HELLENIC FOUNDATION
FOR EUROPEAN & FOREIGN POLICY

ΕΛΙΑΜΕΠ
ELIAMEP

GREEK AND EUROPEAN ECONOMY PROGRAMME

Can Public R&D Drive AI Innovation?

Anastasios EVGENIDIS; Apostolos FASIANOS
& George PAPAPANAGIOTOU

October 2025
Working Paper #137/2025

ELIAMEP | Working Paper # 137/2025

Can Public R&D Drive AI Innovation?

HELLENIC FOUNDATION FOR EUROPEAN & FOREIGN POLICY (ELIAMEP)
49 Vasilissis Sofias Ave., 10676, Athens, Greece
Tel.: +30 210 7257 110 | Fax: +30 210 7257 114 | http://www.eliamep.gr| eliamep@eliamep.gr

***Cover photo**: Unsplash*

Anastasios EVGENIDIS
Research Fellow, ELIAMEP; Senior Lecturer, Royal Holloway University of London

Apostolos FASIANOS
University of Macedonia; Research Fellow, ELIAMEP

George PAPAPANAGIOTOU
University of Macedonia

Abstract

We examine whether public R&D funding stimulates AI innovation in Europe and the UK, using patent data from the European Patent Office (2003–2021). We construct a novel AI indicator based on patent texts and classify innovations using textual analysis. We instrument regional public R&D with plausibly exogenous variation in national defense-related spending. We find that public R&D significantly increases AI patenting: a 1% increase in public R&D raises AI patent output by 0.27%.

Introduction

Artificial intelligence (AI) is now recognized as a transformative technology with the po- tential to boost productivity, reshape labour markets, intensify competition, and accel- erate innovation itself (Autor, 2015; Acemoglu and Restrepo, 2018; Aghion et al., 2019; Cockburn et al., 2018; Mann and Pu¨ttmann, 2023). While AI has existed in various forms for decades, from early rule-based systems to statistical learning models, the re- cent emergence of generative AI has been celebrated by policymakers and investors alike (UNCTAD, 2025). This latest wave has triggered record levels of public R&D investment and sparked a global race to lead in AI innovation.[1]

In the US, the federal government allocates over $4 billion annually to AI through a coordinated R&D strategy involving agencies such as the National Science Foundation (NSF) and the Defense Advanced Research Projects Agency (DARPA), which support innovation across a network of research institutions and regional tech hubs. The UK has committed over £1 billion through its *AI Sector Deal* and *National AI Strategy*,[2] while similar initiatives in Switzerland are supported by the *Digital Switzerland Strategy*. The European Union (EU) has launched a suite of ambitious R&D programmes, including *Horizon Europe* (2021–2027) and the *Digital Europe Programme* to fund the research, development, and deployment of AI.[3]

Given the unprecedented scale of public R&D in AI, in this paper we ask: Does public R&D spending help increasing AI innovations?

Our proposed measure of AI innovations relies on patent grant texts. Recent research

---

[1] Governments traditionally provide subsidies for research and development (R&D) in emerging tech- nologies. One reason for this is that the private sector may fail to internalize the social benefits of innovation. Another is that financial constraints can lead small firms to underinvest in the early stages of R&D (Howell, 2017; Haltiwanger et al., 2013).

2The UK government has committed £7 million from the UK Research and Innovation (UKRI) Technology Missions Fund through the Innovate UK BridgeAI programme to support R&D in key AI areas, as part of its broader strategy to position the UK as a global leader in AI development.

[3] This builds on a policy commitment that began with *Horizon 2020*. Between 2014 and 2020, EUR 8 billion of EU funding, channeled primarily through *Horizon 2020* and Cohesion Policy instru- ments, was allocated to AI investments across European regions, representing approximately 7% of total annual AI investment in recent years (Santos et al., 2025). Under *Horizon Europe* and the *Digital Europe Programme*, the European Commission now invests over EUR 1 billion per year in AI, supporting the development and scaling of innovative technologies. The broader objective is to mobilize additional investment from the private sector and Member States to reach a total annual AI investment of EUR 20 billion by the end of this decade.

highlights the importance of AI-related technologies compared to other forms of innovation. In the US, for instance, AI patents receive significantly more forward citations and command a 9% economic value premium relative to non-AI patents within the same sector (Chen et al., 2024). Moreover, regions in Europe with higher concentrations of AI patent families tend to experience stronger employment growth, with estimated gains of around 2% (Damioli et al., 2024).

We take a practical approach to address the question posed above. If a policymaker had a fixed budget of public R&D funding to allocate, to what extent could they expect it to stimulate AI innovation, as measured by patenting activity, across regions in Europe and the UK? Addressing this question presents several empirical challenges, particularly regarding the classification of AI patents, as well as known endogeneity issues related to public R&D funding and the innovation activities occurring within the regions receiving such funding.[4] We address these issues as follows:

We construct a novel dataset of AI innovations, classified using machine learning techniques applied to patent text data. Specifically, we draw on all patents filed with the European Patent Office (EPO) between 2003 and 2021. The dataset includes biblio- graphic information such as patent titles, abstracts, filing dates, citations, and inventor locations. Using a Naive Bayes Classifier (NBC), we classify patents as either AI-related or non-AI-related, thereby mapping AI innovation production across Europe and the UK over time. Furthermore, using information on inventors' locations, we allocate patents to NUTS-2 regions and construct AI patent series at the regional level.[5]

Our AI patent series reveals a substantial rise in AI-related innovation across European regions over the past two decades. Between 2003 and 2021, the number of AI patents increased by approximately 83%, while the share of AI patents relative to total patents grew by nearly 29%. We further identify the leading regions in AI patent production over the same period. The distribution of AI innovation is highly uneven: the top 10% of patent-producing regions account for approximately 75% of all AI patents in the EU and the UK. ˆIle-de-France, Oberbayern, and Stockholm stand out as leading AI hubs, with these three regions alone accounting for 31% of all AI-related patents.

---

[4] See, for example, Moretti et al. (2023) and Pallante et al. (2023) for detailed discussions on the endogenous nature of R&D spending, albeit with different research questions than ours.

[5] NUTS-2 refers to the second level of the EU's Nomenclature of Territorial Units for Statistics, commonly used in regional economic analysis. Despite the UK's departure from the EU, the dataset includes UK regions, as they remained part of the EPO and the NUTS classification system throughout the sample period.

However, identifying the causal effect of public R&D spending on AI innovation re- quires that R&D allocations be exogenous to regional innovation capacity. To address this endogeneity, we instrument public R&D with plausibly exogenous variation in defense- related R&D spending. Defense R&D is a significant component of public R&D policy and represents a key channel through which governments influence the direction of inno- vation (Draca, 2013; Moretti et al., 2023).

Endogeneity may arise if R&D funding is systematically allocated to regions with ex- isting technological capacity or, conversely, to regions lagging behind innovation activity as part of policy interventions. Such targeting can introduce spurious correlations between public R&D spending and unobserved determinants of AI innovation, potentially overstating the relationship between AI and R&D.

We construct a shift-share instrument based on national variation in defense-related R&D, following established identification strategies in the public spending literature (Autor et al., 2013; Miyamoto et al., 2019; Nakamura and Steinsson, 2018; Borusyak et al., 2022). The choice of our instrument is justified as follows. First, variation in defense R&D over time is primarily driven by exogenous events, such as geopolitical tensions, terrorism, and military buildups, that plausibly exogenous to shifts in regional AI inno- vation (Ramey, 2011; Nakamura and Steinsson, 2014; Moretti et al., 2023; Pallante et al., 2023). Second, the sectoral employment shares used in the shift-share component are predetermined, fixed to the year preceding the start of our sample period. This ensures that variations in defense R&D spending across regions remain independent of contem- poraneous regional economic shocks. Finally, defense R&D in our instrument is observed at the national level, while AI innovation is measured regionally. This reduces the risk of reverse causality, as national defense allocations are unlikely to respond to local economic or technological conditions. Our identification strategy aligns with best practices in the recent literature for shift-share instruments (Borusyak et al., 2025), by ensuring that both components of the instrument, the shift and the share, are exogenous.

Our instrumental variable (IV) regressions show that public regional R&D spending has a significant positive effect on AI-related patent production, after controlling for innovation drivers, sectoral structure, and business cycle conditions. Specifically, a 10% increase in public R&D spending is associated with a 2.7% rise in AI patent output per capita.

Our paper contributes to the following strands of literature. First, it provides a novel addition to the embryonic body of research on the measurement of AI innovation within Europe. Specifically, we offer the first attempt to map regional AI patenting activity across the EU and the UK using machine learning techniques. This represents a step towards aligning the European evidence with recent methodological advances in the U.S. literature, which increasingly relies on textual analysis to classify AI-related innovations (see Giczy et al. (2022)). Prior studies focused on Europe have predominantly used International Patent Classification (IPC) codes or keyword-based search strategies to classify AI patents (Buarque et al., 2020; Xiao and Boschma, 2023). However, these methods face significant limitations in keeping pace with the rapidly evolving and heterogeneous terminology associated with AI technologies (Toole et al., 2020; Baruffaldi et al., 2020; Miric et al., 2023). In contrast, our machine learning approach enables the use of a broader and more flexible definition of AI, capturing a wide array of technologies, in- cluding neural networks, machine learning, and speech recognition, that have emerged over several decades. This approach enables the classification of AI innovations that may be overlooked by conventional query-based methods. Furthermore, it facilitates the con- struction of a regional AI innovation index that can be readily integrated into our panel econometric framework.

Second, we contribute to the broader literature on regional R&D and innovation. Several studies have examined how public investment influences patenting activity in Europe, highlighting the role of public R&D in supporting regional innovation (Bloom et al., 2019; Bottazzi and Peri, 2003; Miguelez and Moreno, 2018). For example, Bot- tazzi and Peri (2003) show that innovation is highly path-dependent, with public R&D

reinforcing existing regional strengths. Miguelez and Moreno (2018) further demonstrate that the effectiveness of public R&D depends on the technological relatedness of regional capabilities. However, this literature has largely overlooked the domain of AI, despite its substantial policy support, as discussed earlier.

We, therefore, add to the emerging literature on public policy and AI innovation. Only a limited number of studies have addressed this area. Igna and Venturini (2023), for instance, emphasize that firm-level factors, such as digital capabilities and ICT in- frastructure, as key drivers of AI adoption and innovation. In contrast, our study shifts attention to the regional level, where R&D policy constitutes the main public policy tool. Santos et al. (2025) also study the allocation of AI-related funding across regions. However, while their emphasis is on identifying the regional characteristics that make regions more likely to receive R&D funding, we focus on assessing the causal impact of exogenous public R&D funding on AI-related patenting activity.

In this context, our third contribution lies in employing defense-related R&D spending as an IV to establish a causal relationship between public R&D investments and AI patenting. While the use of such instruments is well-established (Ramey, 2011; Moretti et al., 2023), their application to the domain of AI innovation is novel. This plausibly exogenous instrument helps address endogeneity in the allocation of public funds and identify the true effect of public R&D on AI innovation.

The paper proceeds as follows: Section 2 describes the data sources and methodology used to classify and regionalize AI innovations. Section 3 outlines the empirical model and identification strategy to establish causality between public R&D and AI innovation. Section 4 presents and discusses our results. Finally, Section 5 concludes the paper.

Data

This section describes our data, the strategy used to classify AI innovations, and our approach to regionalize patents and the mapping of these patents across European regions.

Using textual analysis to measure AI innovations

Patents as a Measurement of Innovation

Patent texts remain one of the most widely used proxies for innovation in empirical re- search (Acs et al., 1994; Kelly et al., 2021). Compared to alternatives, patents offer an objective and structured means of tracking technological developments, as their granting process involves examination criteria that ensure a baseline level of novelty in each invention (OECD, 2004).[6] Furthermore, patent databases provide extensive, long-term coverage across multiple industries and geographical regions, making them particularly valuable for studying the evolution of technological progress over time (Kim and Lee, 2015; Kelly et al., 2021). Additionally, patents capture a significant portion of technological advancements, particularly in sectors with high research intensity, such as AI and computer software (Buesa et al., 2010).

The PATSTAT dataset

To track AI innovations in Europe, we use patent data provided by the European Patent Office's (EPO) PATSTAT. The initial dataset covers all patents registered in EPO, covering patents for 39 European countries, making it one of the comprehensive patent registries in the world. Additionally, PATSTAT contains several features that are in- dispensable at different stages of our analysis. First, it includes information on titles, abstracts, and IPC codes, which we rely on for training our textual analysis algorithm.[7] Second, another advantage is that we have information on the residency of each inventor contributing each patent. This feature allows us to match inventor addresses with their corresponding NUTS-2 regions as explained later. Third, it includes information linking.[6] The main alternative, directly measuring commercialized innovations, faces significant data availability constraints, as records of innovation-derived revenue are scarce and inconsistent across sectors and regions. Other potential innovation indicators, such as survey-based measures, are subject to biases related to firm-level reporting practices, response rates, and variations in the interpretation of what constitutes an innovation (Kleinknecht et al., 2002).

---

[7] IPC is a hierarchically structured code system that classifies patents based on technical content. This classification system is established by Strasbourg Agreement and is maintained by the World Intellectual Property Organization.

each patent to its NACE industrial classification,[8] a feature that allows us to disaggregate AI innovations by sector of origin, e.g., manufacture or Information and Communication Technology (ICT) sectors.

Our final sample comprises 2,862,080 inventors allocated to 2,046,294 patents published over the period 1980-2022. However, for the empirical analysis, as we discuss below, we further restrict the sample to patents filed between 2003 and 2022 to align with the availability of regional R&D spending. The detailed steps leading to the final sample are described in Appendix A.

Defining and Classifying AI patents

Our objective is to classify patents related to AI technologies. However, due to the large size of our dataset, using human coders for this task would be prohibitively expensive. On the other hand, relying on keyword queries and patent code classifications to classify AI technologies, as is often done in the literature,[9] carries its own caveats. The liter- ature highlights key limitations of query-based patent classification, including reliance on extensive synonym lists and poor adaptability to evolving technology terms, such as the shift from 'machine learning' to 'deep learning.' Furthermore, patent classification systems are periodically updated, leading to inconsistencies in measurement over time (Toole et al., 2020; Baruffaldi et al., 2020; Miric et al., 2023).

Against this background, we apply a supervised machine learning method to classify patent texts after constructing a training dataset.[10] For our classification strategy, we would ideally rely on a single definition for AI that remains largely consistent throughout our classification period. This is not the case, however, as there is no consensus on an AI definition (Winfield, 2020).[11] The evolution of artificial intelligence can be roughly described in three waves (UNCTAD, 2025). The first wave (1950s–1980s) focused on rule-based systems, where AI operated through explicitly programmed logic and expert systems. The second wave (1990s–2010s) introduced statistical learning, big data, machine learning, and deep learning to enable more flexible, data-driven models. The current third wave (late 2010s–present) is marked by contextual adaptation, where generative AI systems such as ChatGPT demonstrate the ability to understand, generate, and interact with human-like adaptability across various contexts.

---

[8] NACE stands for *Nomenclature statistique des Activités ´économiques dans la Communauté Européenne.*

[9] See, for example, Fujii and Managi (2018) and Buarque et al. (2020), among others.

[10] There is a growing number of studies using textual analysis to classify AI innovations, most of them focusing on the US. To name a few, Abood and Feltenberger (2018); Giczy et al. (2022); Miric et al. (2023); Kumar and Burns (2025). For a detailed literature review on the different approaches, see Aristodemou and Tietze (2018).

[11] The discrepancies in the literature on AI classification are well documented (Baruffaldi et al., 2020). Some classifications, such as those proposed by Fujii and Managi (2018) and European Patent Office (2017), adopt a more conservative approach, primarily focusing on computational models. By contrast, and closer to the approach of the present study, the OECD (2022) definition takes a broader perspective, encompassing AI applications such as image processing and digital devices with an AI aspect. Similarly, Cockburn et al. (2018) emphasize a more application-oriented view, particularly in relation to robotics.

[12] The assumption of word independence, means that the presence of a specific word in a patent document is assumed to be unrelated to the presence of any other word. In practice, this simplifies the computation of probabilities in the NBC.

Given the evolving terminology and conceptual scope of AI over the period under study, we adopt a broader, component-based approach to our manual classification, fol- lowing Giczy et al. (2022). Instead of relying on a narrow definition of AI, we focus on eight core technological components that have consistently characterized AI innovations throughout their different stages of evolution and are identifiable in patent texts. These components include machine learning, AI hardware, natural language processing, speech recognition, computer vision, knowledge processing, planning/control and evolutionary computation, each described in detail in Appendix B. In brief, our definition of AI-related patents is a technology whose main contribution, as defined in the title and abstract, corresponds to at least one of the above components. Appendix C provides an example of an AI and a non-AI patent which we manually classify in the construction of the training dataset of our model.

To identify AI patents, we employ the NBC, a supervised learning algorithm known for its ease of interpretation and computational scalability, especially with large datasets. This algorithm operates on the assumption that the probability of a word appearing in a document is independent of the presence of other words.[12] Despite its computational simplicity, its performance has been demonstrated to be particularly high (Domingos and Pazzani, 1997). An advantage of the algorithm that makes it suitable for this kind of analysis is that due to the small number of estimated parameters, it is not likely to overfit (Murphy, 2012). Recently, the NBC has been employed in contexts similar to ours, relating to patent classification about emerging and automation technologies (Mann and Pu¨ttmann, 2023; Aristodemou and Tietze, 2018).

We ensure an equal allocation between AI and non-AI classified patents, resulting in a training sample of 2,064 patents, with 1,000 patents classified as AI and 1,064 as non- AI patents.[13] Among the AI-classified patents, approximately 50% pertain to computer vision, 35% to speech recognition, and 16% to natural language processing.[14]

Figure 1 visualizes the 105 tokens selected by the mutual information criterion. Note that the size of each token is proportional to its mutual information score. The three most important tokens appear to be “method”, “data”, and “image”, general terms that could appear in all eight components of AI technology. We observe two broad groups of tokens. The first group is associated with methodological aspects of AI, such as “algorithm”, “learn”, “network”, and “train”. The second group consists of tokens related to the input/output of the respective patent, including “input”, “output”, “audio”, “voice”, “speech”, “object”, and “pixel”.

Note also that we assess the in-sample performance of the NBC by classifying a subsample of the training sample. Specifically, we train the algorithm using 60% (1,239 patents) of manually classified patents and use the remaining 40% (825 patents) as a test sample. Overall, the results indicate that the algorithm achieves a high degree of accuracy in distinguishing between AI-related and non-AI patents.

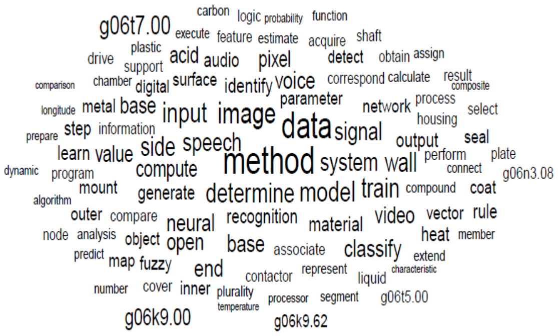
The final step is to employ the NBC to classify the remaining out-of-sample patents. Although we classify all patents since 1980, hereafter we focus only on the patents pub- lished during the period 2003-2021. This period is selected as it aligns with the sample period used in our empirical analysis, which is determined by data availability of re- gional R&D spending. We classify a total of 1,371,864 European patents for the period 2003–2021, of which 208,634 are classified as AI patents (see the detailed steps leading to the final sample in Appendix A). Figure 2 illustrates the aggregate volume of classified

---

[13] Balancing classes in text classification tasks helps prevent bias toward the majority class and im- proves model generalization; see, for example, Chawla (2005).

[14] See also Appendix B for a detailed description of the AI classes.

Figure 1: Tokens selected by the mutual information criterion.

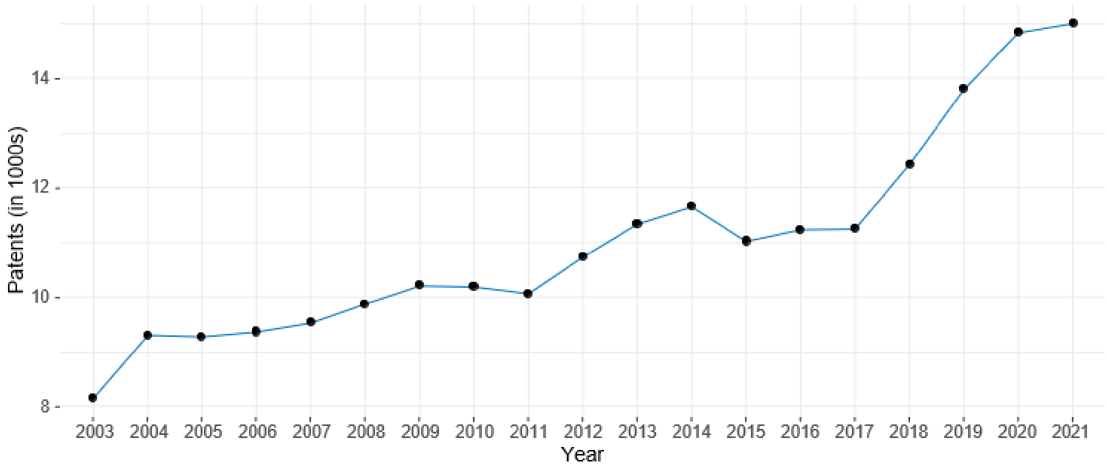


Note: Token size is proportional to the mutual information score.

Source: PATSTAT and own calculations.

AI patents over time. We observe an increasing annual trend that becomes more pro- nounced after 2017. In particular, the number of patents related to AI increased from 8,154 in 2003 to 14,930 in 2021 (an 83% increase). An increase can also be observed in the share of AI patents out of all patents, as shown in E1, Appendix E. Specifically, in 2003 the share of AI patents over the total number of patents is 14% while in 2021 the corresponding share is 18%.

Figure 2: Number of AI patents over time



Note: See text for classification of AI patents and assignment of patents to categories.
Source: PATSTAT and own calculations.

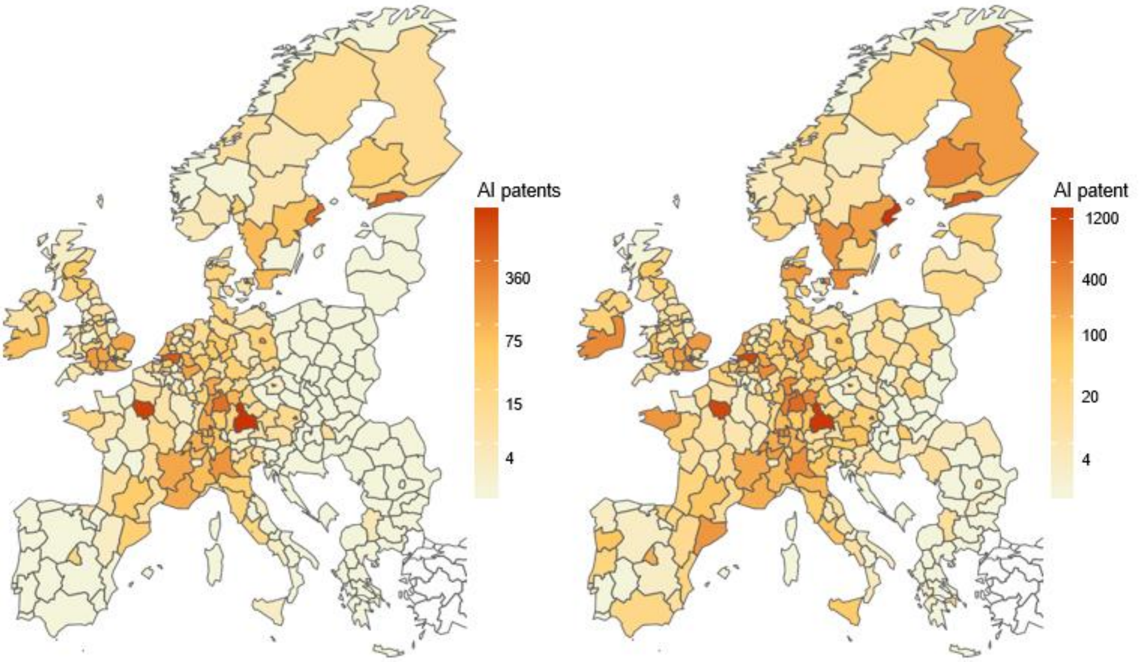
Our approach to classifying AI patents is one of the few approaches tracking AI innovation in Europe (see Buarque et al. (2020); Xiao and Boschma (2023)). Unlike these studies that rely on keyword-based queries to classify AI patents, we use a trained machine learning classifier to analyze patent text. Consequently, our approach identifies a significantly larger number of AI patents. This is expected, as our method allows us to detect innovations not only based on recent definitions of AI, but also those that reflect the evolving terminology and conceptual scope of AI, as explained earlier in Section 2.1. As a result, our method captures a broader range of AI-related advancements. A comparable classification to ours can be found in Giczy et al. (2022), who also used a supervised ML classifier to track US AI patents. Their results contrast with the smaller number of patents classified through traditional keyword searches, with the supervised approach finding almost nine times as many patents (Ho¨tte et al., 2023).

Regionalization of patents

We next disaggregate our AI-related patents to the regional level (NUTS-2). We leverage a key feature of the dataset: approximately 98% of AI patents registered at EPO are geocoded at the NUTS-2 level across EU countries and the UK. Each patent is assigned to the NUTS-2 region of the inventor's residence. For patents with multiple inventors, we attribute the patent to the region with the majority of inventors; if no majority exists, it is assigned to the region of the first-listed inventor.

Figure 3 maps AI patenting activity during this period, illustrating both the overall growth of AI innovation and its evolving spatial distribution. Notably, countries such as Spain and Finland have substantially increased their AI patenting, narrowing the gap with traditional leaders such as Germany and France. Table 1 presents the twenty regions with the highest number of AI-related patents. At the country level, Germany and Switzerland stand out, with six and three regions represented, respectively. The remaining countries are represented by one or two regions each. The leading producer of AI patents over the full period is ˆIle-de-France; however, by 2021, it ranks fourth, behind Stockholm, Oberbayern, and North Brabant.

Figure 3: Count of AI patents (in log) in 2003 and 2021



Note: See text for classification of AI patents and assignment of patents to NUTS-2 regions.
Source: PATSTAT and own calculations.

Table 1: Top AI producing NUTS-2 regions

NUTS-2	Region	AI Patents	Non-AI Patents	AI Patents 2021
FR10	ˆIle de France	24,210	147,615	1,018
DE21	Oberbayern	24,137	175,527	1,395
SE11	Stockholm	16,371	30,042	1,550
NL41	North Brabant	15,600	46,055	1,095
FI1B	Helsinki-Uusimaa	8,475	20,899	573
DE11	Stuttgart	7,037	80,308	460
DE12	Karlsruhe	41,80	32,773	287
UKI3	Inner London - West	3,755	23,597	144
CH04	Zurich	3,207	24,682	201
DE71	Darmstadt	3,068	57,587	222
ITC4	Lombardia	2,803	42,260	192
SE22	South Sweden	2,750	11,002	192
DEA2	Köln	2,738	45,486	208

IE02	Southern and Eastern	2,406	9,297	211
UKJ1	Berkshire	2,336	16,447	146
CH01	Lake Geneva	2,234	23,225	140
CH03	Northwestern Switzerland	2,160	40,610	228
DE30	Berlin	2,153	16,143	173
FR71	Rhône-Alpes	2,136	32,043	95
FI19	West Finland	2,119	5,456	226

Notes: See text for classification of AI patents and assignment of patents to NUTS-2 regions.

Source: PATSTAT and own calculations.

Complementary data used for regressions

We complement our dataset with regional R&D spending data from the Eurostat Regional R&D Expenditure Database (rd_e_gerdreg), our key independent variable.[15] The data concern public R&D expenditure at the NUTS-2 level from 2003 to 2021, reported in millions of national currency and converted to euros where applicable.[16] From the same database, we extract business R&D spending, which measures private sector contributions to R&D, and university R&D spending, which tracks investments in R&D by

---

[15] The data is available at: https://ec.europa.eu/eurostat/databrowser/view/rd_e_gerdreg__custom_14787353/default/table?lang=en.

[16] Although some countries report R&D expenditure outside this period, we focus on this time frame to maximize the number of regions included in the analysis. There are still some missing years, which we linearly interpolate, ensuring that we avoid any forward or backward extrapolation. Note that interpolated data account for between 0% and 34.05% depending on the region. This imputation strategy is commonly used in studies involving regions, which often contain many missing values. For instance, a similar approach to handling missing data of comparable magnitude is used by Moretti et al. (2023).

higher education institutions and public hospitals. Next, we complement our dataset with national defense R&D spending, which we use to construct our IV. This data is sourced from the OECD and measured through government budget allocations for R&D (GBARD). We also incorporate human resources in science and technology (HRST), de- fined as the percentage of the population with tertiary education or employment in science and technology occupations. Other regional-level variables are sourced from the Annual Regional Database of the *European Commission's Directorate General for Regional and Urban Policy* (ARDECO), which include regional GDP, regional employment, and regional population. Regarding national-level variables which are used in our empirical analysis, we source national GDP, GDP deflators, and trade openness from the World Bank, as well as R&D tax credits from the OECD, which capture policy incentives for private R&D investments.

All the variables and their sources are described in detail in Table D1 in the Appendix.

Empirical Strategy

This section outlines our empirical strategy to capture the impact of public R&D spending in generating AI-related patents across European and UK regions. We outline our baseline econometric specification, followed by our identification approach and the IV strategy used to address potential endogeneity concerns.

Main specification

Our specification is the following and accommodates regional and yearly variation. The level of observation is a regional-country-year (*r, c, t*):

$$AI_{r,c,t} = \beta RD_{r,c,t-1} + \gamma X_{c,t-1} + \delta Z_{r,t-1} + d_{r,c} + d_{r,t} + u_{r,c,t} \qquad (1)$$

where $AI_{r,c,t}$ represents the natural logarithm of classified AI-related patents per capita in region *r* of country *c*.[17] $RD_{r,c,t-1}$ denotes lagged public R&D spending per capita in region

---

[17] Following the literature on the determinants of innovation, we adjust for regional population to account for potential scale effects in invention (Furman et al., 2002; Buesa et al., 2010).

r, also expressed in logarithmic terms. $X_{c,t-1}$ and $Z_{r,t-1}$ are vectors of country-by-year and region-by-year variables, respectively, that capture various determinants of AI innovation as well as potential confounding factors. For example, these vectors include controls for regional and national business cycle fluctuations, measured through regional GDP per capita, employment and national fiscal policies. Our specification also incorporates region-by-country fixed effects ($d_{r,c}$) to account for institutional heterogeneities across regions and differences in regional trajectories of R&D investments and AI patenting activity. Additionally, we include time dummies ($d_{r,t}$) and an idiosyncratic error term (ur,c,t).

The focus of our analysis is on estimating the coefficient β, which captures the relationship between annual changes in R&D spending and AI patenting activity in the subsequent year.

Endogeneity concerns and Instrumental Variables

Public R&D policies are unlikely to be random and may be set endogenously as a function of shocks to regional patenting activity. Regional R&D funding may respond to AI innovations emerging in a given region; for example, governments or local authorities may support highly innovative regions to exploit their comparative advantage. Conversely, funding may also be directed toward regions with fewer patents per capita to stimulate innovation activity.

To address these concerns, we adopt a well-established macro identification strategy and rely on R&D defense spending instruments (Autor et al., 2013; Miyamoto et al., 2019; Nakamura and Steinsson, 2018; Moretti et al., 2023). Defense R&D constitutes a substantial component of government R&D in most countries and contributes to significant variations in public R&D over time. In addition, as Daffix and Jacquin (2009) show, there is a strong association between regional public R&D and national military R&D spending, driven by spillover effects, national funding structures, and public-private partnerships, which supports the relevance of our instrument. [18]

---

[18] Defense R&D has historically driven key innovations with widespread civilian applications, such as commercial aerospace, semiconductors, and software (Daffix and Jacquin, 2009). In some countries, particularly France and the UK, publicly funded R&D performed by private firms often originates from defense contracts, linking military and civilian research efforts.

Importantly, defense R&D is largely driven by exogenous shocks, such as geopoliti- cal tensions, terrorism, and war buildups (Mowery, 2010; Moretti, 2021), rather than by countercyclical policies aimed at stabilizing economic conditions at the regional level. To support the validity of defense spending as an instrument, we document fluctuations in military expenditure driven by geopolitical events, suggesting exogenous variation unre- lated to domestic economic conditions. We plot the Geopolitical Risk Index (GPR) from Caldara and Iacoviello (2022) to capture such events. Figure E2 in Appendix E plots military spending as a share of GDP (orange line) alongside the GPR index (blue line). In countries such as Germany, Spain, Belgium, France, Norway, and Poland, military spending increased following periods of elevated geopolitical risk. Taken together, we ar- gue that defense R&D is plausibly exogenous to shifts in AI innovation activity, making it a strong candidate for instrumenting public R&D spending.

In line with Nekarda and Ramey (2011); Dupor and Guerrero (2017); Moretti et al. (2023), we construct a shift-share instrument. Specifically, our shift-share *IV*, $Def_{r,c,t}$, is defined as:

$$Def_{r,c,t} = \frac{\text{sectoral employment}_{r,2002}}{\text{total employment}_{r,2002}} \times DefRD_{c,t-1} \qquad (2)$$

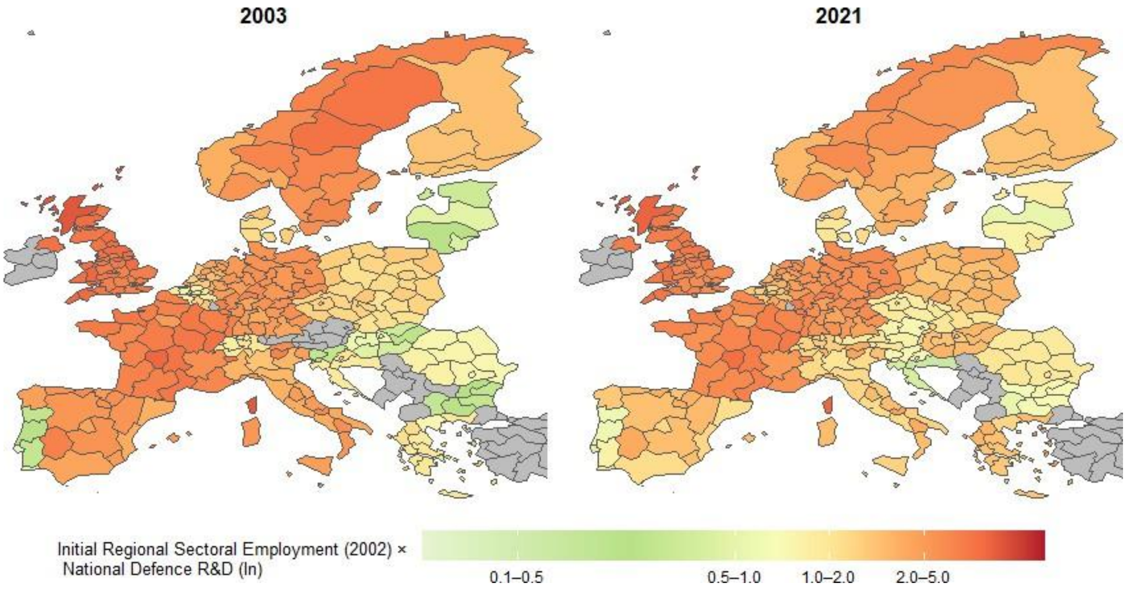
where $Def_{r,c,t}$ represents total military R&D spending in country *c* in year *t* in region *r*. We construct this instrument by weighting it with the share of regional employment in the public sector, which aims to capture cross-regional variations in military sector employment. Here, sectoral employment*r,*2002 represents the number of pre-existing public-sector employees in region *r* in the base year 2002, and the denominator aggregates total em- ployment in the region. The variable $Def\,RD_{c,t-1}$ denotes lagged national defense R&D spending per capita, which captures exogenous variations in government R&D.[19]

---

[19] Pre-existing employment shares and employment growth shares have been widely used in the shift- share instrument literature. For example, Bartik (1991) originally proposed this approach to examine labor market policies. Blanchard and Katz (1992) applied a similar methodology to study regional employment dynamics. Autor et al. (2013) used initial employment shares to assess the local labor market impact of Chinese import competition, while Nunn and Qian (2014) employed a shift-share approach to analyze the effects of US food aid on civil conflict. More recently, Goldsmith-Pinkham et al. (2020) provide an extensive discussion on the identification assumptions and applications of shift-share instruments.

Intuitively, the instrument exploits exogenous national-level fluctuations in defense R&D spending, weighting them by each region's pre-existing public employment share from 2002. This approach relies on the assumption that national defense R&D is largely driven by geopolitical events and security policies rather than local economic conditions. By using a predetermined 2002 employment-based exposure measure, we ensure that variations in defense R&D spending across regions remain independent of contemporaneous regional economic fluctuations, mitigating potential biases from endogenous regional factors (Goldsmith-Pinkham et al., 2020). Overall, our identification strategy aligns with best practices for Bartik-type shift-share instruments, as described by Borusyak et al. (2025), by ensuring that both components of the instrument are exogenous, the shift (national defense R&D spending) and the share (2002 regional public employment).

Figure 4: Defence spending by Initial Sectoral Shares (2002) in NUTS-2 Regions



Note: The instrument is constructed as the interaction between initial (2002) regional sectoral employment shares in public services and the log of national defence R&D expenditure. The two panels reflect variation over time.. Grey areas indicate regions with missing defence data.

Results

Table 2 shows the estimated results using both OLS and IV, employing the instrument described in Section 3.1.1. We include the following controls: regional employment to capture region-specific labor market effects, regional GDP to account for cyclical effects, and national government spending to capture broader fiscal policies. Overall, our baseline controls represent demand shifters which are traditionally thought to influence innovation and patenting activity (Aghion and Howitt, 1990; Acemoglu et al., 2018; Bloom et al., 2019).

The first-stage results of our IV estimates indicate that the instruments are generally strong. Weak instrument diagnostics, reported at the bottom of Table 2, confirm that the instruments have sufficient power. The first-stage F-statistics range from 31.6 to 66.9, consistently exceeding the conventional threshold of 10, indicating strong instrument relevance. Furthermore, the Anderson-Rubin Wald F-test rejects the null hypothesis of weak instruments across all specifications. All regressions exhibit high R^2, which should be attributable to the inclusion of regional and time fixed effects. Finally, the high p-values from the Sargan tests suggest that our IV estimates do not reject the overidentifying restrictions, supporting the validity of our instrument.

The first-stage coefficients are reported in the lower part of Table 2. An increase in regional military R&D would raise regional publicly-funded R&D, reflecting spillover effects, national funding structures, and public-private partnerships, as discussed earlier. Our first-stage coefficients are consistently positive and statistically significant across all specifications. Under the baseline (column 2), a 1% increase in military R&D is associated with a 0.43% increase in public R&D spending. This also confirms the relevance of the instrument and the assumption that military R&D is a meaningful predictor of broader public R&D funding in line with the literature (Daffix and Jacquin, 2009; Miyamoto et al., 2019; Moretti, 2021; Pallante et al., 2023).

Across all IV specifications, the estimated effect of public R&D spending on AI patent- ing is positive and statistically significant. Under our baseline IV specification (column 2), a 10% increase in public R&D spending is associated with a 2.7% increase in AI innovation output per capita. This suggests a substantial impact of public R&D investment in driv- ing AI-related innovation. In contrast, the OLS estimates are consistently smaller, likely reflecting attenuation bias from measurement error in public R&D spending. The IV estimates, provide stronger evidence that public R&D spending significantly contributes to AI innovation.

To address potential reverse causality, where past AI production influences current AI patenting, such as through spillovers from ongoing R&D projects, we control for lagged AI patents in column (3). Incorporating this control reduces the coefficient of public R&D to 0.21, indicating that part of the observed effect is due to persistence in AI innovation. This is further confirmed by the significantly positive coefficient of the lagged variable, suggesting that past innovation contributes to future patenting outcomes. Despite this reduction, the coefficient of public R&D remains positive and statistically significant.

In columns (4) to (6), we introduce controls for regional employment, regional GDP, and national government spending to account for potential confounding factors related to the business cycle and fiscal policies. These additional controls lead to only minor changes in the estimated effect, with coefficients ranging from 0.24 to 0.27. The stability of the R&D coefficients across these specifications indicates that public R&D spending has a robust effect on AI patenting, independent of macroeconomic conditions. Column (7), which includes the lagged AI, shows a drop in the R&D coefficient, again highlighting the role of past innovation.

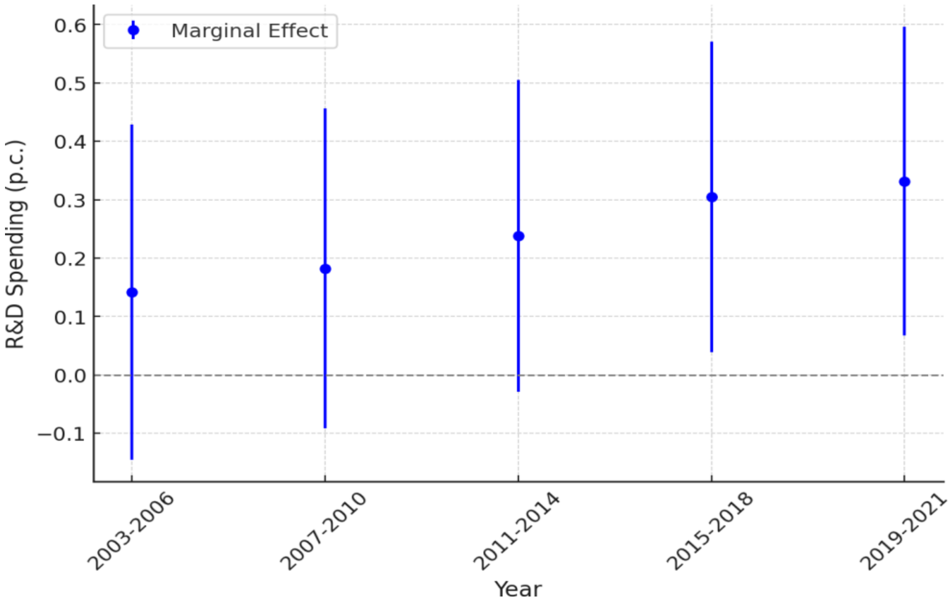
Table 2: Baseline Regression Results

	OLS (1)	IV (2)	IV (3)	IV (4)	IV (5)	IV (6)	IV (7)
R&D Spending$_{t-1}$ (pc, ln)	0.05*	0.27***	0.21***	0.27***	0.24***	0.25***	0.18***
	(0.03)	(0.07)	(0.08)	(0.07)	(0.07)	(0.07)	(0.09)
AI Patents$_{t-1}$ (pc, ln)			0.32***				0.32***
			(0.02)				(0.02)
Regional Employment$_{t-1}$ (ln)				0.00***	0.00***	0.00***	0.00***
				(0.00)	(0.00)	(0.00)	(0.00)
Government Spending$_{t-1}$ (ln)					-0.02**	-0.02**	-0.01
					(0.01)	(0.01)	(0.01)
Regional GDP$_{t-1}$ (ln)						0.00	0.00
						(0.00)	(0.00)
Time Effects	✓	✓	✓	✓	✓	✓	✓
Regional Effects	✓	✓	✓	✓	✓	✓	✓
First-stage coefficient		0.43***	0.27***	0.43***	0.43***	0.45***	0.22***
($R\&D_{def,c,t-1} \times Emp_{share,r,2002}$)		(0.09)	(0.08)	(0.08)	(0.09)	(0.09)	(0.09)
First Stage F		66.91	41.88	66.56	66.56	60.78	37.55
Anderson Rubin-χ^2 (p-val)		0.00	0.00	0.00	0.00	0.00	0.00
Sargan Test (p-val)		0.49	0.65	0.86	0.97	0.93	0.27
R^2	0.92	0.92	0.93	0.92	0.92	0.92	0.92
Observations	3,136	3,136	2,843	3,136	3,136	3,136	3,136

Note: Columns (1)–(7) examine AI patent activity. All models include time and regional fixed effects. ”IV” instruments public R&D using regional defense R&D allocations. Regressions are weighted by the regional share of employment in 2002. The first-stage coefficient refers to the instrumented relationship. Anderson-Rubin tests assess instrument strength; Sargant tests check overidentification. Statistical significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Figure 5 shows how the marginal effect of R&D spending per capita on AI patenting evolves over time. We estimate this by interacting the predicted public R&D variable obtained from the IV with five non-overlapping four-year period dummies. The blue error bars represent 95% confidence intervals. The results reveal a steady increase in this effect over time. Between 2003 and 2014, the impact of R&D spending on AI patenting is borderline statistically insignificant, indicating a moderate influence during that period. However, from 2015 onward, the effect becomes more pronounced and statistically significant, suggesting that public R&D spending has become increasingly effective in driving AI-related innovation.

Figure 5: Marginal effect of R&D spending (p.c.) over time



Note: The figure shows the marginal effect of predicted public R&D spending per capita on AI patenting activity across five-year periods. Estimates are based on interactions of the IV-predicted R&D variable with period dummies. Error bars represent 95% confidence intervals.

Conclusion

This paper examines the impact of public R&D spending on the production of AI-related innovations at the regional level within European and UK regions. We adopt a practical policy-oriented perspective: if a policymaker were to ask us how much additional AI- related innovation could be generated by an additional 1% increase in public R&D, for innovation policy, how would we respond?

Our proposed measure of AI innovation relies on patent grant texts. By leveraging a novel dataset that integrates bibliographic information from patents and applying machine learning techniques to classify AI patents, we construct a regionalized index of AI-related innovation activity across NUTS-2 regions.

Our results are directly relevant to policy debates on effective strategies to improve the cost-effectiveness of R&D funding in boosting AI innovations. We estimate that a 1% increase in public R&D spending per capita leads to an increase in regional AI patenting activity per capita of 0.27%, indicating that public investment plays a direct and measurable role in stimulating technological advancement in AI.

Given the highly uneven distribution of AI innovation across regions—where just 10% of NUTS-2 regions account for 75% of all AI patents, there is a clear need for a more geographically inclusive innovation policy. Policymakers should prioritize expanding and sustaining public R&D funding, especially in regions currently underrepresented in AI innovation. This could help avoid reinforcing existing disparities and instead support the emergence of new AI clusters. A place-based R&D strategy, tailored to regional strengths and institutional capacities, would ensure that funding does not simply flow to areas that are already technologically advanced but also helps lagging regions build foundational capabilities.

The study also highlights the potential of leveraging defense R&D as a source of civilian AI innovation. Since defense-related R&D has historically produced technologies with widespread civilian applications, policymakers should actively design dual-use R&D strategies that facilitate knowledge spillovers from military to commercial and public sectors. Such strategies can be especially effective when coupled with public-private partnerships that align government and industry interests in developing strategic AI technologies.

Last, the results reveal that the impact of R&D spending on AI innovation has grown over time, becoming particularly significant after 2015. This suggests that stable, multi-year commitments to R&D are more effective than short-term funding bursts. Policymakers should, therefore, institutionalize long-term funding strategies that provide certainty for researchers and firms alike, recognizing that innovation outcomes often materialize with a lag.

References

Abood, A. and Feltenberger, D. (2018), ‘Automated patent landscaping’, Artificial Intelligence and Law 26(2), 103–125.

Acemoglu, D., Akcigit, U., Alp, H., Bloom, N. and Kerr, W. (2018), ‘Innovation, reallocation, and growth’, American Economic Review 108(11), 3450–3491.

Acemoglu, D. and Restrepo, P. (2018), ‘Artificial intelligence, automation, and work’, Economics of Innovation and New Technology 28(3), 256–269.

Acs, Z. J., Audretsch, D. B. and Feldman, M. P. (1994), ‘R&d spillovers and recipient firm size’, The Review of Economics and Statistics 76(2), 336–340.

Aghion, P., Antonin, C. and Bunel, S. (2019), ‘Artificial intelligence, growth and employ- ment: The role of policy’, Economie et Statistique 510(1), 149–164.

Aghion, P. and Howitt, P. (1990), ‘A model of growth through creative destruction’.

Aristodemou, L. and Tietze, F. (2018), ‘The state-of-the-art on intellectual property analytics (ipa): A literature review on artificial intelligence, machine learning and deep learning methods for analysing intellectual property (ip) data’, World Patent Information 55, 37–51.

Autor, D. H. (2015), ‘Why are there still so many jobs? the history and future of work- place automation’, Journal of Economic Perspectives 29(3), 3–30.

Autor, D. H., Dorn, D. and Hanson, G. H. (2013), ‘The china syndrome: Local labor market effects of import competition in the united states’, American Economic Review 103(6), 2121–2168.

Bartik, T. J. (1991), Who Benefits from State and Local Economic Development Policies?, W.E. Upjohn Institute for Employment Research.

Baruffaldi, S. H. et al. (2020), ‘Challenges in patent classification and ai queries’, Journal of Intellectual Property.

Blanchard, O. J. and Katz, L. F. (1992), ‘Regional evolutions’, Brookings Papers on Economic Activity 1992(1), 1–61.

Bloom, N., Van Reenen, J. and Williams, H. (2019), ‘A toolkit of policies to promote innovation’, Journal of Economic Perspectives 33(3), 163–184.
Borusyak, K., Hull, P. and Jaravel, X. (2022), ‘Quasi-experimental shift-share research designs’, The Review of economic studies 89(1), 181–213.

Borusyak, K., Hull, P. and Jaravel, X. (2025), ‘A practical guide to shift-share instruments’, American Economic Journal: Applied Economics . Forthcoming.

Bottazzi, L. and Peri, G. (2003), ‘Innovation and spillovers in regions: Evidence from European patent data’, European Economic Review 47(4), 687–710.

Buarque, B. S., Davies, R. B., Hynes, R. M. and Kogler, D. F. (2020), ‘Ok computer: the creation and integration of ai in europe’, Cambridge Journal of Regions, Economy and Society 13(1), 175–192.

Buesa, M., Heijs, J. and Baumert, T. (2010), 'The determinants of regional innovation in Europe: A combined factorial and regression knowledge production function approach', Research policy 39(6), 722–735.

Caldara, D. and Iacoviello, M. (2022), 'Measuring geopolitical risk', American Economic Review 112(4), 1194–1225.

Chawla, N. V. (2005), Data mining for imbalanced datasets: An overview, in O. Maimon and L. Rokach, eds, 'Data Mining and Knowledge Discovery Handbook', Springer,
pp. 853–867.

Chen, W. X., Shi, T. T. and Srinivasan, S. (2024), The value of ai innovations, Working Paper 24-069, Harvard Business School.
URL: https://www.hbs.edu/faculty/Pages/item.aspx?num=65925

Cockburn, I. M., Henderson, R. and Stern, S. (2018), The impact of artificial intelligence on innovation: An exploratory analysis, in 'The Economics of Artificial Intelligence: An Agenda', University of Chicago Press, pp. 115–146.

Daffix, S. and Jacquin, Y. (2009), 'Defense r&d and national r&d systems: a european outlook', The Economics of Peace and Security Journal 4(1).

Damioli, G., Van Roy, V., V´ertesy, D. and Vivarelli, M. (2024), 'Drivers of employment dynamics of ai innovators', Technological Forecasting and Social Change 201, 123249. URL: https://www.sciencedirect.com/science/article/pii/S0040162524000453

Domingos, P. and Pazzani, M. (1997), 'On the optimality of the simple bayesian classifier under zero-one loss', Machine Learning 29(2).

Draca, M. (2013), Reagan's innovation dividend? technological impacts of the 1980s us defense build-up, Technical Report 168, CAGE Working Paper Series.
URL: https://warwick.ac.uk/fac/soc/economics/research/centres/cage/manage/publications/168-2013˙draca.pdf

Dupor, B. and Guerrero, R. (2017), 'Local and aggregate fiscal policy multipliers', Journal of Monetary Economics 92, 16–30.

European Patent Office (2017), Patents and the fourth industrial revolution, Technical report, European Patent Office. Accessed: February 2025.

Fujii, H. and Managi, S. (2018), 'Trends and priority shifts in artificial intelligence technology invention: A global patent analysis', Economic Analysis and Policy 58, 60–69.

Furman, J. L., Porter, M. E. and Stern, S. (2002), 'The determinants of national innova- tive capacity', Research policy 31(6), 899–933.

Giczy, A. V., Pairolero, N. A. and Toole, A. A. (2022), 'Identifying artificial intelli- gence (ai) invention: A novel ai patent dataset', The Journal of Technology Transfer 47(2), 476–505.

Goldsmith-Pinkham, P., Sorkin, I. and Swift, H. (2020), 'Bartik instruments: What, when, why, and how', American Economic Review 110(8), 2586–2624.

Haltiwanger, J., Jarmin, R. S. and Miranda, J. (2013), 'Who creates jobs? small versus large versus young', Review of Economics and Statistics 95(2), 347–361.

Hotte, K., Tarannum, T., Verendel, V. and Bennett, L. (2023), 'Ai technological trajectories in patent data'.
Howell, S. T. (2017), 'Financing innovation: Evidence from r&d grants', American Eco- nomic Review 107(4), 1136–1164.
URL: https://doi.org/10.1257/aer.20150808

Igna, I. and Venturini, F. (2023), 'The determinants of ai innovation across european firms', Research Policy 52(2), 104661.

Kelly, B., Papanikolaou, D., Seru, A. and Taddy, M. (2021), 'Measuring technological innovation over the long run', American Economic Review: Insights 3(3), 303–320.

Kim, J. and Lee, S. (2015), 'Patent databases for innovation studies: A comparative analysis of uspto, epo, jpo and kipo', Technological Forecasting and Social Change 92, 332–345.

Kleinknecht, A., Montfort, K. and Brouwer, E. (2002), 'The non-trivial choice between innovation indicators', Economics of Innovation and New Technology 11, 109–121.

Kumar, A. K. and Burns, M. (2025), Identifying and predicting patent trends in the artificial intelligence domain, in R. J. Howlett and L. C. Jain, eds, 'Advances in Computational Intelligence Systems', Vol. 1462, Springer, pp. 360–372.

Mann, K. and Putman, L. (2023), 'Benign effects of automation: New evidence from patent texts', Review of Economics and Statistics 105(3), 562–579.

Miguelez, E. and Moreno, R. (2018), 'Relatedness, external linkages and regional innova- tion in europe', Regional studies 52(5), 688–701.

Miric, M., Jia, N. and Huang, K. G. (2023), 'Using supervised machine learning for large-scale classification in management research: The case for identifying artificial intelligence patents', Strategic Management Journal 44(2), 491–519.

Miyamoto, W., Nguyen, T. L. and Sheremirov, V. (2019), 'The effects of government spending on real exchange rates: Evidence from military spending panel data', Journal of International Economics 116, 144–157.

Moretti, E. (2021), The New Geography of Jobs, Mariner Books.

Moretti, E., Steinwender, C. and Van Reenen, J. (2023), 'The intellectual spoils of war? defense r&d, productivity, and international spillovers', Review of Economics and Statistics pp. 1–46.

Mowery, D. (2010), Defense and Innovation Policy, University Press.

Murphy, K. P. (2012), Machine Learning: A Probabilistic Perspective, MIT Press, Cambridge, MA.

Nakamura, E. and Steinsson, J. (2014), 'Fiscal stimulus in a monetary union: Evidence from us regions', American Economic Review 104(3), 753–792.

Nakamura, E. and Steinsson, J. (2018), 'Identification in macroeconomics', Journal of Economic Perspectives 32(3), 59–86.

Nekarda, C. J. and Ramey, V. A. (2011), 'Industry evidence on the effects of government spending', American Economic Journal: Macroeconomics 3(1), 36–59.

Nunn, N. and Qian, N. (2014), 'Us food aid and civil conflict', American Economic Review 104(6), 1630–1666.

OECD (2004), 'Patents and innovation: Trends and policy challenges', OECD Report . OECD (2022), 'Oecd framework for the classification of ai systems', OECD Digital Economy Papers, No. 323 .

Pallante, G., Russo, E. and Roventini, A. (2023), 'Does public r&d funding crowd-in pri- vate r&d investment? evidence from military r&d expenditures for us states', Research Policy 52(8), 104807.

Ramey, V. A. (2011), 'Identifying government spending shocks: It's all in the timing',
The Quarterly Journal of Economics 126(1), 1–50.
URL: https://doi.org/10.1093/qje/qjq008

Santos, A. M., Molica, F. and Torrecilla-Salinas, C. (2025), 'Eu-funded investment in artificial intelligence and regional specialization', Regional Science Policy & Practice 17(7), 100190.

Toole, A. A. et al. (2020), 'Limitations in patent classifications for ai innovations', Tech- nology and Innovation Management .

UNCTAD (2025), Chapter I: AI at the technology frontier, in 'Technology and Innovation Report 2025: Inclusive Artificial Intelligence for Development', UNCTAD, pp. 1–30.
URL: https://unctad.org/system/files/official-document/tir2025ch1˙en.pdf

Winfield, A. (2020), 'Intelligence is not one thing', Journal of Artificial General Intelligence 11(2), 97–100.

Xiao, J. and Boschma, R. (2023), 'The emergence of artificial intelligence in European regions: the role of a local ict base', The Annals of Regional Science 71(3), 747–773.

Appendix A - Adjustment of the final sample

Our prediction sample comprises EPO patents filed between 1980 and 2023. Starting from an initial universe of approximately 4.38 million patents linked to 8.45 million inventor records, we implement a series of filtering steps. We first exclude patents with inventors based in Turkey (–4.21%), remove patents lacking abstracts, titles, or IPC codes (–0.64%), and retain only patents with English-language abstracts (–6.71%) to ensure consistency in text processing. Next, we restrict our dataset to patents with inventors located in Europe (–44.27%) and retain only those that can be geocoded at the NUTS-2 level (2,046,294 patents). Finally, we focus on patents filed between 2003 and 2021 to to align with the availability of regional R&D spending available. The resulting sample includes approximately 1.37 million patents. The steps above are summarized in Table A1.

Table A1: Sample Construction from European Patent Office (EPO) Data (1980–2023)

Step	**Inventors**	**Number of Patents**	**Percentage**
Patent Universe (1980–2023): All EP patents	8,453,326	4,379,892	100%
Remove Turkey	8,436,702	4,195,294	–4.21%
Drop data without abstracts, titles, IPC codes	8,393,676	4,168,388	–0.64%
Drop data in non-English abstracts	8,025,195	3,888,611	–6.71%
Remove inventors from outside Europe	3,865,873	2,167,140	–44.27%
NUTS-2 geocoding availability	2,862,080	2,046,294	–5.58%
Restriction to 2003–2021	–	1,371,864	32.96%

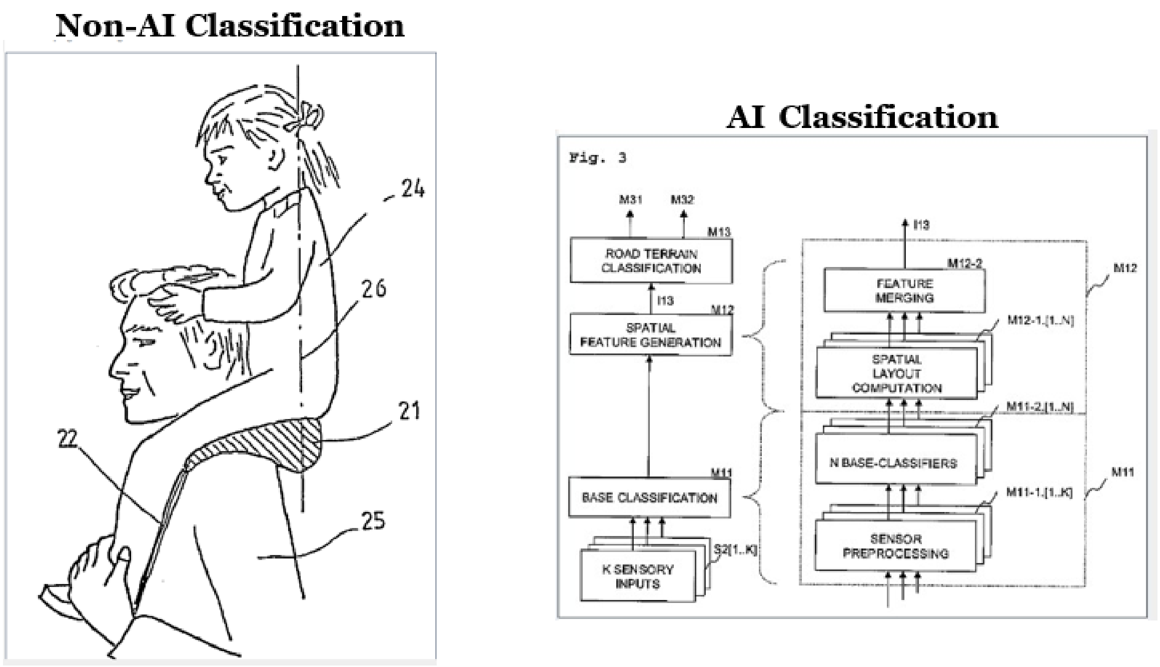
Appendix B - Definitions of AI components

In this research, we do not confine ourselves in a single definition of AI. Rather than that, based on Giczy et al. (2022), we consider eight broad components of AI technologies: ma- chine learning, AI hardware, natural language processing, speech recognition, computer vision, knowledge processing, planning/control, and evolutionary computation. A patent may belong to more than one of these categories. For example, an invention in computer vision could utilize a machine learning method. Next, we provide a definition for each AI technology.

1. Machine learning: The field of machine learning is related to the development and study of statistical algorithms that can learn from data and generalize to unseen data, and thus perform tasks without explicit instructions. Examples of machine learning models are decision trees, support-vectors machines, and neural networks.
2. AI hardware: The field of AI hardware is related to computer hardware designed to execute AI programs. For example, a graphics processing unit may be used to train deep neural networks.
3. Natural language processing is a field of AI concerned with enabling computers to recognize, understand, and generate text and speech. Natural language processing is also related to linguistics. One of the first natural language software is ELIZA, a computer program created to explore communication between humans and machines.
4. Speech recognition concerns the development of methodologies and techniques that enable the translation of spoken (human) language into text. For example, the Hidden Markov Model represents speech as a sequence of hidden states corresponding to phonemes and optimizes the likelihood of word sequences using a dynamic programming algorithm.
5. Computer vision is a field of AI that employs methods enabling computers and systems to process information from visual inputs. For example, blob detection is used to identify two neighboring regions in a digital image with different properties (e.g., different colors).
6. Knowledge processing: This is a field of AI concerned with representing information in a way that enables a system to deduce new knowledge. Specialized hardware and
7. software are usually required for the system to process this information effectively. An inference engine is an example of software used in knowledge processing.
8. Planning/control refers to AI technologies concerned with implementing strategies to achieve a specified goal. Given a description of an initial state, a desired outcome, and a set of possible actions, the planning problem involves determining an order of actions that will guarantee the desired outcome. Combinatorial optimization is typically employed, especially in dynamically changing or unknown environments.
9. Evolutionary computation is a family of algorithms for global optimization that mimic biological evolution. An initial set of potential solutions is defined and iteratively updated. In each iteration, less desirable solutions are discarded, while small changes are introduced to the remaining ones. Evolutionary algorithms include, among others, genetic algorithms and programming, differential evolution, and neuroevolution.

Appendix C Example of Patent Classification as AI and Non-AI

Figure C1: Manual Patent Classification Example



Non-AI Classification

AI Classification

Non-AI Example:
DEVICE FOR CARRYING A CHILD
Abstract: The invention concerns a configuration for carrying a child upon the shoulders of an adult. The device comprises a seat section (21) partly encircling the neck of the adult, and leg sections (22) extending from the end portion's semi-circular seat section...

AI Example:
Road-terrain detection method and system for driver assistance systems
Abstract: The present invention describes a road terrain detection system that comprises a method for classifying selected locations in the environment of a vehicle based on sensory input signals such as pixel values of a camera image...

Note*:* Patent examples were retrieved from EPO's Espacenet online query database and correspond to observations in our classification sample. The classification as AI or Non-AI is based on the patent's core technological contribution as described in its abstract above. The Non-AI patent is a mechanical device with no computational or intelligent processing component, whereas the AI patent explicitly involves classification algorithms, sensory input processing, and system-based decision-making, which align with the fundamental components of AI innovation.

Appendix D - Variable Descriptions

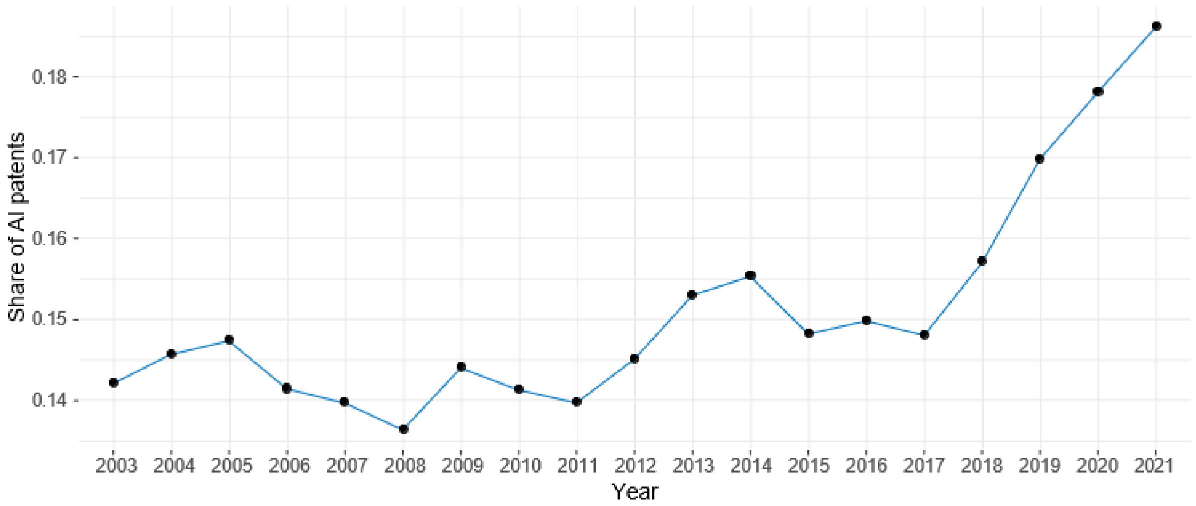
Table D1: Descriptions of Variables Used in the Study

Variable Name	Description	Level	Source
Public R&D Spending	Gross domestic expenditure on R&D (GERD) by the government sector (Million Euros)	NUTS-2	Eurostat
Business R&D Spending	Gross domestic expenditure on R&D (GERD) by the business enterprise sector (Million Euros)	NUTS-2	Eurostat
University R&D Spending	Gross domestic expenditure on R&D (GERD) by the Higher Education Sector (Million Euros)	NUTS-2	Eurostat
Human Resources in Science and Technology (HRST)	Percentage of the population who fulfil one or the other of the following conditions: (i) successfully completed education at the third level; (ii) not formally qualified as above, but employed in a S&T occupation where the above qualifications are normally required.	NUTS-2	Eurostat
Regional population	Average annual population in persons	NUTS-2	ARDECO
Regional employment	Total Employment (workplace based, employed persons)	NUTS-2	ARDECO
Regional GDP	GDP at current market prices deflated by a GDP deflator	NUTS-2	ARDECO
Migration	Net migration in persons	NUTS-2	ARDECO
Defense R&D	Government budget allocations for R&D (GBARD) by socio-economic objective: defense spending	National	OECD
R&D Tax Subsidies	Implied tax subsidy rates on R&D expenditures	National	OECD

Country GDP	GDP (current US$) deflated using the GDP deflator and converted into Euros	National	World Bank
Country population	Total population based on the de facto definition of population, which counts all residents regardless of legal status or citizenship in midyear estimates.	National	World Bank
Trade Openness	Trade (% of GDP)	National	World Bank
Corporate Tax Rate	Corporate tax rate (%)	National	Tax Foundation
Non-R&D Procurement	TED provides details about public procurement notices, including both contract notices and contract award notices in the EU. The non-R&D procurement series (values expressed in euros) excludes con- tracts associated with the procurement of R&D services.	National	Tenders Electronic Daily (TED), European Commission
Composite Risk In- dex	Annual average score that aggregates the political, financial, and economic ratings for each country's overall risk.	National	UNIDO Industrial Statistics Database
Manufacturing Sec- tor Value Added	Total, including Computer, Machinery, Transportation, and Pharmaceuticals.	National	UNIDO Indus- trial Statistics Database
GDP Deflator	Ratio of GDP in current local currency to GDP in constant local currency.	National	World Bank

Appendix E - Additional Tables and Figures

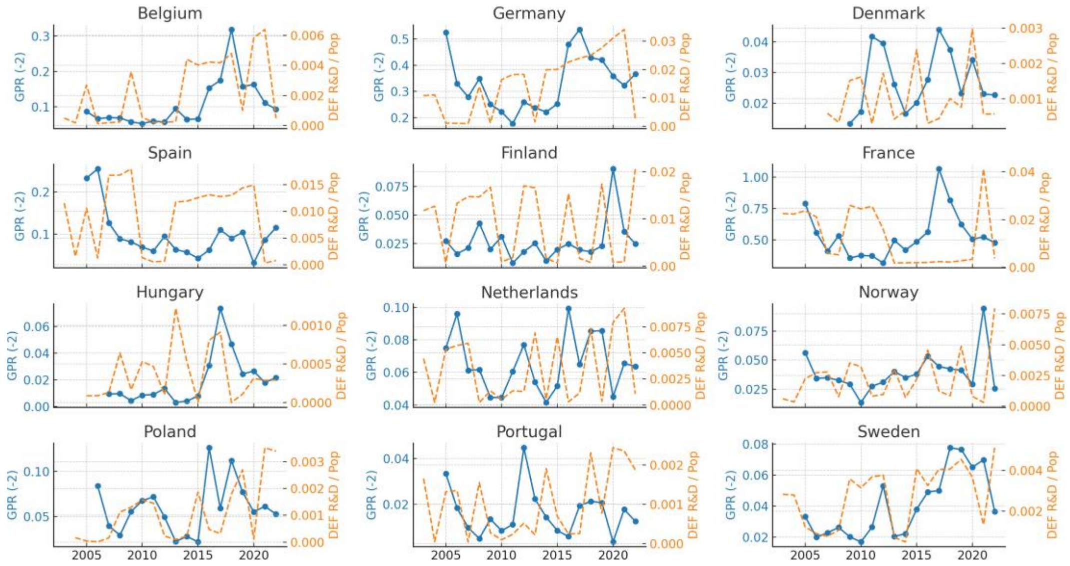
Figure E1: Share of AI patents over time



Note: See text for classification of AI patents.
Source: PATSTAT and own calculations.

]

Figure E2: Military Spending and Geopolitical Events in Selected EU Countries.



Note: Authors' calculations using the country-specific geopolitical risk index by Caldara and Iacoviello (2022) and Eurostat defense data.