		
21: Citing government assignee	0.00	0.01	0.05	0.08	0.03	-0.02	0.24	0.05	0.06	0.08	0.08	0.05	0.06	0.09	0.08	0.10	0.04	0.03	0.57	0.43	1.00	
22: Citing US university	0.03	0.02	0.17	0.12	0.06	0.08	0.23	0.05	0.06	0.11	0.11	0.12	0.13	0.08	0.08	0.01	0.12	0.11	0.55	0.55	0.29	1.00

Table B5: Correlation matrix

Assignee	Number of patents	%
International Business Machines Corporation	268	7.45
University of California	122	3.39
Massachusetts Institute of Technology	99	2.75
HRL Laboratories	78	2.17
SRI International	73	2.03
Honeywell International	65	1.81
California Institute of Technology	61	1.70
University of Southern California	60	1.67
The Boeing Company	57	1.58
United Technologies Corporation	49	1.36

Table B6: Main federal contractors: assignees funded by federal agencies and state departments for supporting the development of AI patents.

C Main path patents

Table C1: Patents in the AI main path. Node numbers link this table to Figure 1. Patent numbers are hyperlinks that lead to patent documents.

Node number	Patent number	Patent title
0	US2432123A	Translation of visual symbols
1	US2615992A	Apparatus for indicia recognition
2	US2897481A	Apparatus for reading
3	US2932006A	Symbol recognition system
4	US2889535A	Recognition of recorded intelligence
5	US2928074A	Method and apparatus for reading handwritten symbols, particularly numerals
6	US2964734A	Method and apparatus for sensing handwriten or printed characters
7	US3105956A	Character recognition system
8	US3069079A	Automatic character recognition method
9	US2959769A	Data consolidation systems
10	US3025495A	Automatic character recognition
11	US3112468A	Character recognition system
12	US3108254A	Machine reading of handwritten characters
13	US3179923A	Scanning system for large areas
14	US3173126A	Reading machine with core matrix
15	US3234513A	Character recognition apparatus
16	US3165717A	Character recognition system
17	US3200373A	Handwritten character reader
18	US3104369A	High-speed optical identification of printed matter
19	US3289164A	Character normalizing reading machine
20	US3496542A	Multifont character reading machine
21	US3601802A	Pattern matching character recognition system
22	US3816722A	Computer for calculating the similarity between patterns and pattern recognition system comprising the similarity computer
23	US4049913A	System for recognizing speech continuously spoken with number of word or words preselected
24	US4092493A	Speech recognition system
25	US4060694A	Speech recognition method and apparatus adapted to a plurality of different speakers
26	US4156868A	Syntactic word recognizer
27	US4059725A	Automatic continuous speech recognition system employing dynamic programming
28	US4256924A	Device for recognizing an input pattern with approximate patterns used for reference patterns on mapping
29	US4181821A	Multiple template speech recognition system
30	US4336421A	Apparatus and method for recognizing spoken words
31	US4277644A	Syntactic continuous speech recognizer
32	US4349700A	Continuous speech recognition system
33	US4319221A	Similarity calculator comprising a buffer for a single input pattern feature vector to be pattern matched with reference patterns
34	US4504970A	Training controller for pattern processing system
35	US4355302A	Spelled word recognizer
36	US4384273A	Time warp signal recognition processor for matching signal patterns
37	US4400788A	Continuous speech pattern recognizer
38	US4286115A	System for recognizing words continuously spoken according to a format
39	US4400828A	Word recognizer
40	US4593367A	Probabilistic learning element
41	US4618983A	Speech recognition with preliminary matching

Continued on next page

42	US4580241A	Graphic word spelling correction using automated dictionary comparisons with phonetic skeletons
43	US4481593A	Continuous speech recognition
44	US4852173A	Design and construction of a binary-tree system for language modelling
45	US4754489A	Means for resolving ambiguities in text based upon character context
46	US4670848A	Artificial intelligence system
47	US4674066A	Textual database system using skeletonization and phonetic replacement to retrieve words matching or similar to query words
48	US4730269A	Method and apparatus for generating word skeletons utilizing alpha set replacement and omission
49	US4587670A	Hidden Markov model speech recognition arrangement
50	US4559604A	Pattern recognition method
51	US4805225A	Pattern recognition method and apparatus
52	US4796199A	Neural-model, information-handling architecture and method
53	US4881178A	Method of controlling a classifier system
54	US4821333A	Machine learning procedures for generating image domain feature detector structuring elements
55	US4837689A	Inputting and editing system in a knowledge based inquiry and answer system
56	US4931926A	Inputting system and an editing system in an inquiry-and-answer system
57	US4866635A	Domain independent shell for building a diagnostic expert system
58	US4815005A	Semantic network machine for artificial intelligence computer
59	US4835690A	Integrated expert system for medical imaging scan, set-up, and scheduling
60	US4771401A	Apparatus and method for linguistic expression processing
61	US4783758A	Automated word substitution using numerical rankings of structural disparity between misspelled words & candidate substitution words
62	US4713778A	Speech recognition method
63	US4713777A	Speech recognition method having noise immunity
64	US4718092A	Speech recognition activation and deactivation method
65	US4718093A	Speech recognition method including biased principal components
66	US4718094A	Speech recognition system
67	US4712242A	Speaker-independent word recognizer
68	US4712243A	Speech recognition apparatus
69	US4715004A	Pattern recognition system
70	US4975961A	Multi-layer neural network to which dynamic programming techniques are applicable
71	US4876731A	Neural network model in pattern recognition using probabilistic contextual information
72	US4965725B1	Neural network based automated cytological specimen classification system and method
73	US5053974A	Closeness code and method
74	US5067095A	SPANN: Sequence processing artificial neural network
75	US5056037A	Analog hardware for learning neural networks
76	US4897811A	N-dimensional coulomb neural network which provides for cumulative learning of internal representations
77	US4918617A	Neural-model computational system with multi-directionally overlapping broadcast regions
78	US4935877A	Non-linear genetic algorithms for solving problems
79	US4994967A	Information retrieval system with means for analyzing undefined words in a natural language inquiry
80	US5103498A	Intelligent help system
81	US5041976A	Diagnostic system using pattern recognition for electronic automotive control systems
82	US5274801A	Artificial intelligence delivery system
83	US4864501A	Word annotation system
84	US4887212A	Parser for natural language text
85	US4849898A	Method and apparatus to identify the relation of meaning between words in text expressions
86	US4759068A	Constructing Markov models of words from multiple utterances
87	US5046099A	Adaptation of acoustic prototype vectors in a speech recognition system
88	US4803729A	Speech recognition method
89	US5058166A	Method of recognizing coherently spoken words

Continued on next page

90	US4987596A	Knowledge-guided automatic speech recognition apparatus and method
91	US4833712A	Automatic generation of simple Markov model stunted baseforms for words in a vocabulary
92	US4827521A	Training of Markov models used in a speech recognition system
93	US4852180A	Speech recognition by acoustic/phonetic system and technique
94	US4783803A	Speech recognition apparatus and method
95	US4837831A	Method for creating and using multiple-word sound models in speech recognition
96	US5040215A	Speech recognition apparatus using neural network and fuzzy logic
97	US5175793A	Recognition apparatus using articulation positions for recognizing a voice
98	US5046019A	Fuzzy data comparator with neural network postprocessor
99	US5058180A	Neural network apparatus and method for pattern recognition
100	US5052043A	Neural network with back propagation controlled through an output confidence measure
101	US5060278A	Pattern recognition apparatus using a neural network system
102	US5048100A	Self organizing neural network method and system for general classification of patterns
103	US5086479A	Information processing system using neural network learning function
104	US5058184A	Hierarchical information processing system
105	US5333239A	Learning process system for use with a neural network structure data processing apparatus
106	US5067164A	Hierarchical constrained automatic learning neural network for character recognition
107	US5170463A	Neuro-computer
108	US5140530A	Genetic algorithm synthesis of neural networks
109	US5390281A	Method and apparatus for deducing user intent and providing computer implemented services
110	US5497319A	Machine translation and telecommunications system
111	US5068789A	Method and means for grammatically processing a natural language sentence
112	US5060155A	Method and system for the representation of multiple analyses in dependency grammar and parser for generating such representation
113	US5099425A	Method and apparatus for analyzing the semantics and syntax of a sentence or a phrase
114	US4817156A	Rapidly training a speech recognizer to a subsequent speaker given training data of a reference speaker
115	US4829577A	Speech recognition method
116	US5222147A	Speech recognition LSI system including recording/reproduction device
117	US5054074A	Optimized speech recognition system and method
118	US4926488A	Normalization of speech by adaptive labelling
119	US4941178A	Speech recognition using preclassification and spectral normalization
120	US5072452A	Automatic determination of labels and Markov word models in a speech recognition system
121	US5208897A	Method and apparatus for speech recognition based on subsyllable spellings
122	US5202952A	Large-vocabulary continuous speech prefiltering and processing system
123	US5033087A	Method and apparatus for the automatic determination of phonological rules as for a continuous speech recognition system
124	US5018201A	Speech recognition dividing words into two portions for preliminary selection
125	US5146503A	Speech recognition
126	US4866778A	Interactive speech recognition apparatus
127	US5278911A	Speech recognition using a neural net
128	US5251286A	Method for estimating formation permeability from wireline logs using neural networks
129	US5162997A	Control system for automotive vehicle for controlling vehicle driving behavior with feature of harmonization of vehicular driving condition dependent control and driver's driving tendency adapted control
130	US5247584A	Signal processing unit for classifying objects on the basis of signals from sensors
131	US5155801A	Clustered neural networks
132	US5239594A	Self-organizing pattern classification neural network system
133	US5105468A	Time delay neural network for printed and cursive handwritten character recognition
134	US5265224A	Recognition unit and recognizing and judging apparatus employing same
135	US5179596A	Analog pattern categorization system having dual weighted connectivity between nodes
136	US5220640A	Neural net architecture for rate-varying inputs

Continued on next page

137	US5271090A	Operational speed improvement for neural network
138	US5317675A	Neural network pattern recognition learning method
139	US5500920A	Semantic co-occurrence filtering for speech recognition and signal transcription applications
140	US5243520A	Sense discrimination system and method
141	US5128865A	Method for determining the semantic relatedness of lexical items in a text
142	US5148489A	Method for spectral estimation to improve noise robustness for speech recognition
143	US5150449A	Speech recognition apparatus of speaker adaptation type
144	US5027406A	Method for interactive speech recognition and training
145	US5278942A	Speech coding apparatus having speaker dependent prototypes generated from nonuser reference data
146	US5031217A	Speech recognition system using Markov models having independent label output sets
147	US5050215A	Speech recognition method
148	US5220639A	Mandarin speech input method for Chinese computers and a mandarin speech recognition machine
149	US5315689A	Speech recognition system having word-based and phoneme-based recognition means
150	US5195167A	Apparatus and method of grouping utterances of a phoneme into context-dependent categories based on sound-similarity for automatic speech recognition
151	US5129001A	Method and apparatus for modeling words with multi-arc Markov models
152	US5170432A	Method of speaker adaptive speech recognition
153	US5168524A	Ch-recognition circuitry employing nonlinear processing, speech element modeling and phoneme estimation
154	US5133012A	Speech recognition system utilizing both a long-term strategic and a short-term strategic scoring operation in a transition network thereof
155	US5193142A	Training module for estimating mixture Gaussian densities for speech-unit models in speech recognition systems
156	US5293584A	Speech recognition system for natural language translation
157	US5526463A	System for processing a succession of utterances spoken in continuous or discrete form
158	US5202926A	Phoneme discrimination method
159	US5526465A	Methods and apparatus for verifying the originator of a sequence of operations
160	US5680509A	Method and apparatus for estimating phone class probabilities a-posteriori using a decision tree
161	US4977598A	Efficient pruning algorithm for hidden Markov model speech recognition
162	US4984178A	Chart parser for stochastic unification grammar
163	US5199077A	Wordspotting for voice editing and indexing
164	US5075896A	Character and phoneme recognition based on probability clustering
165	US5007081A	Speech activated telephone
166	US5136654A	Vocabulary partitioned speech recognition apparatus
167	US5065431A	Pattern recognition using stored N-tuple occurrence frequencies
168	US5475798A	Speech-to-text translator
169	US5517667A	Neural network that does not require repetitive training
170	US5285523A	Apparatus for recognizing driving environment of vehicle
171	US5408588A	Artificial neural network method and architecture
172	US5517597A	Convolutional expert neural system (ConExNS)
173	US5461696A	Decision directed adaptive neural network
174	US5276771A	Rapidly converging projective neural network
175	US5541836A	Word disambiguation apparatus and methods
176	US5321607A	Automatic translating machine
177	US5212821A	Machine-based learning system
178	US5307444A	Voice analyzing system using hidden Markov model and having plural neural network predictors
179	US5329609A	Recognition apparatus with function of displaying plural recognition candidates
180	US5649056A	Speech recognition system and method which permits a speaker's utterance to be recognized using a hidden Markov model with subsequent calculation reduction
181	US5425129A	Method for word spotting in continuous speech

Continued on next page

182	US5502791A	Speech recognition by concatenating fenonic allophone hidden Markov models in parallel among subwords
183	US5345536A	Method of speech recognition
184	US5309547A	Method of speech recognition
185	US5222146A	Speech recognition apparatus having speech coder outputting acoustic prototype ranks
186	US5613036A	Dynamic categories for a speech recognition system
187	US5390278A	Phoneme based speech recognition
188	US5444617A	Method and apparatus for adaptively generating field of application dependent language models for use in intelligent systems
189	US5679001A	Children's speech training aid
190	US5233681A	Context-dependent speech recognizer using estimated next word context
191	US5390280A	Speech recognition apparatus
192	US5276766A	Fast algorithm for deriving acoustic prototypes for automatic speech recognition
193	US5452397A	Method and system for preventing entry of confusingly similar phases in a voice recognition system vocabulary list
194	US5608841A	Method and apparatus for pattern recognition employing the hidden Markov model
195	US5455889A	Labelling speech using context-dependent acoustic prototypes
196	US5329608A	Automatic speech recognizer
197	US5333275A	System and method for time aligning speech
198	US5825978A	Method and apparatus for speech recognition using optimized partial mixture tying of HMM state functions
199	US5459815A	Speech recognition method using time-frequency masking mechanism
200	US5268990A	Method for recognizing speech using linguistically-motivated hidden Markov models
201	US5640490A	User independent, real-time speech recognition system and method
202	US5477451A	Method and system for natural language translation
203	US5418717A	Multiple score language processing system
204	US5864810A	Method and apparatus for speech recognition adapted to an individual speaker
205	US5440662A	Keyword/non-keyword classification in isolated word speech recognition
206	US5526259A	Method and apparatus for inputting text
207	US5428707A	Apparatus and methods for training speech recognition systems and their users and otherwise improving speech recognition performance
208	US5222121A	Voice recognition dialing unit
209	US5386492A	Speech recognition system utilizing vocabulary model preselection
210	US5510981A	Language translation apparatus and method using context-based translation models
211	US5481644A	Neural network speech recognition apparatus recognizing the frequency of successively input identical speech data sequences
212	US5796921A	Mapping determination methods and data discrimination methods using the same
213	US5301257A	Neural network
214	US5704013A	Map determination method and apparatus
215	US5528491A	Apparatus and method for automated natural language translation
216	US5477450A	Machine translation method and apparatus
217	US5608623A	Special cooccurrence processing method and apparatus
218	US5805771A	Automatic language identification method and system
219	US5502774A	Automatic recognition of a consistent message using multiple complimentary sources of information
220	US5737485A	Method and apparatus including microphone arrays and neural networks for speech/speaker recognition systems
221	US5475792A	Telephony channel simulator for speech recognition application
222	US5513298A	Instantaneous context switching for speech recognition systems
223	US5488652A	Method and apparatus for training speech recognition algorithms for directory assistance applications
224	US5487133A	Distance calculating neural network classifier chip and system
225	US5668929A	Speech activated security systems and methods

Continued on next page

226	US5615296A	Continuous speech recognition and voice response system and method to enable conversational dialogues with microprocessors
227	US5638425A	Automated directory assistance system using word recognition and phoneme processing method
228	US5566272A	Automatic speech recognition (ASR) processing using confidence measures
229	US5758319A	Method and system for limiting the number of words searched by a voice recognition system
230	US5794198A	Pattern recognition method
231	US5825977A	Word hypothesizer based on reliably detected phoneme similarity regions
232	US5684925A	Speech representation by feature-based word prototypes comprising phoneme targets having reliable high similarity
233	US5822728A	Multistage word recognizer based on reliably detected phoneme similarity regions
234	US5515475A	Speech recognition method using a two-pass search
235	US5623609A	Computer system and computer-implemented process for phonology-based automatic speech recognition
236	US5594834A	Method and system for recognizing a boundary between sounds in continuous speech
237	US5634086A	Method and apparatus for voice-interactive language instruction
238	US5623578A	Speech recognition system allows new vocabulary words to be added without requiring spoken samples of the words
239	US5799279A	Continuous speech recognition of text and commands
240	US5524169A	Method and system for location-specific speech recognition
241	US5497447A	Speech coding apparatus having acoustic prototype vectors generated by tying to elementary models and clustering around reference vectors
242	US5590242A	Signal bias removal for robust telephone speech recognition
243	US5768603A	Method and system for natural language translation
244	US5581655A	Method for recognizing speech using linguistically-motivated hidden Markov models
245	US5274739A	Product code memory Itakura-Saito (MIS) measure for sound recognition
246	US5748841A	Supervised contextual language acquisition system
247	US5649057A	Speech recognition employing key word modeling and non-key word modeling
248	US5509104A	Speech recognition employing key word modeling and non-key word modeling
249	US5621859A	Single tree method for grammar directed, very large vocabulary speech recognizer
250	US5991721A	Apparatus and method for processing natural language and apparatus and method for speech recognition
251	US5864788A	Translation machine having a function of deriving two or more syntaxes from one original sentence and giving precedence to a selected one of the syntaxes
252	US5850627A	Apparatuses and methods for training and operating speech recognition systems
253	US5450525A	Vehicle accessory control with manual and voice response
254	US5983179A	Speech recognition system which turns its voice response on for confirmation when it has been turned off without confirmation
255	US5764853A	Voice recognition device and method using a (GGM) Guaranteed Global minimum Mapping
256	US5867811A	Method, an apparatus, a system, a storage device, and a computer readable medium using a bilingual database including aligned corpora
257	US5907821A	Method of computer-based automatic extraction of translation pairs of words from a bilingual text
258	US5752232A	Voice activated device and method for providing access to remotely retrieved data
259	US5765132A	Building speech models for new words in a multi-word utterance
260	US5819220A	Web triggered word set boosting for speech interfaces to the world wide web
261	US5987414A	Method and apparatus for selecting a vocabulary sub-set from a speech recognition dictionary for use in real time automated directory assistance
262	US5749072A	Communications device responsive to spoken commands and methods of using same
263	US5983186A	Voice-activated interactive speech recognition device and method
264	US6061654A	System and method of recognizing letters and numbers by either speech or touch tone recognition utilizing constrained confusion matrices
265	US5787394A	State-dependent speaker clustering for speaker adaptation

Continued on next page

266	US6055498A	Method and apparatus for automatic text-independent grading of pronunciation for language instruction
267	US5774628A	Speaker-independent dynamic vocabulary and grammar in speech recognition
268	US5721808A	Method for the composition of noise-resistant hidden Markov models for speech recognition and speech recognizer using the same
269	US5583965A	Methods and apparatus for training and operating voice recognition systems
270	US5963892A	Translation apparatus and method for facilitating speech input operation and obtaining correct translation thereof
271	US5561722A	Pattern matching method and pattern recognition apparatus
272	US6085162A	Translation system and method in which words are translated by a specialized dictionary and then a general dictionary
273	US6161083A	Example-based translation method and system which calculates word similarity degrees, a priori probability, and transformation probability to determine the best example for translation
274	US5950157A	Method for establishing handset-dependent normalizing models for speaker recognition
275	US5960399A	Client/server speech processor/recognizer
276	US6078886A	System and method for providing remote automatic speech recognition services via a packet network
277	US6195641B1	Network universal spoken language vocabulary
278	US6125341A	Speech recognition system and method
279	US6070140A	Speech recognizer
280	US5715367A	Apparatuses and methods for developing and using models for speech recognition
281	US7020609B2	Voice activated apparatus for accessing information on the World Wide Web
282	US5860062A	Speech recognition apparatus and speech recognition method
283	US5617509A	Method, apparatus, and radio optimizing Hidden Markov Model speech recognition
284	US5664058A	Method of training a speaker-dependent speech recognizer with automated supervision of training sufficiency
285	US6266642B1	Method and portable apparatus for performing spoken language translation
286	US6366886B1	System and method for providing remote automatic speech recognition services via a packet network
287	US6453290B1	Method and system for network-based speech recognition
288	US5799065A	Call routing device employing continuous speech
289	US7099824B2	Speech recognition system, speech recognition server, speech recognition client, their control method, and computer readable memory
290	US6463413B1	Speech recognition training for small hardware devices
291	US5970446A	Selective noise/channel/coding models and recognizers for automatic speech recognition
292	US6134527A	Method of testing a vocabulary word being enrolled in a speech recognition system
293	US6101472A	Data processing system and method for navigating a network using a voice command
294	US6061646A	Kiosk for multiple spoken languages
295	US6377922B2	Distributed recognition system having multiple prompt-specific and response-specific speech recognizers
296	US6260012B1	Mobile phone having speaker dependent voice recognition method and apparatus
297	US6192338B1	Natural language knowledge servers as network resources
298	US7203651B2	Voice control system with multiple voice recognition engines
299	US6456974B1	System and method for adding speech recognition capabilities to java
300	US7409349B2	Servers for web enabled speech recognition
301	US7610547B2	Markup language extensions for web enabled recognition
302	US7506022B2	Web enabled recognition architecture
303	US8229753B2	Web server controls for web enabled recognition and/or audible prompting
304	US7003464B2	Dialog recognition and control in a voice browser
305	US7260535B2	Web server controls for web enabled recognition and/or audible prompting for call controls
306	US8311835B2	Assisted multi-modal dialogue
307	US7552055B2	Dialog component re-use in recognition systems

Continued on next page

308	US9083798B2	Enabling voice selection of user preferences
309	US7917365B2	Synchronizing visual and speech events in a multimodal application
310	US8090584B2	Modifying a grammar of a hierarchical multimodal menu in dependence upon speech command frequency
311	US9208785B2	Synchronizing distributed speech recognition
312	US7848314B2	VOIP barge-in support for half-duplex DSR client on a full-duplex network
313	US7676371B2	Oral modification of an ASR lexicon of an ASR engine
314	US8145493B2	Establishing a preferred mode of interaction between a user and a multimodal application
315	US8374874B2	Establishing a multimodal personality for a multimodal application in dependence upon attributes of user interaction
316	US8086463B2	Dynamically generating a vocal help prompt in a multimodal application
317	US7827033B2	Enabling grammars in web page frames
318	US8612230B2	Automatic speech recognition with a selection list
319	US8055504B2	Synchronizing visual and speech events in a multimodal application
320	US8069047B2	Dynamically defining a VoiceXML grammar in an X+V page of a multimodal application
321	US7840409B2	Ordering recognition results produced by an automatic speech recognition engine for a multimodal application
322	US8938392B2	Configuring a speech engine for a multimodal application based on location
323	US8713542B2	Pausing a VoiceXML dialog of a multimodal application
324	US9208783B2	Altering behavior of a multimodal application based on location
325	US7809575B2	Enabling global grammars for a particular multimodal application
326	US7822608B2	Disambiguating a speech recognition grammar in a multimodal application
327	US8909532B2	Supporting multi-lingual user interaction with a multimodal application
328	US9973450B2	Methods and systems for dynamically updating web service profile information by parsing transcribed message strings
329	US9349367B2	Records disambiguation in a multimodal application operating on a multimodal device
330	US8326636B2	Using a physical phenomenon detector to control operation of a speech recognition engine
331	US8352261B2	Use of intermediate speech transcription results in editing final speech transcription results
332	US8355914B2	Mobile terminal and method for correcting text thereof
333	US8352264B2	Corrective feedback loop for automated speech recognition
334	US8494852B2	Word-level correction of speech input
335	US8676577B2	Use of metadata to post process speech recognition output
336	US8478590B2	Word-level correction of speech input
337	US8626511B2	Multi-dimensional disambiguation of voice commands
338	US8560301B2	Apparatus and method for language expression using context and intent awareness
339	US9858925B2	Using context information to facilitate processing of commands in a virtual assistant
340	US9117447B2	Using event alert text as input to an automated assistant
341	US8799000B2	Disambiguation based on active input elicitation by intelligent automated assistant
342	US10134385B2	Systems and methods for name pronunciation
343	US10176167B2	System and method for inferring user intent from speech inputs
344	US9548066B2	Voice application architecture
345	US9767091B2	Methods for understanding incomplete natural language query
346	US9899020B2	Machine learning dialect identification
347	US10133738B2	Translation confidence scores
348	US9734143B2	Multi-media context language processing
349	US10002125B2	Language model personalization
350	US9805029B2	Predicting future translations
351	US9747283B2	Predicting future translations
352	US10002131B2	Classifying languages for objects and entities
353	US10275459B1	Source language content scoring for localizability
354	US10223356B1	Abstraction of syntax in localization through pre-rendering
355	US10229113B1	Leveraging content dimensions during the translation of human-readable languages

Continued on next page

356	US10261995B1	Semantic and natural language processing for content categorization and routing
357	US10289681B2	Predicting future translations
358	US10013417B2	Classifying languages for objects and entities
359	US10089299B2	Multi-media context language processing
360	US10346537B2	Universal translation
361	US10180935B2	Identifying multiple languages in a content item

D Supplementary results

D.1 Effect of timing of government funding

Figures D1 and D2 show the effect of timing of government funding (and its division into government interest and government assignee) on the trajectory indicator.

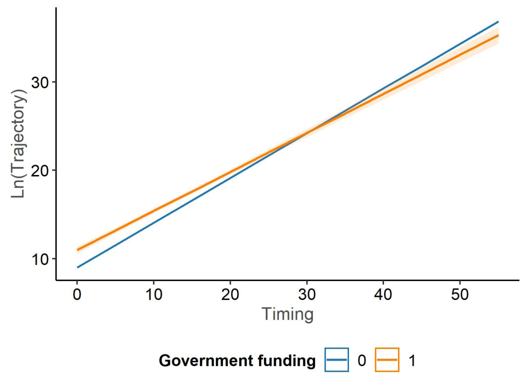
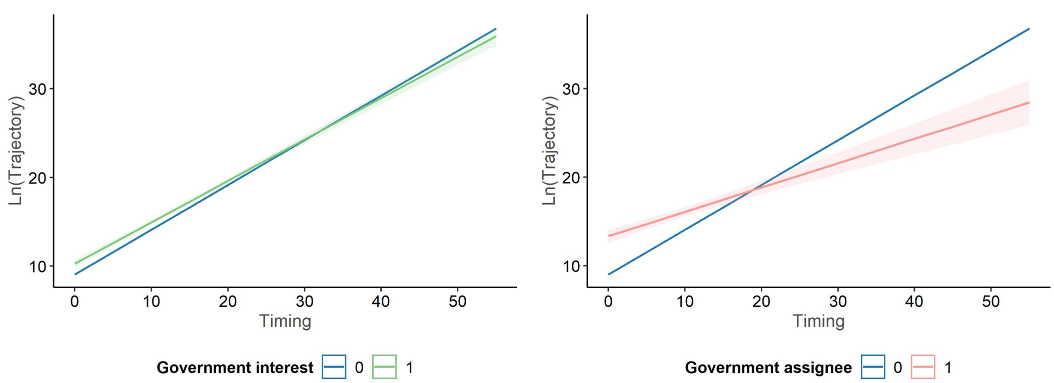


Figure D1: Timing of government funding. Marginal effects – with 95% confidential intervals – of government funding on the trajectory indicator (log) at different levels of the variable timing. Patents supported by government funding are in orange, while all the other patents in blue. Predictions are retrieved by specification (3) in Table 3.

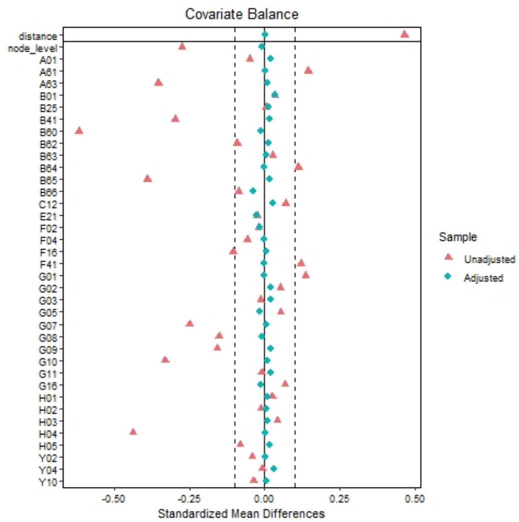


(a) Government interest

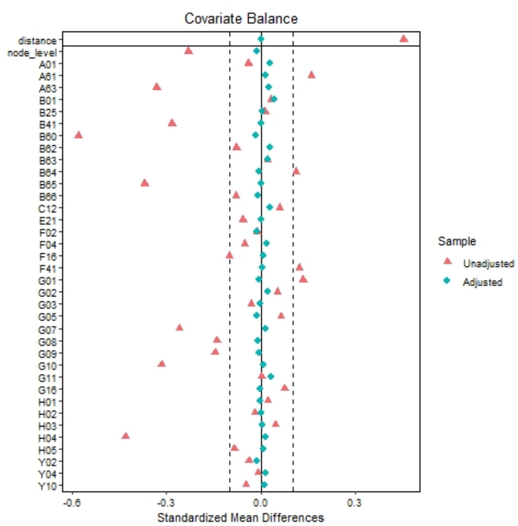
(b) Government assignee

Figure D2: Timing of government-backed patents. Marginal effects – with 95% confidential intervals – of government funding on the trajectory indicator (log) at different levels of the variable timing. Patents supported by federal contractors (government interest) are in green, those with a federal agency or state department as assignee are in red (government assignee), while all the other patents in blue. Predictions are retrieved by specifications (4) and (5) in Table 4.

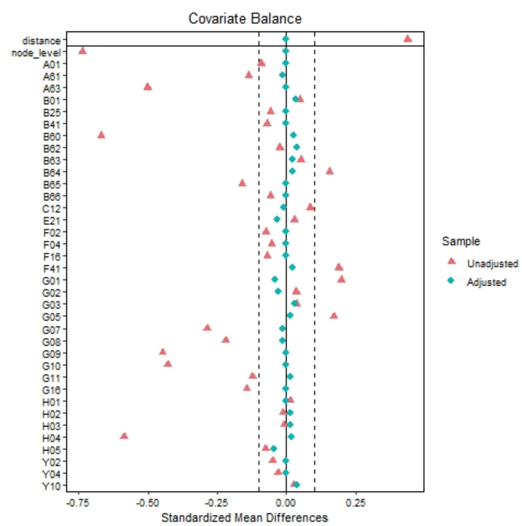
D.2 Propensity score matching



(a) Government funding



(b) Government interest



(c) Government assignee

Figure D3: Covariate balance before (red triangles) and after (green rhombuses) the 1-1 propensity score matching without replacement.

D.3 Instrumental variable

To address possible selection bias in our estimations, we design a quasi experiment based on the use of an instrumental variable. More specifically, we instrument government funding by using the predicted number of patents connected to defense R&D.

We identify patents related to defense R&D by selecting USPTO patents that received government funding from the US Department of Defense or have this department (or one of its divisions, such as Army, Navy, or Air Force) as assignee. Each patent related to defense R&D is then associated to 4-digit CPC classes. Since each patent may be associated with more than one CPC class, we introduce weights proportional to the importance of these classes in the patent. Then, we compute the weighted number of patents related to the US Department of Defense for each 4-digit CPC class. To obtain results that are comparable over time, we normalized the number of patents associated to defense R&D in each CPC class by the total number of patents in that class. The resulting indicator can be interpreted as a measure of the importance of defense R&D in each 4-digit CPC class. Moreover, since we are interested in capturing the predicted number of patents, we introduce a one-year lag. Therefore, for each 4-digit CPC class i at the time t, we compute:

$$\text{Predicted defense patents in CPC}_{i,t} = \frac{\text{Number of defense-related patents}_{i,t-1}}{\text{Number of patents}_{i,t-1}}. \tag{8}$$

Then, we define the instrumental variable *Predicted defense patents*$_{p,t}$ for each patent p with application year t as the weighted average of *Predicted defense patents in* $CPC_{i,t}$ over the collection CPC_p of 4-digit CPC classes related to the patent:

$$\text{Predicted defense patents}_{p,t} = \sum_{i \in CPC_p} \text{share}_i \cdot \text{Predicted defense patents in CPC}_{i,t}, \tag{9}$$

where share_i is the weight of each 4-digit CPC i connected to the patent.

Table D1 reports the results of the first-stage estimations.

Table D1: First stage - Instrumental variable

	Dependent variable:		
	Government funding	Government interest	Government assignee
	(1)	(2)	(3)
Predicted defense patents	1.359***	1.165***	0.610***
	(0.101)	(0.094)	(0.065)
US university	0.427***	0.428***	−0.004***
	(0.009)	(0.009)	(0.001)
Timing	−0.0003***	−0.0002***	−0.0003***
	(0.00004)	(0.00004)	(0.00002)
Number of claims	−0.0001	−0.00002	−0.0002***
	(0.00004)	(0.00004)	(0.00002)
Number of inventors	0.001***	0.001***	−0.001***
	(0.0003)	(0.0002)	(0.0001)
Intercept	0.015***	0.011***	0.013***
	(0.002)	(0.002)	(0.001)
3-digit CPC	Yes	Yes	Yes
Observations	114,670	114,670	114,670
R^2	0.160	0.170	0.016
Adjusted R^2	0.159	0.169	0.016
Residual Std. Error	0.167	0.159	0.089
F Statistic	495.199***	532.877***	43.222***

Note: All the models are estimated using OLS.
Robust standard errors are reported in parenthesis.
Legend: *p<0.1; **p<0.05; ***p<0.01

D.4 Additional robustness checks

Tables D2 and D3 present the results of econometric estimations with an alternative dependent variable: the longest path length associate to the patent. In more detail, for each node p, we construct the sub-graph that includes the node p and all its ancestors and descendants. In such a graph, all possible paths from sources to sinks must be through p by construction. We then compute the longest path in the sub-graph considering edge weight w_{uv}, as defined in Section 4. Finally, we associate the length of this path to the node p.

Table D2: Longest path length. Impact of government funding on the longest path length through patents.

	Dependent variable:		
	log(Longest path length)		
	(1)	(2)	(3)
Government funding	2.275***	2.057***	3.350***
	(0.191)	(0.206)	(0.460)
Government funding*Timing			-0.096***
			(0.022)
US university		0.676***	0.690***
		(0.237)	(0.238)
Timing	0.800***	0.800***	0.803***
	(0.004)	(0.004)	(0.004)
Number of claims	0.049***	0.049***	0.049***
	(0.003)	(0.003)	(0.003)
Number of inventors	-0.261***	-0.262***	-0.262***
	(0.018)	(0.018)	(0.018)
Intercept	25.358***	25.353***	25.308***
	(0.118)	(0.118)	(0.119)
3-digit CPC	Yes	Yes	Yes
Observations	114,670	114,670	114,670
R^2	0.428	0.428	0.428
Adjusted R^2	0.428	0.428	0.428
Residual Std. Error	11.299	11.299	11.297
F Statistic	1301.714***	1272.379***	1245.434***

Note: All the models are estimated using OLS.
Robust standard errors are reported in parenthesis.
Legend: *p<0.1; **p<0.05; ***p<0.01

Table D3: Longest path length. Impact of different government funding sources (through government assignee or grants - government interest statement) on the longest path length through patents.

	Dependent variable:					
	log(Longest path length)					
	(1)	(2)	(3)	(4)	(5)	(6)
Government interest	1.976***		0.839***	1.833***	0.848***	1.734***
	(0.198)		(0.226)	(0.503)	(0.226)	(0.521)
Government interest*Timing				-0.069***		-0.062**
				(0.023)		(0.024)
Government assignee		4.933***	4.439***	4.271***	5.436***	4.872***
		(0.405)	(0.429)	(0.434)	(0.851)	(0.902)
Government assignee*Timing					-0.101**	-0.059
					(0.050)	(0.052)
US university			1.211***	1.194***	1.212***	1.196***
			(0.239)	(0.239)	(0.239)	(0.239)
Timing	0.799***	0.800***	0.801***	0.803***	0.801***	0.803***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Number of claims	0.049***	0.050***	0.049***	0.049***	0.049***	0.049***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Number of inventors	-0.261***	-0.253***	-0.258***	-0.258***	-0.258***	-0.258***
	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)
Intercept	25.386***	25.353***	25.308***	25.281***	25.299***	25.279***
	(0.118)	(0.118)	(0.118)	(0.119)	(0.119)	(0.119)
3-digit CPC	Yes	Yes	Yes	Yes	Yes	Yes
Observations	114,670	114,670	114,670	114,670	114,670	114,670
R^2	0.428	0.428	0.428	0.428	0.428	0.428
Adjusted R^2	0.427	0.428	0.428	0.428	0.428	0.428
Residual Std. Error	11.301	11.298	11.295	11.294	11.295	11.294
F Statistic	1300.764***	1301.629***	1245.419***	1219.045***	1218.476***	1193.123***

Note: All the models are estimated using OLS.
Robust standard errors are reported in parenthesis.
Legend: *p<0.1; **p<0.05; ***p<0.01

Tables D4 and D5 summarize the results when we replace direct indicators of government funding with indirect ones used in the previous literature. In particular, we include independent dummy variables that detect patents that cite government-funded inventions. In the same vein, we replace the control *US university* with the dummy variable *citing US university*.

Table D4: Citing government funding. Influence on the trajectory of citing government funded patents.

	Dependent variable:		
	log(Trajectory)		
	(1)	(2)	(3)
Citing government funding	1.204***	0.995***	1.303***
	(0.046)	(0.056)	(0.110)
Timing:citing government funding			-0.018***
			(0.004)
Citing US university		0.412***	0.428***
		(0.056)	(0.056)
Timing	0.494***	0.493***	0.498***
	(0.002)	(0.002)	(0.002)
Number of claims	0.039***	0.038***	0.038***
	(0.002)	(0.002)	(0.002)
Number of inventors	-0.113***	-0.114***	-0.114***
	(0.011)	(0.011)	(0.011)
Intercept	8.536***	8.529***	8.454***
	(0.077)	(0.077)	(0.081)
3-digit CPC	Yes	Yes	Yes
Observations	114,670	114,670	114,670
R^2	0.437	0.438	0.438
Adjusted R^2	0.437	0.437	0.438
Residual Std. Error	7.276	7.275	7.274
F Statistic	3095.255***	3028.102***	2980.272***

Note: All the models are estimated using OLS.
Robust standard errors are reported in parenthesis.
Legend: *p<0.1; **p<0.05; ***p<0.01

Table D5: Citing government funding. Influence on the trajectory of patents citing inventions with a government interest statement or government assignees.

	Dependent variable:					
	log(Trajectory)					
	(1)	(2)	(3)	(4)	(5)	(6)
Citing gov interest	1.128***		0.498***	0.607***	0.481***	0.260**
	(0.047)		(0.059)	(0.114)	(0.059)	(0.124)
Timing:citing gov interest				-0.006		0.012***
				(0.004)		(0.005)
Citing gov assignee		1.781***	1.358***	1.354***	2.209***	2.347***
		(0.066)	(0.072)	(0.072)	(0.153)	(0.168)
Timing:citing gov assignee					-0.048***	-0.056***
					(0.006)	(0.007)
Citing US university			0.417***	0.422***	0.448***	0.443***
			(0.057)	(0.057)	(0.057)	(0.057)
Timing	0.495***	0.499***	0.493***	0.495***	0.499***	0.496***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Number of claims	0.039***	0.039***	0.037***	0.037***	0.037***	0.037***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Number of inventors	-0.112***	-0.109***	-0.115***	-0.114***	-0.114***	-0.114***
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Intercept	8.567***	8.617***	8.562***	8.539***	8.470***	8.503***
	(0.078)	(0.077)	(0.077)	(0.081)	(0.079)	(0.081)
3-digit CPC	Yes	Yes	Yes	Yes	Yes	Yes
Observations	114,670	114,670	114,670	114,670	114,670	114,670
R^2	0.437	0.438	0.439	0.439	0.439	0.439
Adjusted R^2	0.437	0.438	0.439	0.439	0.439	0.439
Residual Std. Error	7.279	7.273	7.266	7.266	7.264	7.264
F Statistic	3092.791***	3102.593***	2980.869***	2949.544***	2919.599***	2889.547***

Note: All the models are estimated using OLS.
Robust standard errors are reported in parenthesis.
Legend: *p<0.1; **p<0.05; ***p<0.01

We also estimate the effect of government-funded patents in AI on the number of citations, a widespread measure of patent relevance. Since the estimates of the effect of government funding on the number of citations can be affected by endogeneity issues, such as reverse causality and selection bias, we present the results based on quasi-experimental designs: results of the 1-1 matching procedure are in Tables D6, D7, and D8, while those resulting from an instrumental variable approach are in Tables D9, D10, and D11. We count the number of citations in six different ways: considering only citations from patents in our network, considering citations from all patents in our original sample (USPTO patents granted after 1976), considering citations received up to five years after the earliest date of publication (both in the network and in the entire sample), and considering only citations in five years by excluding patents after 2012 since we do not have a complete record of citations for them (both in the network and in the entire sample). We remove from controls the timing because it is not necessary for this kind of analysis, which can be agnostic of the network. We replace this indicator of time evolution with the application year (closer to the time of invention than the grant year).

Table D6: Number of citations. Influence of government funding on the number of citations - Matching 1-1 without replacement (PS)

	Dependent variable:					
	log(Number of citations)					
	network	all	network up to 5 years	all up to 5 years	network 5 years	all 5 years
	(1)	(2)	(3)	(4)	(5)	(6)
Government funding	−0.288***	−0.537***	−0.215***	−0.483***	0.049	−0.110***
	(0.030)	(0.029)	(0.026)	(0.028)	(0.031)	(0.031)
US university	0.203***	0.246***	0.148***	0.208***	0.143***	0.163***
	(0.039)	(0.039)	(0.035)	(0.038)	(0.043)	(0.043)
Number of claims	0.014***	0.019***	0.011***	0.017***	0.010***	0.015***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Number of inventors	0.050***	0.066***	0.044***	0.062***	0.062***	0.083***
	(0.008)	(0.007)	(0.007)	(0.007)	(0.008)	(0.008)
Application year	−0.056***	−0.089***	−0.004***	−0.020***	0.019***	0.020***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Constant	114.251***	181.023***	9.002***	42.182***	−37.181***	−38.520***
	(3.871)	(4.308)	(3.025)	(3.637)	(3.754)	(3.859)
3-digit CPC	Yes	Yes	Yes	Yes	Yes	Yes
Observations	7,864	7,864	7,864	7,864	6,140	6,140
R^2	0.184	0.305	0.093	0.127	0.115	0.141
Adjusted R^2	0.180	0.301	0.088	0.123	0.109	0.136
Residual Std. Error	1.192	1.155	1.053	1.101	1.041	1.025
F Statistic	43.066***	83.676***	19.565***	27.871***	19.727***	25.123***

Note: All the models are estimated using OLS on matched patents (1-1 propensity score matching). Robust standard errors are reported in parenthesis.
Legend: *p<0.1; **p<0.05; ***p<0.01

Table D7: Number of citations. Influence of government interest on the number of citations - Matching 1-1 without replacement (PS)

	Dependent variable:					
	log(Number of citations)					
	network	all	network up to 5 years	all up to 5 years	network 5 years	all 5 years
	(1)	(2)	(3)	(4)	(5)	(6)
Government interest	−0.270***	−0.500***	−0.204***	−0.457***	0.060*	−0.086***
	(0.032)	(0.031)	(0.028)	(0.029)	(0.033)	(0.033)
US university	0.186***	0.232***	0.135***	0.201***	0.123***	0.146***
	(0.040)	(0.040)	(0.036)	(0.039)	(0.044)	(0.044)
Number of claims	0.014***	0.018***	0.011***	0.016***	0.010***	0.015***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Number of inventors	0.057***	0.073***	0.049***	0.067***	0.064***	0.088***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.009)	(0.008)
Application year	−0.060***	−0.095***	−0.006***	−0.025***	0.020***	0.019***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Constant	121.321***	193.715***	12.995***	50.948***	−39.505***	−36.635***
	(4.248)	(4.706)	(3.284)	(3.956)	(4.142)	(4.272)
3-digit CPC	Yes	Yes	Yes	Yes	Yes	Yes
Observations	7,194	7,194	7,194	7,194	5,530	5,530
R^2	0.192	0.316	0.093	0.126	0.111	0.130
Adjusted R^2	0.188	0.313	0.088	0.121	0.104	0.123
Residual Std. Error	1.193	1.155	1.061	1.107	1.051	1.033
F Statistic	41.492***	80.773***	17.843***	25.215***	16.713***	19.988***

Note: All the models are estimated using OLS on matched patents (1-1 propensity score matching). Robust standard errors are reported in parenthesis.
Legend: *p<0.1; **p<0.05; ***p<0.01

Table D8: Number of citations. Influence of government assignee on the number of citations - Matching 1-1 without replacement (PS)

	Dependent variable:					
	log(Number of citations)					
	network	all	network up to 5 years	all up to 5 years	network 5 years	all 5 years
	(1)	(2)	(3)	(4)	(5)	(6)
Government assignee	−0.378***	−0.646***	−0.264***	−0.550***	−0.118**	−0.311***
	(0.054)	(0.054)	(0.046)	(0.051)	(0.053)	(0.057)
US university	0.481***	0.448***	0.320**	0.389**	0.305*	0.175
	(0.139)	(0.153)	(0.141)	(0.162)	(0.166)	(0.184)
Number of claims	0.007***	0.013***	0.004*	0.011***	0.003	0.009***
	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	(0.003)
Number of inventors	0.051***	0.059***	0.047***	0.057***	0.061***	0.068***
	(0.017)	(0.015)	(0.015)	(0.015)	(0.016)	(0.016)
Application year	−0.049***	−0.070***	−0.002	−0.006*	0.009***	0.016***
	(0.003)	(0.004)	(0.003)	(0.003)	(0.003)	(0.004)
Constant	99.718***	142.174***	4.911	12.860**	−17.129***	−31.203***
	(6.771)	(7.496)	(5.225)	(6.334)	(6.300)	(7.082)
3-digit CPC	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,858	1,858	1,858	1,858	1,636	1,636
R^2	0.176	0.277	0.099	0.153	0.113	0.158
Adjusted R^2	0.160	0.262	0.081	0.136	0.092	0.139
Residual Std. Error	1.131	1.107	0.960	1.021	0.959	0.986
F Statistic	10.533***	18.856***	5.422***	8.877***	5.494***	8.129***

Note: All the models are estimated using OLS on matched patents (1-1 propensity score matching). Robust standard errors are reported in parenthesis.
Legend: *p<0.1; **p<0.05; ***p<0.01

Table D9: Number of citations. Influence of government funding on the number of citations - Instrumental variable

	Dependent variable:					
	log(Number of citations)					
	network	all	network up to 5 years	all up to 5 years	network 5 years	all 5 years
	(1)	(2)	(3)	(4)	(5)	(6)
Government funding	−3.544***	−6.389***	−1.754***	−4.668***	−2.165***	−5.773***
	(0.366)	(0.525)	(0.241)	(0.397)	(0.345)	(0.647)
US university	1.712***	2.950***	0.916***	2.190***	1.024***	2.461***
	(0.163)	(0.235)	(0.107)	(0.178)	(0.144)	(0.270)
Number of claims	0.015***	0.020***	0.012***	0.019***	0.010***	0.015***
	(0.001)	(0.001)	(0.0004)	(0.001)	(0.0005)	(0.001)
Number of inventors	0.042***	0.056***	0.037***	0.051***	0.052***	0.071***
	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)	(0.003)
Application year	−0.082***	−0.118***	−0.027***	−0.049***	0.007***	0.009***
	(0.001)	(0.001)	(0.0004)	(0.001)	(0.001)	(0.001)
Intercept	165.407***	237.920***	54.039***	99.267***	−13.276***	−15.873***
	(1.271)	(1.662)	(0.904)	(1.337)	(1.245)	(1.890)
3-digit CPC	Yes	Yes	Yes	Yes	Yes	Yes
Observations	114,670	114,670	114,670	114,670	63,100	63,100
F-test	170.1***	170.1***	170.06***	170.1***	67.64***	67.64***

Note: All the models are estimated using 2SLS.
Robust standard errors are reported in parenthesis.
Legend: *p<0.1; **p<0.05; ***p<0.01

Table D10: Number of citations. Influence of government interest on the number of citations - Instrumental variable

	Dependent variable:					
	log(Number of citations)					
	network	all	network up to 5 years	all up to 5 years	network 5 years	all 5 years
	(1)	(2)	(3)	(4)	(5)	(6)
Government interest	−4.064***	−7.327***	−2.012***	−5.353***	−2.752***	−7.339***
	(0.444)	(0.638)	(0.288)	(0.485)	(0.464)	(0.895)
US university	1.939***	3.361***	1.029***	2.490***	1.265***	3.105***
	(0.197)	(0.285)	(0.127)	(0.216)	(0.192)	(0.373)
Number of claims	0.015***	0.021***	0.012***	0.019***	0.010***	0.015***
	(0.001)	(0.001)	(0.0004)	(0.001)	(0.0005)	(0.001)
Number of inventors	0.043***	0.058***	0.038***	0.053***	0.053***	0.075***
	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)	(0.004)
Application year	−0.081***	−0.116***	−0.026***	−0.047***	0.008***	0.011***
	(0.001)	(0.001)	(0.0004)	(0.001)	(0.001)	(0.001)
Intercept	163.351***	234.213***	53.021***	96.559***	−14.764***	−19.840***
	(1.214)	(1.603)	(0.861)	(1.290)	(1.213)	(1.947)
3-digit CPC	Yes	Yes	Yes	Yes	Yes	Yes
Observations	114,670	114,670	114,670	114,670	63,100	63,100
F-test	148.8***	148.8***	148.77***	148.8***	54.36***	54.36***

Note: All the models are estimated using 2SLS.
Robust standard errors are reported in parenthesis.
Legend: *p<0.1; **p<0.05; ***p<0.01

Table D11: Number of citations. Influence of government assignee on the number of citations - Instrumental variable

	Dependent variable:					
	log(Number of citations)					
	network	all	network up to 5 years	all up to 5 years	network 5 years	all 5 years
	(1)	(2)	(3)	(4)	(5)	(6)
Government assignee	-8.343^{***}	-15.043^{***}	-4.130^{***}	-10.991^{***}	-3.440^{***}	-9.175^{***}
	(1.095)	(1.795)	(0.642)	(1.327)	(0.603)	(1.295)
US university	0.164^{***}	0.161^{***}	0.150^{***}	0.152^{***}	0.125^{***}	0.064^{**}
	(0.025)	(0.032)	(0.021)	(0.027)	(0.026)	(0.030)
Number of claims	0.014^{***}	0.018^{***}	0.012^{***}	0.017^{***}	0.009^{***}	0.014^{***}
	(0.001)	(0.001)	(0.0004)	(0.001)	(0.0005)	(0.001)
Number of inventors	0.033^{***}	0.041^{***}	0.033^{***}	0.040^{***}	0.046^{***}	0.057^{***}
	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)	(0.003)
Application year	-0.085^{***}	-0.124^{***}	-0.028^{***}	-0.054^{***}	0.006^{***}	0.004^{***}
	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
Intercept	172.605^{***}	250.898^{***}	57.602^{***}	108.749^{***}	-9.995^{***}	-7.122^{***}
	(2.010)	(3.049)	(1.284)	(2.316)	(1.562)	(2.713)
3-digit CPC	Yes	Yes	Yes	Yes	Yes	Yes
Observations	114,670	114,670	114,670	114,670	63,100	63,100
F-test	75.1^{***}	75.1^{***}	75.1^{***}	75.1^{***}	47.89^{***}	47.89^{***}

Note: All the models are estimated using 2SLS.
Robust standard errors are reported in parenthesis.
Legend: *p<0.1; **p<0.05; ***p<0.01

To further test the relevance of funding timing, we introduce the forward trajectory indicator w_p^+, where p is a patent, defined in Equation 2. Results are reported in Tables D12 and D13.

Table D12: Forward trajectory indicator. Influence of government funding on the forward trajectory.

	Dependent variable:		
	log(Forward trajectory)		
	(1)	(2)	(3)
Government funding	1.243***	1.128***	2.001***
	(0.131)	(0.146)	(0.266)
Government funding*Timing			-0.065***
			(0.011)
US university		0.356**	0.366**
		(0.164)	(0.164)
Timing	-0.204***	-0.204***	-0.202***
	(0.002)	(0.002)	(0.002)
Number of claims	0.037***	0.037***	0.037***
	(0.002)	(0.002)	(0.002)
Number of inventors	-0.131***	-0.132***	-0.132***
	(0.011)	(0.011)	(0.011)
Intercept	8.456***	8.454***	8.423***
	(0.078)	(0.078)	(0.078)
3-digit CPC	Yes	Yes	Yes
Observations	114,670	114,670	114,670
R^2	0.150	0.150	0.150
Adjusted R^2	0.150	0.150	0.150
Residual Std. Error	7.236	7.236	7.235
F Statistic	401.026***	392.002***	384.002***

Note: All the models are estimated using OLS.
Robust standard errors are reported in parenthesis.
Legend: *p<0.1; **p<0.05; ***p<0.01

Table D13: Forward trajectory indicator. Influence on the forward trajectory of patents with a government interest statement or government assignees.

	Dependent variable:					
	log(Forward trajectory)					
	(1)	(2)	(3)	(4)	(5)	(6)
Government interest	1.028***		0.428***	0.958***	0.446***	0.618**
	(0.132)		(0.155)	(0.283)	(0.155)	(0.291)
Government interest*Timing				-0.037***		-0.012
				(0.011)		(0.012)
Government assignee		2.542***	2.291***	2.201***	4.365***	4.255***
		(0.319)	(0.336)	(0.337)	(0.540)	(0.565)
Government assignee*Timing					-0.210***	-0.202***
					(0.029)	(0.030)
US university			0.663***	0.654***	0.663***	0.660***
			(0.166)	(0.166)	(0.166)	(0.166)
Timing	-0.204***	-0.204***	-0.203***	-0.202***	-0.202***	-0.202***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Number of claims	0.037***	0.037***	0.037***	0.037***	0.037***	0.037***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Number of inventors	-0.131***	-0.127***	-0.130***	-0.130***	-0.130***	-0.130***
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Intercept	8.473***	8.457***	8.433***	8.418***	8.415***	8.411***
	(0.078)	(0.078)	(0.078)	(0.078)	(0.078)	(0.078)
3-digit CPC	Yes	Yes	Yes	Yes	Yes	Yes
Observations	114,670	114,670	114,670	114,670	114,670	114,670
R^2	0.150	0.150	0.150	0.150	0.151	0.151
Adjusted R^2	0.149	0.150	0.150	0.150	0.150	0.150
Residual Std. Error	7.237	7.236	7.234	7.234	7.233	7.233
F Statistic	400.513***	400.164***	383.594***	375.723***	375.500***	367.933***

Note: All the models are estimated using OLS.
Robust standard errors are reported in parenthesis.
Legend: *p<0.1; **p<0.05; ***p<0.01

Tables D14 and D15 present results for a sub-sample of patents. We select patents using only the criteria suggested by the WIPO (2019) report on artificial intelligence inventions. The sample includes 111,525 patents. In Tables D16 and D17, instead, we select only patents granted after 1980 (not included).

Table D14: Sample selection. Influence of government funding on the trajectory. Patents belong to a sub-sample of AI inventions.

	Dependent variable:		
	log(Trajectory)		
	(1)	(2)	(3)
Government funding	1.114***	1.040***	1.860***
	(0.134)	(0.148)	(0.265)
Government funding*Timing			-0.061***
			(0.011)
US university		0.230	0.235
		(0.167)	(0.167)
Timing	0.500***	0.500***	0.502***
	(0.002)	(0.002)	(0.002)
Number of claims	0.042***	0.042***	0.042***
	(0.002)	(0.002)	(0.002)
Number of inventors	-0.103***	-0.103***	-0.103***
	(0.012)	(0.012)	(0.012)
Intercept	8.772***	8.770***	8.741***
	(0.080)	(0.079)	(0.080)
3-digit CPC	Yes	Yes	Yes
Observations	111,525	111,525	111,525
R^2	0.432	0.432	0.432
Adjusted R^2	0.432	0.432	0.432
Residual Std. Error	7.331	7.331	7.330
F Statistic	2984.451***	2916.457***	2861.327***

Note: All the models are estimated using OLS.
Robust standard errors are reported in parenthesis.
Legend: *p<0.1; **p<0.05; ***p<0.01

Table D15: Sample selection. Influence on the trajectory of patents with a government interest statement or government assignees. Patents belong to a sub-sample of AI inventions.

	Dependent variable:					
	log(Trajectory)					
	(1)	(2)	(3)	(4)	(5)	(6)
Government interest	0.915***		0.415***	0.905***	0.436***	0.523*
	(0.135)		(0.158)	(0.282)	(0.157)	(0.290)
Government interest*Timing				-0.034***		-0.006
				(0.012)		(0.012)
Government assignee		2.257***	2.012***	1.929***	4.274***	4.218***
		(0.323)	(0.341)	(0.342)	(0.541)	(0.565)
Government assignee*Timing					-0.230***	-0.225***
					(0.030)	(0.031)
US university			0.502***	0.491***	0.502***	0.500***
			(0.168)	(0.169)	(0.168)	(0.169)
Timing	0.499***	0.500***	0.500***	0.501***	0.501***	0.501***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Number of claims	0.042***	0.043***	0.043***	0.043***	0.043***	0.043***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Number of inventors	-0.103***	-0.099***	-0.101***	-0.101***	-0.102***	-0.102***
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
Intercept	8.788***	8.773***	8.752***	8.739***	8.732***	8.730***
	(0.080)	(0.080)	(0.080)	(0.080)	(0.080)	(0.080)
3-digit CPC	Yes	Yes	Yes	Yes	Yes	Yes
Observations	111,525	111,525	111,525	111,525	111,525	111,525
R^2	0.432	0.432	0.432	0.432	0.432	0.432
Adjusted R^2	0.431	0.432	0.432	0.432	0.432	0.432
Residual Std. Error	7.332	7.331	7.330	7.330	7.328	7.328
F Statistic	2981.144***	2985.782***	2854.623***	2798.106***	2803.314***	2745.277***

Note: All the models are estimated using OLS.
Robust standard errors are reported in parenthesis.
Legend: *p<0.1; **p<0.05; ***p<0.01

Table D16: Patents granted after 1980. Influence of government funding on the trajectory. We select only patents granted after 1980.

	Dependent variable:		
	log(Trajectory)		
	(1)	(2)	(3)
Government funding	1.104***	1.008***	1.707***
	(0.130)	(0.144)	(0.257)
Government funding*Timing			-0.051***
			(0.011)
US university		0.295*	0.298*
		(0.162)	(0.162)
Timing	0.516***	0.516***	0.517***
	(0.002)	(0.002)	(0.002)
Number of claims	0.046***	0.046***	0.046***
	(0.002)	(0.002)	(0.002)
Number of inventors	-0.092***	-0.093***	-0.093***
	(0.011)	(0.011)	(0.011)
Intercept	8.168***	8.165***	8.141***
	(0.077)	(0.077)	(0.077)
3-digit CPC	Yes	Yes	Yes
Observations	113,835	113,835	113,835
R^2	0.453	0.453	0.453
Adjusted R^2	0.452	0.452	0.452
Residual Std. Error	7.144	7.144	7.143
F Statistic	3353.781***	3277.453***	3214.805***

Note: All the models are estimated using OLS.
Robust standard errors are reported in parenthesis.
Legend: *p<0.1; **p<0.05; ***p<0.01

Table D17: Patents granted after 1980. Influence on the trajectory of patents with a government interest statement or government assignees. We select only patents granted after 1980.

	Dependent variable:					
	log(Trajectory)					
Government interest	0.954***		0.478***	0.942***	0.489***	0.612**
	(0.132)		(0.153)	(0.273)	(0.153)	(0.280)
Government interest*Timing				-0.032***		-0.008
				(0.011)		(0.012)
Government assignee		2.108***	1.821***	1.743***	3.837***	3.756***
		(0.319)	(0.336)	(0.337)	(0.541)	(0.564)
Government assignee*Timing					-0.197***	-0.191***
					(0.029)	(0.030)
US university			0.528***	0.518***	0.531***	0.528***
			(0.164)	(0.164)	(0.164)	(0.164)
Timing	0.515***	0.516***	0.516***	0.517***	0.517***	0.517***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Number of claims	0.046***	0.047***	0.046***	0.046***	0.046***	0.046***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Number of inventors	-0.092***	-0.089***	-0.091***	-0.091***	-0.091***	-0.091***
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Intercept	8.181***	8.172***	8.149***	8.137***	8.133***	8.130***
	(0.077)	(0.077)	(0.077)	(0.077)	(0.077)	(0.077)
3-digit CPC	Yes	Yes	Yes	Yes	Yes	Yes
Observations	113,835	113,835	113,835	113,835	113,835	113,835
R^2	0.452	0.452	0.453	0.453	0.453	0.453
Adjusted R^2	0.452	0.452	0.452	0.452	0.453	0.453
Residual Std. Error	7.145	7.144	7.143	7.143	7.142	7.142
F Statistic	3350.587***	3354.324***	3207.717***	3144.336***	3148.587***	3083.619***

Note: All the models are estimated using OLS.
Robust standard errors are reported in parenthesis.
Legend: *p<0.1; **p<0.05; ***p<0.01

We also change regression controls to test the robustness of results. Tables D18 and D19 show estimates in which we control for worldwide-university assignees instead of US-university assignees. Finally, Tables D20 and D21 present the effect of government funding on the trajectory when we control also for the number of patents' backward citations.

Table D18: University. Influence of government funding on the trajectory when we control for inventions assigned to worldwide universities.

	Dependent variable:		
	log(Trajectory)		
	(1)	(2)	(3)
Government funding	1.184***	1.486***	2.335***
	(0.132)	(0.139)	(0.257)
Government funding*Timing			-0.063***
			(0.011)
University		-0.988***	-0.982***
		(0.114)	(0.114)
Timing	0.503***	0.503***	0.505***
	(0.002)	(0.002)	(0.002)
Number of claims	0.043***	0.043***	0.043***
	(0.002)	(0.002)	(0.002)
Number of inventors	-0.106***	-0.101***	-0.101***
	(0.011)	(0.011)	(0.011)
Intercept	8.594***	8.620***	8.590***
	(0.078)	(0.078)	(0.078)
3-digit CPC	Yes	Yes	Yes
Observations	114,670	114,670	114,670
R^2	0.435	0.435	0.435
Adjusted R^2	0.435	0.435	0.435
Residual Std. Error	7.292	7.290	7.289
F Statistic	3078.115***	3017.297***	2958.976***

Note: All the models are estimated using OLS.
Robust standard errors are reported in parenthesis.
Legend: *p<0.1; **p<0.05; ***p<0.01

Table D19: University. Influence on the trajectory of patents with a government interest statement or government assignees when we control for inventions assigned to worldwide universities.

	Dependent variable:					
	log(Trajectory)					
	(1)	(2)	(3)	(4)	(5)	(6)
Government interest	0.983***		0.996***	1.565***	1.016***	1.192***
	(0.134)		(0.146)	(0.272)	(0.146)	(0.279)
Government interest*Timing				-0.040***		-0.012
				(0.011)		(0.012)
Government assignee		2.322***	1.699***	1.604***	3.971***	3.859***
		(0.321)	(0.338)	(0.339)	(0.537)	(0.561)
Government assignee*Timing					-0.230***	-0.222***
					(0.030)	(0.031)
University			-0.851***	-0.857***	-0.850***	-0.852***
			(0.115)	(0.115)	(0.115)	(0.115)
Timing	0.503***	0.503***	0.503***	0.504***	0.504***	0.505***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Number of claims	0.043***	0.043***	0.043***	0.043***	0.043***	0.043***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Number of inventors	-0.106***	-0.102***	-0.100***	-0.100***	-0.100***	-0.100***
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Intercept	8.610***	8.597***	8.608***	8.592***	8.588***	8.584***
	(0.078)	(0.078)	(0.078)	(0.078)	(0.078)	(0.079)
3-digit CPC	Yes	Yes	Yes	Yes	Yes	Yes
Observations	114,670	114,670	114,670	114,670	114,670	114,670
R^2	0.435	0.435	0.435	0.435	0.436	0.436
Adjusted R^2	0.435	0.435	0.435	0.435	0.435	0.435
Residual Std. Error	7.294	7.293	7.290	7.290	7.288	7.288
F Statistic	3074.472***	3078.966***	2950.681***	2891.702***	2897.804***	2837.505***

Note: All the models are estimated using OLS.
Robust standard errors are reported in parenthesis.
Legend: *p<0.1; **p<0.05; ***p<0.01

Table D20: Number of backward citations. Influence of government funding on the trajectory when we control for the number of backward citations.

	Dependent variable:		
	log(Trajectory)		
	(1)	(2)	(3)
Government funding	1.197***	1.104***	1.948***
	(0.132)	(0.147)	(0.263)
Government funding*Timing			-0.063***
			(0.011)
US university		0.291*	0.300*
		(0.166)	(0.166)
Timing	0.499***	0.499***	0.501***
	(0.002)	(0.002)	(0.002)
Number of claims	0.041***	0.041***	0.041***
	(0.002)	(0.002)	(0.002)
Number of inventors	-0.110***	-0.110***	-0.110***
	(0.011)	(0.011)	(0.011)
Number of backward citations	0.003***	0.003***	0.003***
	(0.000)	(0.000)	(0.000)
Intercept	8.642***	8.640***	8.610***
	(0.078)	(0.078)	(0.078)
3-digit CPC	Yes	Yes	Yes
Observations	114,670	114,670	114,670
R^2	0.436	0.436	0.436
Adjusted R^2	0.435	0.435	0.435
Residual Std. Error	7.288	7.288	7.287
F Statistic	3012.269***	2945.150***	2891.006***

Note: All the models are estimated using OLS.
Robust standard errors are reported in parenthesis.
Legend: *p<0.1; **p<0.05; ***p<0.01

Table D21: Number of backward citations. Influence on the trajectory of patents with a government interest statement or government assignees when we control for the number of backward citations.

	Dependent variable: log(Trajectory)					
	(1)	(2)	(3)	(4)	(5)	(6)
Government interest	0.997***		0.467***	0.984***	0.487***	0.611**
	(0.134)		(0.157)	(0.281)	(0.156)	(0.288)
Government interest*Timing				-0.036***		-0.009
				(0.012)		(0.012)
Government assignee		2.333***	2.057***	1.969***	4.303***	4.224***
		(0.321)	(0.338)	(0.340)	(0.537)	(0.562)
Government assignee*Timing					-0.228***	-0.222***
					(0.030)	(0.031)
US university			0.570***	0.561***	0.570***	0.568***
			(0.167)	(0.168)	(0.167)	(0.168)
Timing	0.499***	0.499***	0.500***	0.501***	0.501***	0.501***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Number of claims	0.041***	0.042***	0.041***	0.041***	0.041***	0.041***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Number of inventors	-0.109***	-0.106***	-0.108***	-0.108***	-0.108***	-0.108***
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Number of backward citations	0.003***	0.003***	0.003***	0.003***	0.003***	0.003***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Intercept	8.658***	8.645***	8.622***	8.608***	8.603***	8.600***
	(0.078)	(0.078)	(0.078)	(0.078)	(0.078)	(0.078)
3-digit CPC	Yes	Yes	Yes	Yes	Yes	Yes
Observations	114,670	114,670	114,670	114,670	114,670	114,670
R^2	0.435	0.436	0.436	0.436	0.436	0.436
Adjusted R^2	0.435	0.435	0.435	0.436	0.436	0.436
Residual Std. Error	7.290	7.289	7.287	7.287	7.285	7.285
F Statistic	3008.681***	3012.962***	2883.873***	2828.293***	2833.315***	2776.047***

Note: All the models are estimated using OLS.
Robust standard errors are reported in parenthesis.
Legend: *p<0.1; **p<0.05; ***p<0.01

Table D22: Weighted average of lagged 3-digit CPC growth. Influence of government funding on the trajectory when we control for weighted average of one-year lagged 3-digit CPC growth over the previous three years.

	Dependent variable:		
	log(Trajectory)		
	(1)	(2)	(3)
Government funding	1.061***	1.000***	1.923***
	(0.130)	(0.144)	(0.258)
Government funding*Timing			-0.068***
			(0.011)
US university		0.188	0.198
		(0.161)	(0.161)
Timing	0.540***	0.540***	0.542***
	(0.002)	(0.002)	(0.002)
Nb claims	0.038***	0.038***	0.038***
	(0.002)	(0.002)	(0.002)
Inventors number	-0.078***	-0.078***	-0.078***
	(0.011)	(0.011)	(0.011)
Avg CPC growth rate$_{t-1}$	6.676***	6.675***	6.678***
	(0.079)	(0.079)	(0.079)
Intercept	7.208***	7.207***	7.174***
	(0.075)	(0.075)	(0.076)
CPC 3d	Yes	Yes	Yes
Observations	114,670	114,670	114,670
R^2	0.461	0.461	0.461
Adjusted R^2	0.460	0.460	0.461
Residual Std. Error	7.125	7.125	7.123
F Statistic	3312.803***	3239.197***	3178.929***

Note: All the models are estimated using OLS.
Robust standard errors are reported in parenthesis.
Legend: *p<0.1; **p<0.05; ***p<0.01

Table D23: Weighted average of lagged 3-digit CPC growth. Influence on the trajectory of patents with a government interest statement or government assignees when we control for weighted average of one-year lagged 3-digit CPC growth over the previous three years.

	Dependent variable:					
	log(Trajectory)					
	(1)	(2)	(3)	(4)	(5)	(6)
Government interest	0.887***		0.454***	1.096***	0.475***	0.707**
	(0.131)		(0.154)	(0.274)	(0.153)	(0.282)
Government interest*Timing				-0.045***		-0.016
				(0.012)		(0.012)
Government assignee		2.034***	1.766***	1.657***	4.155***	4.007***
		(0.319)	(0.336)	(0.337)	(0.537)	(0.562)
Government assignee*Timing					-0.242***	-0.231***
					(0.031)	(0.032)
US university			0.428***	0.416**	0.428***	0.424***
			(0.163)	(0.164)	(0.163)	(0.163)
Timing	0.540***	0.540***	0.540***	0.541***	0.541***	0.542***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Nb claims	0.038***	0.039***	0.038***	0.038***	0.039***	0.039***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Inventors number	-0.078***	-0.074***	-0.077***	-0.077***	-0.077***	-0.077***
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Avg CPC growth rate$_{t-1}$	6.681***	6.675***	6.670***	6.672***	6.674***	6.675***
	(0.079)	(0.079)	(0.079)	(0.079)	(0.079)	(0.079)
Intercept	7.221***	7.211***	7.193***	7.175***	7.171***	7.166***
	(0.075)	(0.075)	(0.075)	(0.076)	(0.075)	(0.076)
CPC 3d	Yes	Yes	Yes	Yes	Yes	Yes
Observations	114,670	114,670	114,670	114,670	114,670	114,670
R^2	0.460	0.461	0.461	0.461	0.461	0.461
Adjusted R^2	0.460	0.460	0.461	0.461	0.461	0.461
Residual Std. Error	7.126	7.125	7.124	7.124	7.122	7.121
F Statistic	3309.556***	3313.671***	3171.309***	3109.185***	3117.085***	3053.601***

Note: All the models are estimated using OLS.
Robust standard errors are reported in parenthesis.
Legend: *p<0.1; **p<0.05; ***p<0.01