Discussion Paper Series

IZA DP No. 18817 July 2026

The Interplay Between AI and Technological Relatedness in Shaping Regional Innovation in Europe

Francesco D'Alessandro
University of Bologna

Enrico Santarelli
University of Bologna
and University of Luxembourg

Marco Vivarelli
Catholic University of the Sacred Heart,
UNU-MERIT and IZA@LISER

IZA@LISER NETWORK
ISSN: 2365-9793

IZA DP No. 18817 July 2026

The Interplay Between AI and Technological Relatedness in Shaping Regional Innovation in Europe*

Abstract

This study examines how regional technological relatedness and local AI knowledge influence regional innovative activity, as measured by patenting activity. Using a novel three-way longitudinal dataset (670 four-digit CPC classes × 302 NUTS-2 regions × nine four-year periods, 1986–2021) and leveraging a deep learning-based identification of AI patents, we show that two broad mechanisms operate in parallel. First, in accordance with the extant literature, technologies that are cognitively close to a region's existing patent portfolio enjoy higher patenting activity, confirming that relatedness remains a strong and persistent predictor of innovative output. Second, local AI endowments are positively associated with patenting across technological fields, even after conditioning on relatedness, indicating that AI plays an enabling and cross-cutting role in a given regional innovation system. Moreover, the interaction between relatedness and AI turns out to be negative and statistically significant, implying that AI attenuates the extent to which local innovative efforts depend on the technology's proximity to the regional portfolio. In sum, AI appears to enhance overall local innovative activity while reducing its reliance on pre-existing regional knowledge structures.

JEL classification

O31, R11

Keywords

Artificial Intelligence, AI, technological change, regional innovation, relatedness

Corresponding author

Marco Vivarelli
marco.vivarelli@unicatt.it

* *Acknowledgments* Marco Vivarelli acknowledges the support by the Italian Ministero dell'Università e della Ricerca – MUR (PRIN-2022, project 2022P499ZB: "Innovation and labor market dynamics"; principal investigator: Marco Vivarelli; funded by the European Union - Next Generation EU). Enrico Santarelli acknowledges the support by the Italian Ministero dell'Università e della Ricerca – MUR (PRIN-2022 project 2022RRBJPL_001: "Automation and digitalization in SMEs: Competition, upskilling and labor policies"; principal investigator: Emanuela Carbonara). The views and opinions expressed are only those of the authors and do not necessarily reflect those of the European Union or the European Commission. Neither the European Union nor the European Commission can be held responsible for them.

1. Introduction

Over the past few decades, Artificial Intelligence (AI) has evolved from a niche research field into a pervasive, transformative force permeating nearly every domain of economic and technological activity. Indeed, it is increasingly regarded as a General-Purpose Technology (GPT) (Bresnahan & Trajtenberg, 1995; Brynjolfsson & McAfee, 2014; Trajtenberg, 2019) and as an Invention in the Methods of Invention (IMI) (Griliches, 1957; Agrawal et al., 2019; Cockburn et al., 2019). Like electricity, steam power, or information and communication technologies, AI is expected to induce long-term structural transformations, not only in how economies grow and regions compete, but more crucially in how innovation is conceived, produced, and diffused across space.

GPTs are historically associated with waves of complementary and related innovations that reshape industries and technological paradigms (Dosi, 1982; Bresnahan & Trajtenberg, 1995). Likewise, AI has the potential to transform the innovation process itself by augmenting knowledge discovery and enabling the discovery of new technological trajectories (Agrawal et al., 2019; Cockburn et al., 2019). Its significance, therefore, extends beyond productivity gains and labor-market effects (Vivarelli, 1995; Acemoglu, 2025; Staccioli & Virgillito, 2025; Vivarelli & Arenas Díaz, 2025; Lábaj et al., 2025), as it may fundamentally reshape the mechanisms through which technological change occurs.

Economists have argued that AI's enabling characteristics (Trajtenberg, 2019; Brynjolfsson, 2023) create unprecedented opportunities for innovation, fostering new ventures (D'Alessandro et al., 2026), new sectors (Neffke et al., 2018), and the process of creative destruction (Schumpeter, 1939). By enabling the introduction of new products and services, these dynamics may ultimately offset the short-term adjustment costs associated with technological change (Carbonara & Santarelli, 2023). While early studies focused on the AI–innovation nexus at the firm level (Grashof & Kopka, 2023; Babina et al., 2024), AI is increasingly viewed as a key component of regional innovation systems (Iammarino, 2005) and a catalyst for structural transformation (Lazzeretti et al., 2022). This raises the question of how AI shapes regional innovation trajectories and conditions the influence of pre-existing local capabilities.

This paper investigates how AI interacts with regional technological relatedness in shaping innovation across European regions and technological fields. The analysis is grounded in the Evolutionary Economic Geography (EEG) literature, which explains the spatial unevenness of innovation through historical legacies, cognitive proximity, and path dependence (Boschma & Lambooy, 1999; Iammarino, 2005; Boschma & Frenken, 2006). From this perspective, we posit that regions are better able to generate innovations in technologies that are closely related to their existing knowledge base, as local search and cognitive proximity reduce recombinatorial costs and uncertainty (Boschma, 2005; Boschma & Frenken, 2011; Balland et al., 2015). Regional technological portfolios, therefore, both enable and constrain innovative efforts, as history matters (David, 1985; Dosi & Nelson, 2023).

AI may alter this path-dependent process. Beyond generating technological complementarities, it can facilitate search, experimentation, and knowledge recombination across cognitively distant domains, potentially enabling regions to innovate beyond the boundaries of existing expertise (Cockburn et al., 2019; Cicerone et al., 2023). AI may thus broaden the scope of regional innovation while loosening the dependence of inventive activity on the pre-existing technological portfolio.

To the best of our knowledge, this study provides the first comprehensive analysis of how regional technological relatedness and local AI endowments jointly shape innovation across technological fields, proxied by regional patent production in 670 four-digit Cooperative Patent Classification (CPC) classes. We examine whether AI fosters regional innovation directly, while also moderating the relationship between relatedness and innovation. Empirically, we combine three complementary patent databases to geolocate patent families filed by European inventors between 1986 and 2021, map regional AI endowments using a deep-learning procedure extending Van Roy et al.'s (2020) framework, measure technological relatedness through a density index following EEG studies, and estimate three-way fixed-effects Pseudo Poisson maximum-likelihood (PPML) models to test our hypotheses.

We contribute to the literature at the intersection of EEG and the Economics of AI (Boschma & Frenken, 2006; Agrawal et al., 2019; Agrawal et al., 2022) in three ways. First, we show that regions innovate more intensively in technological fields that are closely related to their existing knowledge base, in line with the EEG literature and the notion of local search. Second, we develop a conceptual framework in which AI not only supports innovation across technological fields but also alters the extent to which inventive activity depends on technological relatedness. Third, we test these arguments using a patent-based dataset comprising more than one million region–technology–period observations across European regions between 1986 and 2021. Supported by extensive robustness checks, we find that larger local AI endowments increase patenting across technological fields and negatively moderate the relationship between relatedness and innovation. These findings support the view of AI as an enabling technology that not only raises innovative output across technological fields but also relaxes the constraints imposed by local proximity, allowing regions to generate new technological knowledge in more distant domains.

The paper is organized as follows. Section 2 sets out the theoretical framework. Section 3 describes the data construction and the empirical strategy. Section 4 reports the main results along with robustness checks. Section 5 concludes by summarizing the outcomes of this study, discussing its limitations, and outlining avenues for future research.

2. Theoretical setting

From Schumpeter's theory of creative destruction ([1911] 1934, 1939) to more recent work on GPTs (Bresnahan & Trajtenberg, 1995) and technological paradigms (Dosi, 1982), scholars have recognized that transformative technologies reshape not only production processes but also the geography and organization of innovation. In this lineage, AI might represent a paradigmatic shift (Damioli et al., 2025), given its pervasive application and capacity to accelerate knowledge recombination, potentially altering the conditions under which innovation emerges within regions. Yet, as historical experience has shown, technological potential alone does not determine its impact (Solow, 1987); rather, outcomes depend on how new GPTs interact with pre-existing structures of knowledge and capabilities historically accumulated within places. This section examines AI as an enabling technological system and situates its possible effects within the EEG tradition, where the historical structure of the regional knowledge base is seen as a critical determinant of innovation.

2.1. Cognitive Proximity, Relatedness, and Technological Innovation

The production of new knowledge is commonly understood as a recombinatorial and cumulative process that builds on previously accumulated capabilities (Schumpeter, 1939; Weitzman, 1998). Because technological knowledge is localized and partly tacit (Audretsch & Feldman, 1996; Breschi & Malerba, 1997), regional capabilities are unevenly distributed and shape the range of technological combinations that regions can pursue (Hidalgo & Hausmann, 2009). As a result, regional innovation is inherently path-dependent (Boschma & Frenken, 2011), following trajectories constrained by existing knowledge structures (Boschma, 2005).

Building on Jacobs (1969), scholars have argued that local technological variety enlarges the pool of knowledge available for recombination, thereby fostering innovation (Iammarino, 2005; Colombelli & Quatraro, 2018). However, the innovative value of such variety depends on the degree of cognitive proximity among knowledge inputs (Boschma, 2005). The related-variety literature argues that innovation is most likely to emerge when knowledge inputs are sufficiently diverse to generate novelty yet sufficiently related to be effectively combined (Frenken et al., 2007; Boschma & Iammarino, 2009). The local availability of related knowledge inputs (i.e., cognitively close)[1] reduces uncertainty and search costs, facilitating the recombination of existing capabilities across connected technological domains (Boschma, 2005; Caragliu & Nijkamp, 2016). This mechanism is of crucial importance because search costs are unevenly distributed across the technological space. As search costs increase sharply with cognitive distance and beyond the boundaries of existing expertise (Fleming, 2001), proximate technological domains provide more accessible opportunities for recombination, favoring cross-fertilization and the creation of new combinations[2].

[1] Knowledge inputs are considered related when they are either complementary (i.e., frequently combined in the production of new knowledge) or cognitively similar, requiring comparable capabilities for their application (Breschi et al., 2003; Hidalgo et al., 2018).

[2] See Kogler et al. (2013), Tavassoli and Carbonara (2014), Castaldi et al. (2015), and Miguélez and Moreno (2018), who consistently find that regions with higher levels of related technological variety exhibit stronger innovative performance.

The concept of relatedness builds on this logic, shifting attention from the overall composition of the regional knowledge base to the position of specific technologies within that technological space (Boschma, 2017; Whittle & Kogler, 2020). It captures the extent to which a focal technological domain can draw on related capabilities, routines, and knowledge inputs already present in the regional portfolio. Because agents tend to favor local over distant search during the discovery of new combinations (Nelson & Winter, 1982), new innovative efforts are generally closely related to and reliant on pre-existing regional knowledge and capabilities (Boschma & Gianelle, 2014; Boschma et al., 2015; Boschma, 2017; Hidalgo et al., 2018). In other words, since innovation is inherently recombinant, inventors tend to explore domains that are cognitively close to the regional technological portfolio, as they are more likely to recognize relevant opportunities, mobilize complementary knowledge inputs, and combine existing capabilities into new inventions. Conversely, when a technological field is distant from the regional knowledge base, inventive activity is expected to be more difficult, as actors face higher cognitive barriers, weaker complementarities, and greater uncertainty in the search process. While this argument has been most extensively developed in the regional branching literature[3], the implications of its underlying mechanisms also apply to the intensive margin of knowledge production. Recent evidence by Boschma et al. (2023) supports this idea, showing that even breakthrough innovations, despite their transformative nature, often result from local search and are produced more frequently and intensively when they are technologically related to a region's existing knowledge base.

These arguments suggest that relatedness shapes the intensity of innovative activity at the region–technology level. When a focal technology is cognitively close to the regional portfolio, local actors can draw on familiar capabilities, leverage complementary knowledge inputs, and search more effectively for new combinations. Regions are therefore expected to generate more innovative output in technological fields that are more closely related to their pre-existing knowledge base. This leads to our first hypothesis:

Hypothesis 1: *Regions exhibit higher levels of innovative output (e.g., patent counts) in technological domains that are more closely related to their preexisting technological portfolio.*

2.2. Artificial Intelligence and Innovation

AI encompasses technologies that integrate computational, learning, and perceptual capabilities to process data, adapt to new information, and perform cognitively demanding tasks such as perception, reasoning, decision-making, and physical actuation (LeCun et al., 2015; Russell & Norvig, 2016;

[3] The diversification literature examines how regions develop new technological specializations, supporting the relatedness hypothesis. Using USPTO patent data, Boschma et al. (2015) show that regions are more likely to diversify into technologies related to their existing knowledge base. Balland et al. (2019) further integrate the notion of complexity (Hidalgo & Hausmann, 2009) to guide diversification strategies. Relatedness has also been shown to foster regional specialization in specific technological domains, including nanotechnologies (Colombelli et al., 2014) or fuel cell technologies (Tanner, 2014).

Sejnowski, 2018; High-level Expert Group on Artificial Intelligence, 2019). Following previous studies (Cockburn et al., 2019; Van Roy et al., 2020; Damioli et al., 2024; D'Alessandro et al., 2026), we adopt a broad definition of AI that includes both software-based technologies (e.g., machine learning, natural language processing, computer vision) and hardware-based tools (e.g., robotics, autonomous systems).

As a GPT, AI combines pervasive applicability with rapid technological progress, giving it the potential to reshape the overall invention–innovation–diffusion process. Rather than acting solely as an input into existing processes, AI also functions as an enabling technology that influences the direction, scope, and intensity of innovation across domains through several interconnected mechanisms.

AI is often characterized as a "macroinvention" (Mokyr, 1990), distinguished by pervasive applicability, continuous technical improvement, and "innovational complementarities" that enhance research and innovation across core, complementary, and using sectors (David, 1990; Bresnahan & Trajtenberg, 1995; Trajtenberg, 2019). Advances in learning, symbolic, perceptual, robotic, and autonomous system technologies extend the capabilities of intelligent systems, simultaneously spurring innovation within AI and raising the returns to developing complementary technologies and application-specific systems in other domains. At the same time, the technical requirements and problems emerging from application-oriented and complementary domains stimulate further improvements in AI systems, generating cumulative, cross-sectoral feedback loops (Bekar et al., 2018; Goldfarb et al., 2023). In sum, AI acts as a "technological bridge" that connects often distant knowledge bases and accelerates knowledge creation beyond the AI field itself, expanding the technological opportunity set available to innovators (Liu et al., 2020; Grashof & Kopka, 2023).

AI has also been described as a general-purpose IMI, namely a "meta-technology" that transforms not only the direction of technological change but also the process through which new knowledge is produced (Griliches, 1957; Rosenberg, 1998; Cockburn et al., 2019). Since innovation is inherently recombinatorial (Weitzman, 1998) and scientific discovery involves multi-stage search over complex combinatorial spaces (Fleming, 2001; Fleming & Sorenson, 2004; Agrawal et al., 2023), the accelerating accumulation of scientific knowledge raises search costs, uncertainty, and the "knowledge burden" (Bloom et al., 2020), thereby exacerbating the "needle-in-a-haystack" problem (Agrawal et al., 2019; Bianchini et al., 2022)[4]. AI holds promise for mitigating both the knowledge burden and the uncertainty associated with innovative activities (Agrawal et al., 2022). Machine and deep learning methods mark a shift from inquiry- to data-driven scientific search, organizing the search space, prioritizing the most valuable (or, successful) combinations, and discovering unexploited technological opportunities (OECD, 2019, 2023; Agrawal et al., 2023, 2024)[5]. At the same time, robotics and autonomous systems increasingly automate laboratory experimentation and hypothesis testing, as illustrated by robot scientists such as

[4] For instance, Wang et al. (2023) argue that the wide variety of hypotheses to be formulated and tested might seriously hamper the scientific process in the biochemistry field, where an estimated 10^{60} drug-like molecules exist to explore.
[5] Machine and deep-learning technologies have already demonstrated value in preclinical drug discovery (Catacutan et al., 2024), gravitational-wave detection (Gabbard et al, 2022), or biomolecular interactions analysis (Abramson et al., 2024).

Adam in genomics and Eve in drug discovery (King et al., 2009; Williams et al., 2015), which autonomously test hypotheses in parallel and execute closed-loop scientific research through cycles of experimentation. This has led to the emergence of "self-driving labs" (Purificato et al., 2025; Agrawal et al., 2024). By reducing search, experimentation, and coordination costs, AI can strengthen knowledge spillovers, absorptive capacity, and R&D productivity across scientific and technological fields (Cohen & Levinthal, 1990; Johnson et al., 2022; Truong & Papagiannidis, 2022).

Finally, AI adoption can raise labor productivity and operational efficiency, allowing firms to reallocate labor and capital toward more knowledge-intensive activities, including explorative R&D (Damioli et al., 2021; Xu et al., 2025). These gains may increase firms' capacity to invest in innovation, absorb external knowledge, and experiment with new technological trajectories (Liu et al., 2020).

The innovative benefits of AI are unlikely to be evenly distributed across space. Regions with stronger AI-related technological capabilities are better positioned to exploit AI's enabling properties, whereas those lacking such endowments may experience slower and more limited gains. The relationship between AI and innovation should therefore be strongest where AI knowledge and capabilities are locally abundant (Liu et al., 2020; Cicerone et al., 2023; Rodríguez-Pose & You, 2024). These effects are reinforced by the localized and path-dependent accumulation of capabilities (Audretsch & Feldman, 1996): regions with an established record of AI development are more likely to host specialized skills, firms, institutions, and networks that attract further AI-related investments. Spatial proximity may also foster diffusion and adoption: geographical distance between technology producers and users can intensify cognitive barriers and hinder knowledge transfer (Felice et al., 2022), while firms located in AI-production hubs may face stronger mimetic pressures and exhibit higher rates of "bandwagon adoption" (Dahlke et al., 2024). Greater regional AI endowments may therefore reinforce innovation both directly, by enhancing local technological capabilities, and indirectly, by accelerating the diffusion and adoption of AI. Overall, this leads to our second hypothesis:

Hypothesis 2: *The local endowment of Artificial Intelligence is positively correlated with regional innovation across technological fields.*

2.3. AI, relatedness, and innovation: a complex relationship

The previous sections establish that relatedness facilitates recombinant search by reducing uncertainty and search costs while maximizing recombinatorial payoffs. AI, as a horizontal and general-purpose meta-technology, can expand the range of combinations that actors are able to identify and exploit. The question is, therefore, whether AI adds to the innovative advantages associated with relatedness or also alters the extent to which innovation depends on the existing regional knowledge base.

First, AI's pervasive toolkit generates technological complementarities that sit atop existing regional portfolios and materially expand the space of viable recombinations. Machine-learning models and autonomous systems, for example, can connect previously unlinked data sources and technological domains, surfacing latent synergies that pure "relatedness" logics would miss. Even in fields already characterized by strong local relatedness, the local endowment of AI tools can uncover non-obvious complementarities among locally available technologies, thereby providing an "extra buffer" of ideas (Montresor & Quatraro, 2017) and making new combinations practically feasible (Agrawal et al., 2023). By acting as a "bridging platform" (Grashof & Kopka, 2023) across technologies, AI can therefore uncover combinations that are not fully captured by the measured proximity of a focal technology to the regional portfolio. Therefore, even after conditioning on local relatedness, greater AI endowments should consequently be associated with additional, extra innovative output. In this respect, we hypothesize:

Hypothesis 3: *Given the average regional relatedness, higher AI endowments continue to exert a positive effect on regional innovation across technological fields.*

At the same time, AI may weaken the extent to which inventive activity depends on relatedness by lowering the cognitive barriers to exploring distant segments of the technological space (Montresor & Quatraro, 2017; Cockburn et al., 2019; Agrawal et al., 2023). Innovation in technologically distant fields is typically more difficult because local actors lack proximate capabilities, complementary inputs, and established cognitive linkages. AI can partly compensate for these limitations by organizing vast knowledge spaces, connecting otherwise distant knowledge pools, and identifying productive combinations that would be difficult to detect through conventional local search (Fleming & Sorenson, 2004; Agrawal et al., 2019, 2023). In this sense, AI may produce a "cognitive expansion" effect, allowing regions to explore technological domains that are weakly connected to their existing portfolios. AI may therefore operate in a relatedness-substituting manner (Montresor & Quatraro, 2017), enabling actors to produce innovations in technological fields that may not build on a solid local knowledge infrastructure, from which they would normally emerge through a canonical recombinatorial process. In this respect, AI tools constitute a new "playbook" for innovation itself (Cockburn et al., 2019) that facilitates search beyond established trajectories and opens access to high-potential but previously under-exploited technological niches. By reducing the constraints imposed by local search, AI may broaden the regional innovation frontier and help regions move beyond path dependence and potential technological lock-ins. Therefore, we hypothesize:

Hypothesis 4: *The local endowment of Artificial Intelligence negatively moderates the positive effect of relatedness on regional innovation across technological fields.*

3. Data and methodology

We assemble a panel of region–technology–period observations (*i.e.,* a three-way panel), drawing on patent data (Pavitt, 1985; Griliches, 1998; Acs et al., 2002). Patents are particularly suited to this purpose, as they provide a widely used and fine-grained measure of inventive activity as well as a consistent proxy for technological innovation (Pavitt, 1985; Griliches, 1998; Acs et al., 2002). Moreover, they enable us to geolocate the technological knowledge they embody and classify it across technological domains.

We leverage three complementary data sources: EPO's PATSTAT, USPTO's PatentsView, and OECD's Regpat. Starting from PATSTAT, we create a Patent Universe (PU) that comprises all DOCDB families[6] with at least one application filed at either the EPO, WIPO, or USPTO[7], and having a priority year between 1986 and 2021[8]. For our analysis, this PU is then merged with PatentsView and Regpat databases, which provide detailed information on the location of both patents' applicants and inventors[9].

The construction of the three-way panel relies on two key pieces of information contained in patent data. First, we geolocate patent families using inventors' addresses, which usually indicate the locus where the invention was created, such as a laboratory, research establishment, or the place of residence of the inventor (Jaffe et al., 1993; Maraut et al., 2008; de Rassenfosse et al., 2019). Following the geolocation exercise, we only retained patent families having at least one inventor residing in Europe (irrespective of nationality). Second, we exploit the technological classifications assigned to patents by independent examiners. Since each patent is associated with one or more CPC codes, we identify all four-digit CPC technological classes represented in the sample. The use of CPC subclasses is consistent with previous studies (Boschma et al., 2023) and offers a suitable level of aggregation, avoiding both excessive technological detail and overly broad technological categories. To smooth spikes and idiosyncratic annual fluctuations, we aggregate the dataset into nine non-overlapping four-year periods[10], following prior work (Balland et al., 2019). As a result, the final sample covers 670 distinct 4-digit CPC classes in 302 European NUTS-2 regions[11], observed through nine four-year periods spanning 1986–2021.

Our decision to employ such a detailed level of analysis is consistent with previous studies (Boschma et al., 2023; Cicerone et al., 2023) and is grounded in both theoretical and empirical considerations. First,

---

[6] The DOCDB ("Documentation Database") patent family groups all documents that share exactly the same priority or combination of priorities (i.e., they originate from the same original filing). Since applications within the same DOCDB families claim the same "active" priority, they are considered to have the same technological content (Martínez, 2010). Considering patent families, rather than simply patents, avoids double-counting effects.

[7] Considering US patents, along with those applied at EPO and WIPO, is of crucial importance for better identifying the local endowment of AI (see Section 3.3). Indeed, software-based patents are commonly accepted in the US, while, as per EPO's guidelines, they need to have a clear technical character that distinguishes them from computer programs "as such" (EPO, 2020; Dernis et al., 2021).

[8] The priority year reflects the date of the first patent filing, making it the most accurate proxy for the moment in which the invention was originally conceived.

[9] PatentsView offers data for patents applied at the USPTO, while Regpat for filings at the EPO and WIPO.

[10] Periods are defined as follows: 1986-1989; 1990-1993; 1994-1997; 1998-2001; 2002-2005; 2006-2009; 2010-2013; 2014-2017; 2018-2021.

[11] We adopted the 2021 NUTS version.

by examining each four-digit CPC class within each NUTS-2 region and period, we can test whether AI behaves as a true enabler, facilitating innovation not just in aggregate, but across the full spectrum of distinct technological domains. Second, this structure allows us to construct a highly precise measure of knowledge relatedness (see Section 3.2).

3.1. Innovative output

We proxy the innovative output of European NUTS-2 regions using patent family counts. As emphasized by Pavitt (1985) and Griliches (1998), patent data are affected by several limitations. In addition to strategic usage, not all inventions are patented because firms can rely on alternative appropriability mechanisms such as secrecy or lead-time advantages, and not all patented inventions are subsequently commercialized as innovations. Moreover, patenting propensities vary considerably across sectors and technologies. As a result, patent data provide only a partial representation of innovative activity. However, as far as our analysis is concerned, patenting activity is the most reliable indicator, capable of providing a consistent mapping of regional innovative efforts.

We leverage the structure of our three-way panel to quantify regional technological output in each technological domain. Therefore, our dependent variable PAT_{jrt} is the count of patent families that region *r* produces in a four-digit CPC class *j* during period *t*[12]. In this regard, this variable is region- and technology-specific, as in Boschma et al. (2023). To avoid endogeneity concerns, we exclude all AI-classified patents from the variable PAT_{jrt}. These patents instead serve exclusively to construct our measure of regional AI endowment (Section 3.3).

3.2. Relatedness Density Indicator

As elaborated above, our empirical strategy hinges on precisely measuring how close each technology in a region aligns with its existing knowledge base. We operationalize this concept through the relatedness density indicator (Hidalgo et al., 2007), which enables us to quantify the cognitive proximity between a focal four-digit CPC class and the region's technological portfolio (Boschma et al., 2005; Boschma et al., 2015). To this purpose, we leverage the purpose-built *R* package EconGeo (Balland, 2017).

The starting point involves computing a technological space for each period, which incorporates the actual proximity between technologies. As noted by Qiao & Li (2025), technological proximity can be inferred in multiple ways: via geographic co-specialization (how often regions are simultaneously specialized in technologies *i* and *j*), inventor-level overlaps (how frequently the same inventors master both *i* and *j*), or co-occurrence within the same patent (how frequently technologies *i* and *j* appear within

[12] Following Boschma et al. (2023) and Caviggioli et al. (2023), patent families assigned to more than one CPC sub-class have been counted once for each technological field.

the same patent). As in Boschma et al. (2023), we perform a patent-level co-occurrence analysis to quantify the proximity among technologies. The proximity ϕ_{ijt} between technologies (four-digit CPC classes) *i* and *j*, in each period *t*, is given by:

$$\phi_{ijt} = \frac{m_t * c_{ijt}}{s_{it} * s_{jt}} \tag{1}$$

where c_{ijt} denotes the number of times technologies *i* and *j* co-occur in the same patent document during period *t*. Furthermore, s_{it} and s_{jt} are the total co-occurrence counts of technologies *i* and *j* with all other technologies, respectively (i.e., the degrees of nodes *i* and *j* in the undirected co-occurrence network over period *t*) defined as $s_{it} = \sum_{k \neq i} c_{ikt}$ and $s_{jt} = \sum_{k \neq j} c_{jkt}$. Finally, m_t denotes the aggregate number of edges in the network in period *t*. This measure, called "Association Strength", was first proposed by van Eck & Waltman (2009) and has been proven to better capture similarity (proximity) among items without being biased by "size" effects. Given the properties of this measure, a value of $\phi_{ijt} > 1$ means that technologies *i* and *j* co-occur more than expected in period *t*, reflecting a high proximity among the two. For each period, we build proximity matrices of dimension 670 × 670, where entries $\gamma_{ijt} = 1$ if $\phi_{ijt} > 1$[13]. Building on the technology-level proximities, we construct the region–technology–period relatedness density in two steps. First, we define each region's technological portfolio as the set of technologies in which it has developed relatively strong capabilities, following prior studies (Boschma et al., 2015; Balland et al., 2019). To this end, we employ the revealed technology advantage (RTA), a Balassa-style specialization index that identifies the technologies in which a region is comparatively specialized (Balassa, 1965). This is given by:

$$RTA_{jrt} = 1 \quad if \quad \frac{PAT_{jrt} \Big/ \sum_j PAT_{jrt}}{\sum_r PAT_{jrt} \Big/ \sum_r \sum_j PAT_{jrt}} > 1 \quad otherwise\ 0 \tag{2}$$

In this expression, the numerator is region *r*'s share of its own patenting activity devoted to technology *j*, while the denominator is the share of technology *j* in the total European patenting pool. Thus, an RTA_{jrt} that equals 1 indicates that region *r* is relatively specialized in technology *j*, as its share is higher than the European one. The technological portfolio is given by all technologies *j* for which region *r* has an $RTA_{jrt} = 1$. For each period *t*, RTA matrices of dimension 302 × 670 are thus built, where entries equal 1 if a region *r* is specialized in technology *j*. Second, we combine proximity and RTA matrices to derive a final indicator at the technology-region-period level (REL_DENS_{jrt}). Formally:

[13] This procedure entails the dichotomization of the proximities.

$$REL\_DENS_{jrt} = \frac{\sum_{i,i\neq j}\gamma_{ijt} * RTA_{irt}}{\sum_{i,i\neq j}\gamma_{ijt}} \times 100 \quad (3)$$

The numerator aggregates the proximities between *j* and those technologies the region already "masters" ($RTA_{irt} = 1$), while the denominator normalizes by *j*'s total proximity to all other technologies. Higher values indicate that a larger fraction of *j*'s related technologies overlaps with the region's existing technological portfolio[14]. Therefore, a higher relatedness density value means that the region is, in cognitive terms, better positioned to develop innovations in *j* due to a higher proximity and recombinatorial scope.

The advantage of this measure is that it does not define technological relatedness ex-ante (Boschma et al., 2012); rather, it is inferred from the structure of the technological space. Therefore, such a measure is better able to capture the full range of potential relationships among technologies. Second, this index further allows for an assessment of whether technologies available at the regional level are related to those underpinning new regional innovative efforts.

3.3. The local endowment of AI technologies

AI-related developments have been proxied by extant studies using a variety of data sources, each capturing a different dimension of the phenomenon[15]. As we aim to measure regional technological endowments related to AI, patent data are particularly suited to our empirical setting. Patents provide codified information on applied inventions, offering insights into which AI technologies are moving toward market deployment. This enables us to map the trajectory of applied AI research and innovation. The local endowment of AI technologies is proxied using a newly constructed dataset of AI-related patent families, identified through a deep-learning procedure described in the online Appendix A.

The procedure builds on the broad and comprehensive definition of AI proposed by previous studies (Van Roy et al., 2020; Damioli et al., 2024), which considers AI as a heterogeneous technological field encompassing learning, symbolic, data-processing, perception, robotics, control, and autonomous-system technologies. In the present study, rather than relying exclusively on predefined patent classification codes or simple keyword searches, AI patents are identified through an automated patent landscaping strategy (Abood & Feltenberger, 2018) combined with a BERT-based text classifier.

[14] Suppose technology *j* is strongly related ($\gamma = 1$) to ten other technologies (i.e., the denominator is the sum of their proximity weights), and region *r* is specialized in five of those ten. This means that 50% of the technologies related to *j* are already present in region *r*'s portfolio.

[15] AI has been measured using publication-, labor market-, web-, and patent-based indicators. Publications capture advances at the scientific frontier (Dunham et al., 2020; Bianchini et al., 2022); labor-market data measure firms' demand for AI skills and AI-related human capital (Alekseeva et al., 2021; Squicciarini & Nachtigall, 2021; Borgonovi et al., 2023; Babina et al., 2024); web-based approaches identify firms' AI adoption and diffusion patterns (Colombelli et al., 2023; Dernis et al., 2023; Dahlke et al., 2024, 2025); and patent-based indicators capture AI-related inventive activity and technological capabilities (Cockburn et al., 2019; Van Roy et al., 2020; Giczy et al., 2022; Miric et al., 2023; Cicerone et al., 2023; Grashof & Kopka, 2023; Pairolero et al., 2025). Some recent studies combine multiple sources to obtain a more comprehensive picture of AI diffusion (Calvino et al., 2022; Calvino et al., 2024).

The procedure starts from a hybrid top-down and bottom-up expansion of an existing AI keyword list. Keywords drawn from Van Roy et al. (2020) and Damioli et al. (2024) are used to retrieve AI-related scientific publications from Scopus, focusing on AI-related research fields[16]. The retrieved publications are then analysed through text-mining techniques to construct an extended vocabulary of 221 AI-related n-grams. This new, extended vocabulary is subsequently used to construct a high-precision seed set of AI patents from the PATSTAT-based Patent Universe[17], prioritizing patents whose titles and abstracts display a clear and semantically coherent AI signal. The seed set is then expanded through automated patent landscaping using keyword signals, CPC classifications, and family citations, while patents outside the resulting landscape provide the pool from which anti-seed documents are sampled (Abood & Feltenberger, 2018). The resulting seed and anti-seed documents are used to fine-tune a binary classifier based on the pre-trained BERT for Patents model (Srebrovic & Yonamine, 2020), which is applied to the expanded patent landscape to distinguish patents with substantive AI-related technological content from weakly related or false-positive candidates. As detailed in the online Appendix A, the final model achieves strong test-set performance, with an F1-score above 0.9. The final AI patent dataset is then constructed at the DOCDB family level, with a family classified as AI-related if at least one of its applications is retained by the model. As a result, the present deep-learning-based identification expands the keyword-only methodology used in prior empirical exercises (Van Roy et al., 2020; Damioli et al., 2021; Damioli et al., 2024).

From the resulting dataset, we identify all AI patent families produced by European inventors and having a priority year between 1986 and 2021, which serve as the basis for constructing the regional AI endowment measure. As emphasized in Section 3.1, these patents are deliberately excluded from the dependent variable to avoid endogeneity. To proxy the local AI endowments, we compute the AI technological stock employing the "permanent inventory method" (P.I.M.). The AI stock in region *r*, for each period *t*, is defined recursively as the previous period's stock after depreciation, plus newly produced AI-related knowledge in *t*. Formally:

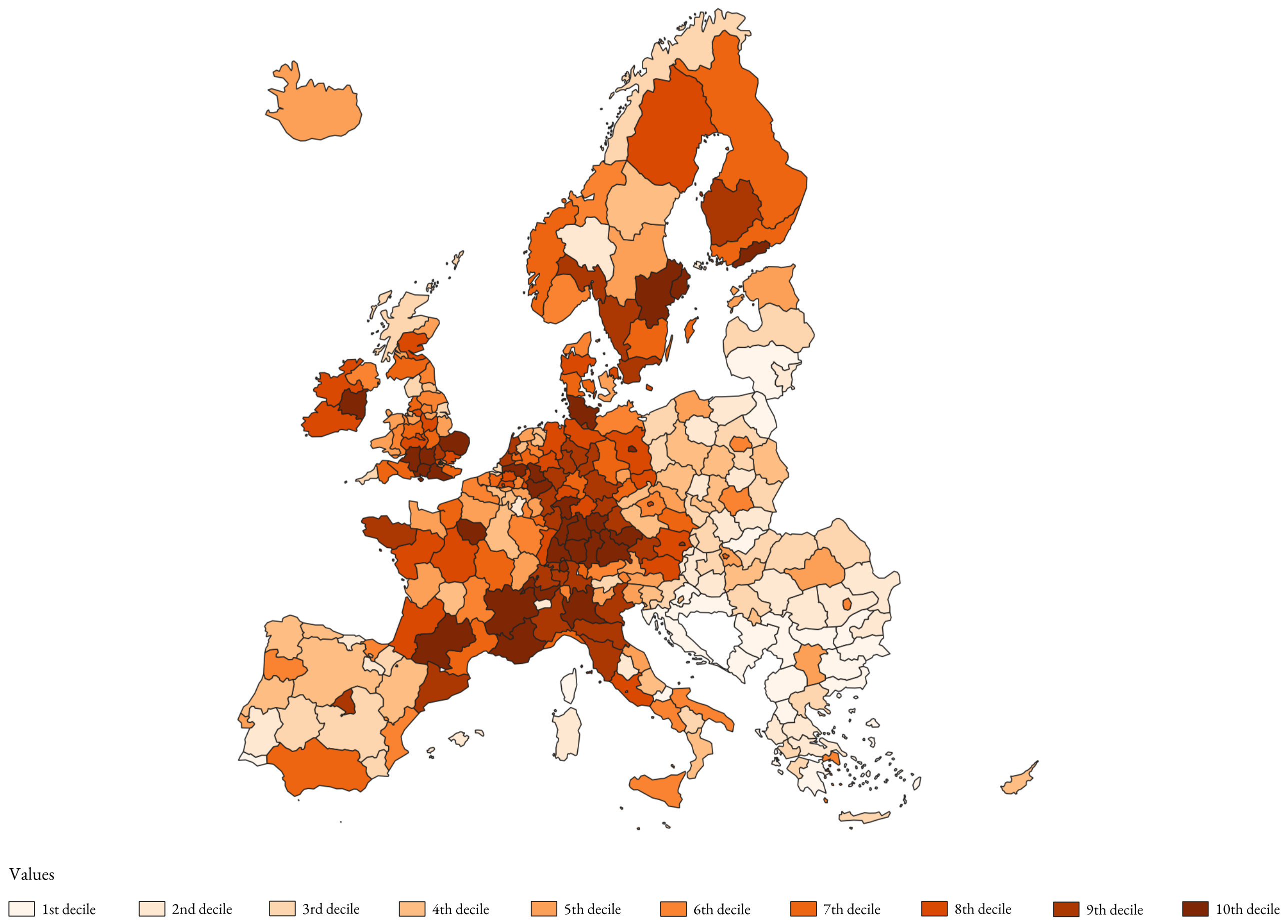
$$K\_AI_{rt} = \frac{K\_AI_{rt-1}}{(1+\delta)} + AI\_PAT_{rt} \tag{4}$$

with the depreciation rate $\delta = 0.15$ following Hall et al. (2005). Figure 1 displays the AI technological stock for the period 2014-2017, and shows a consistent picture compared to existing studies (Cicerone et al., 2023; Xiao & Boschma, 2023; Minniti et al., 2025).

16 AI-related fields are Computer Science, Engineering, and Mathematics (Bianchini et al., 2022).

17 As described in the online Appendix, the full Patent Universe assembled for the fine-tuning pipeline comprises patent families with priority years 1980–2021. For the empirical analysis reported here, however, we start the observation window in 1986 because earlier years exhibit substantial missingness in several crucial regional control variables.

Figure 1. Regional AI technological stock – period 2014-1017



3.4. Control variables

Following previous studies, we control for several factors affecting regional patenting capacity by introducing covariates at each level of analysis.

At the region–technology level, we include the revealed technology advantage RTA_{jrt} index introduced in equation (2) as a control. This variable captures regional specialization in technology j, thereby absorbing the portion of variation in patenting that is driven by the fact that the region is already comparatively strong in that field[18]. At the technology level, we control for the average number of claims per patent within each technological class and period ($TECH\_CLAIMS_{jt}$). On one side, it proxies the average inventive and recombinatorial scope within a technology. On the other side, it adjusts for differences in strategic behaviour of patenting across fields, as patents with more claims often seek

[18] We include both RTA_{jrt} and REL_DENS_{jrt} in the regressions to test whether the production of patents still relies on relatedness even after controlling for within-technology specialization. As innovation is understood as a recombinatorial process, we expect the REL_DENS_{jrt} indicator to retain an independent, positive association with the production of new technological knowledge (e.g., patents).

broader or more layered protection, which can be both a signal of perceived novelty/value and of legal positioning (e.g., defensive or value-maximizing strategies).

At the regional level, population density (POP_DENS_{rt}) controls for agglomeration economies and the localized exchange of knowledge, as denser regions may benefit from more frequent interactions, lower coordination costs, and stronger innovation clusters (Frenken et al., 2007; Mewes, 2019). In addition, per capita gross domestic product (GDP_PC_{rt}) is included to control for the overall economic development and resource availability of regions. Higher GDP per capita proxies for greater capacity to invest in innovation (both public and private R&D), better physical and institutional infrastructure, more developed human capital endowments, and stronger local demand for new technologies, all of which are well-established drivers of inventive activity. It also captures broader absorptive capacity: wealthier regions tend to have organizations and individuals better equipped to recognize, assimilate, and apply new knowledge across technological fields. We also include the share of industrial employment[19] (IND_S_{rt}) to account for differences in regional productive structure, since manufacturing and other industrial sectors tend to be comparatively R&D-intensive and more likely to rely on formal intellectual-property protection. This variable helps isolate the effects of relatedness and AI endowment from confounding variation due to the underlying industrial structure. In the robustness analyses, we replace the industrial-employment measure with the regional share of employment in high-technology manufacturing and knowledge-intensive services[20] (HT_S_{rt}), which should serve as a closer proxy for advanced production and knowledge-intensive industry. Prior work has also shown that high-tech sectors rely more heavily on IP protection (Mezzanotti & Simcoe, 2023): this alternative control helps ensure our results are not driven by the regional sectoral composition at the technological frontier. We also control for the regional R&D intensity ($R\&D\_INT_{rt}$), measured as R&D expenditure relative to regional gross value added, which reflects a greater commitment to knowledge creation through investments in experimentation and skilled labor, and, via enhanced absorptive capacity, improves the region's ability to recognize, assimilate, and apply new knowledge (Griliches, 1979; Cohen & Levinthal, 1990; Acs et al., 2002). Because the high-technology employment and R&D-intensity series contain substantial missing values, they are excluded from the preferred baseline specification and introduced only in robustness checks by estimating the model on the subset with non-missing data to assess the sensitivity of our results. As detailed in Section 4.1, results remain unaffected.

Table 1 provides a comprehensive list of variables. As already mentioned, after removing missing values, the final dataset resulted in an unbalanced three-way panel with 670 distinct 4-digit CPC classes in 302 European NUTS-2 regions[21], observed from 1986 to 2021, for a total of 1,473,330 observations.

---

[19] The industrial sector includes the following NACE rev. 2 one-digit codes: B (Mining and quarrying), C (Manufacturing), D (Electricity, gas, steam and air conditioning supply), E (Water supply; sewerage, waste management and remediation activities).
[20] The sectors included are the following NACE rev. 2 two-digit codes: 21, 26, 59 to 63, 72.
[21] The empirical analysis includes 34 countries: AT, BE, BG, CH, CY, CZ, DE, DK, EE, EL, ES, FI, FR, HR, HU, IE, IS, IT, LT, LU, LV, ME, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, UK.

Table 1. Variables description

Variable	Description	Source
PAT_{jrt}	Total number of patents in region *r* and technology *j* (4-digits CPC codes)	Authors' own elaboration based on EPO's PATSTAT, USPTO's PatentsView and OECD's Regpat databases
REL_DENS_{jrt}	Relatedness density of technology *j* in region *r*	
K_AI_{rt}	AI technological stock in region *r*	
$TECH\_CLAIMS_{jt}$	Average number of claims in technology *j*	
RTA_{jrt}	Dummy = 1 indicating region *r*'s specialization in technology *j*	
POP_DENS_{rt}	Regional Population density	Authors' own elaboration based on EU Directorate General for Regional and Urban Policy's ARDECO and Eurostat's Regional Statistics databases
GDP_PC_{rt}	Regional Gross Domestic Product per capita	
IND_S_{rt}	Share of industrial employment	
$R\&D\_INT_{rt}$	Ratio between regional R&D expenditures and Gross Value Added	
HT_S_{rt}	Share of High-Tech employment	

Table 2 reports summary statistics. To reduce skewness and account for multiplicative effects, the AI stock, Population Density, and GDP per capita are log-transformed. Finally, as a standard practice when empirically estimating interaction effects (as detailed in Section 3.5), we mean-center all our variables; this step is also crucial for testing our hypothesis 3.

Table 2. Summary Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
PAT_{jrt}	1473330	4.382	24.618	0	2904
REL_DENS_{jrt}	1473330	24.171	16.42	0	100
K_AI_{rt} (log)	1473330	2.921	1.962	0	7.976
$TECH\_CLAIMS_{jt}$	1473330	13.153	2.773	0	35.2
RTA_{jrt}	1473330	.196	.397	0	1
POP_DENS_{rt} (log)	1473330	4.957	1.264	.766	9.28
GDP_PC_{rt} (log)	1473330	9.964	.704	7.668	12.359
IND_S_{rt}	1473330	.181	.079	.015	.447
$R\&D\_INT_{rt}$	1293770	.013	.011	0	.092
HT_S_{rt}	905840	.039	.019	.006	.121

3.5. Econometric strategy

Given our dependent variable, our empirical analysis is based on count-data models (Cameron & Trivedi, 2013). While Poisson and Negative Binomial models are the standard alternatives (Wooldridge, 2010), the conditional fixed-effects Negative Binomial does not generally deliver a full set of fixed-effects purging time-invariant heterogeneity in panel settings and therefore has well-known limitations for true fixed-effects identification (Allison & Waterman, 2002). Moreover, a standard Poisson model with multiple high-dimensional fixed effects becomes computationally a hard task. We therefore estimate Pseudo-Poisson Maximum Likelihood models with high-dimensional fixed effects (PPML-HDFE) (Correia et al., 2020), which efficiently accommodate our three-way panel structure, absorb region, technology, and time fixed effects, and are robust to heteroskedasticity and overdispersion (Wooldridge, 1999; Guimarães, 2008). Robust standard errors are clustered at the NUTS-2 level.

We propose two models in which all the independent variables are lagged by one period to further mitigate endogeneity concerns[22]. In addition, as discussed above, all models include region, technology, and time-period fixed effects. To test the first two hypotheses, the following model is estimated:

$$E[PAT_{jrt} \mid \theta] = \exp(\beta_1 REL\_DENS_{jrt-1} + \beta_2 K\_AI_{rt-1} + \sum_{i=1}^{6} \beta X_{jrt-1} + \varphi_j + \omega_r + \tau_t + \varepsilon_{rjt}) \tag{5}$$

The coefficients of interest are β_1 and β_2, through which Hypotheses 1 and 2 are tested, respectively. A positive and statistically significant β_1 would lend support to Hypothesis 1, whereas a positive and statistically significant β_2 would provide evidence in favor of Hypothesis 2.

To test hypotheses 3 and 4, we extend the model described in (5) by including an interaction term between the relatedness density indicator and the local AI technological stock variable. Formally:

$$\begin{aligned} E[PAT_{jrt} \mid \theta] = \exp(&\beta_1 REL\_DENS_{jrt-1} + \beta_2 K\_AI_{rt-1} + \beta_3 REL\_DENS_{jrt-1} \times K\_AI_{rt-1} + \sum_{i=1}^{6} \beta X_{jrt-1} \\ &+ \varphi_j + \omega_r + \tau_t + \varepsilon_{rjt}) \end{aligned} \tag{6}$$

In this case, the coefficients of interest are β_2 and β_3, which serve to test Hypotheses 3 and 4, respectively. A positive and statistically significant β_2 would corroborate Hypothesis 3: because all variables are mean-centred, β_2 identifies the effect of the local endowment of AI technologies when the relatedness density equals its mean (i.e., AI's conditional effect). A positive and significant β_2, therefore, implies that AI is still positively associated with regional patenting activity across technologies, even after conditioning on average relatedness. Finally, β_3 captures whether the local endowment of AI technologies amplifies ($\beta_3 > 0$) or substitutes for ($\beta_3 < 0$) the effect of relatedness. The substantive claim tested in the present study refers to a "relatedness-substituting" effect; therefore, we expect β_3 to be negative and significant.

4. Results

In this section, we begin by discussing the estimates from the preferred specification. Table 3 reports our baseline results: columns 1–3 report progressively augmented specifications of the model proposed in (5), and column 4 presents the specification that includes the interaction between relatedness density and the regional AI stock, as in equation (6).

In column 1, the coefficient on the relatedness density indicator is positive and statistically significant ($p<0.01$), indicating that regions produce more innovations in technological fields that were closer to

[22] While we do not claim causality, a lagged model is employed to mitigate simultaneity concerns, which in turn defines a clear temporal ordering between explanatory variables and patenting activity. In addition, as stated before, we further reduce endogeneity concerns by deliberately excluding AI-classified patents from the dependent variable.

their pre-existing portfolios. This extends the EEG argument on relatedness to the intensive margin of regional knowledge production and supports Hypothesis 1. Column 2 reports a positive and significant coefficient for the regional AI stock ($p<0.01$), consistent with the view of AI as a general-purpose and enabling technology that stimulates innovation across technological fields. When both variables are included in column 3, their coefficients remain positive and significant, supporting Hypotheses 1 and 2. The relatedness coefficient is virtually unchanged, whereas the AI coefficient declines but remains significant, suggesting that the two variables capture distinct, although partly overlapping, dimensions of regional inventive capacity. Column 4 introduces the interaction between relatedness and AI. Since both variables are mean-centred, the positive AI coefficient indicates that stronger AI endowments are associated with higher expected patenting, even after conditioning on average relatedness, supporting Hypothesis 3. Put differently, AI contributes as an independent enabler of inventive activity that is not simply a by-product of measured technological proximity. This pattern is consistent with the notion that AI's ability to surface latent complementarities across knowledge modules creates additional opportunities for novel combinations that pure "related search" would not reveal. Column 4 also highlights a negative and significant ($p<0.01$) interaction term between relatedness and AI, which implies that the returns to relatedness decline as local AI endowments rise. This finding supports Hypothesis 4 and, theoretically, is consistent with a relatedness-substituting mechanism: AI may facilitate search and recombination in technologically distant fields, thereby reducing the extent to which regional knowledge production depends on a technology's proximity to the regional pre-existing portfolio. Taken together, the results point to a dual role for AI: it raises inventive activity across fields while also relaxing the constraints imposed by local technological proximity and established technological trajectories. Therefore, AI does not reinforce existing paths or simply replace them; it both expands the frontier of feasible recombinations and reshapes the functional importance of relatedness and cognitive proximity in regional knowledge production.

Finally, the estimated coefficients of control variables conform to theoretical priors and are, overall, statistically significant. The positive and significant coefficient on the RTA index confirms that regions with prior specialization in a given technology tend to produce more patents in that field. Importantly, the relatedness density coefficient remains positive and significant after controlling for RTA, indicating that innovation in a focal technological field depends not only on specialization but also on the availability of cognitively proximate knowledge within the regional portfolio. The positive coefficient on average claims is consistent with greater recombinatorial scope within technological fields, whereas those on population density, GDP per capita, and industrial employment point respectively to the role of agglomeration economies, regional resources and absorptive capacity, and the stronger propensity of industrial regions to engage in patenting. Overall, these results strengthen confidence that the estimated effects of relatedness and AI endowments are not driven by omitted regional characteristics.

Table 3. Total Patents Specification; Pseudo Poisson Fixed-Effect model.

	(1)	(2)	(3)	(4)
REL_DENS	0.0198***		0.0197***	0.0240***
	(0.00127)		(0.00129)	(0.00119)
K_AI		0.232***	0.200***	0.206***
		(0.0295)	(0.0343)	(0.0345)
REL_DENS * K_AI				-0.00184***
				(0.000592)
RTA	0.815***	1.072***	0.816***	0.819***
	(0.0174)	(0.0323)	(0.0174)	(0.0171)
TECH_CLAIMS	0.0416***	0.0370***	0.0416***	0.0425***
	(0.00582)	(0.00443)	(0.00601)	(0.00605)
POP_DENS	1.343***	1.159***	1.280***	1.321***
	(0.267)	(0.202)	(0.252)	(0.260)
GDP_PC	0.827***	0.575***	0.620***	0.555***
	(0.130)	(0.105)	(0.130)	(0.129)
IND_S	1.730**	1.781***	1.911**	1.816**
	(0.819)	(0.677)	(0.743)	(0.742)
Constant	1.572***	1.434***	1.265***	1.263***
	(0.195)	(0.160)	(0.195)	(0.202)
Observations	1,473,330	1,473,330	1,473,330	1,473,330
Period FE	YES	YES	YES	YES
NUTS-2 FE	YES	YES	YES	YES
Tech. Field FE	YES	YES	YES	YES

Note: Robust Standard errors clustered at the NUTS-2 level in parenthesis. P-values are represented by: *** p<0.01, ** p<0.05, * p<0.10. The dependent variable is the count of patents in NUTS-2 regions by four-digit CPC technological classes, excluding those classified as AI. Independent variables are centred around their mean. All regressors are lagged by one period. All models include Period, NUTS-2 and Technological Field Fixed Effects.

4.1. Additional Analyses and Robustness Checks

We tested the sensitivity of our baseline findings through a series of robustness checks and complementary analyses, all of which confirm the main results.

First, we test whether the effects in the baseline and following estimates still hold once we exclude from the analysis the technological fields where complementarities are naturally strongest. In this respect, we can test whether AI at the local level really acts as a general-purpose framework to drive innovation above and beyond core and complementary AI-related fields. We re-estimate the baseline models excluding the most frequent AI-related technological classes, defined using the AI dataset described in the online Appendix (Table A6). Excluding the top 10 and, more stringently, the top 20 AI-related CPC codes removes the technological fields most likely to contain direct AI–related complementarities. The estimates reported in Table 4 show that our main findings still hold, even when patenting in core AI-related technological fields is omitted from the sample. AI's positive association with regional patenting and its moderating role vis-à-vis relatedness hold broadly across the technological spectrum.

We next examine whether the baseline relationships are confirmed after accounting more explicitly for regional innovativeness. To this end, we augment the preferred specification with regional R&D intensity and the high-tech employment share, first separately and then jointly. The estimated coefficients in Tables 5, 6, and 7 generally conform to theoretical expectations: R&D intensity is positive, although with varying significance, while the high-tech employment share is positive and significant, indicating that regions with stronger research effort and a larger concentration of high-tech activities tend to produce more patents. Importantly, the positive associations of relatedness and local AI endowments with patenting, as well as the negative interaction between them, remain stable across these extended specifications.

A third robustness check further probes whether the moderating role of AI extends beyond recombinatorial proximity and may also compensate for limited field-specific inventive capabilities. Evidence from the baseline specification suggests that AI-rich regions are less constrained by the structure of their existing technological portfolios when generating new patents, enabling them to navigate more distant segments of the technological space. Building on this finding, stronger AI endowments may also reduce the importance of established capabilities within a specific technological field for producing innovation in that domain. To investigate this possibility, we interact the regional AI stock with the RTA indicator, which captures prior specialization, experience, and routines in a given technology. The results reported in Table 8 show a negative and significant interaction coefficient across specifications. Consistent with the baseline interaction between AI and relatedness, this finding suggests that AI-rich regions are better able to navigate the technological space even when established capabilities in the focal technology are relatively limited. It therefore provides further evidence that AI reduces the dependence of regional patenting activity on pre-existing technological structures and capabilities.

A fourth robustness check addresses the intertwined relationship between AI and ICT technologies. AI development is deeply rooted in the broader digital, ICT-related knowledge base, as advances in software, computing, communication, and data-processing technologies provide many of the capabilities on which AI systems rely (Buarque et al., 2020; Xiao & Boschma, 2023). At the same time, the development and diffusion of AI can generate additional demand for complementary ICT innovations in hardware, software, networks, and digital platforms, creating mutually reinforcing technological trajectories at the regional level (Lee & Lee, 2021; Santarelli et al., 2023)[23]. Micro-level evidence further documents dense horizontal and vertical linkages between AI and ICT firms, including joint R&D partnerships, strategic alliances, and supplier–buyer relationships, which support their co-evolution within regional ecosystems (Turkina et al., 2021). To ensure our results are not driven by ICT-centric dynamics, we proceed in two complementary steps. First, as done in prior studies (D'Alessandro et al., 2026; Minniti et al., 2025), we identify ICT patent families using the taxonomy proposed by Inaba & Squicciarini (2017), and exclude them from the dependent variable. Second, we construct a regional ICT stock (analogous to our AI stock) and re-estimate the baseline by including it as an additional control. As reported in Table 9, the AI stock remains positively associated with patenting, while its interaction with relatedness remains negative, although at a lower significance level ($p<0.05$). These results indicate that the role of AI is not merely a by-product of ICT-centric dynamics but retains an independent association with innovation across technological fields.

A fifth additional analysis investigates the temporal stability of our results: do the relationships we document hold across time, and have they changed as AI matured and diffused? To investigate this, we split the sample into two subperiods (1986–2005 versus 2006–2021) and re-estimate our models (Table 10). Two notable patterns emerge, signalling a strengthening of AI's role in more recent years. In the earlier period, the AI stock and its interaction with relatedness are significant in the baseline model. Yet, both lose significance once ICT patents are excluded, and the specification is saturated with all our controls. This suggests that, in the formative phase of AI, its contribution to innovation remained closely intertwined with the broader ICT technological paradigm. By contrast, in the later period, the AI coefficient is larger and consistently significant, while the negative interaction with relatedness is stronger and remains significant across all specifications. These results indicate that, as AI matured, it developed a more autonomous and cross-cutting role, increasingly stimulating patenting across technological fields and reducing the dependence of inventive activity on pre-existing capabilities.

Finally, we assess whether our findings extend beyond the quantity of inventive output to the production of high-quality innovations. Patents differ substantially in their technological and economic significance, and only a small share exerts a disproportionate influence on subsequent technological developments (Hall et al., 2005; Squicciarini et al., 2013). In this robustness exercise, we therefore focus

[23] All of this is further confirmed by our empirical validation exercise presented in the online Appendix.

Table 4. Total Patents Specification excluding AI-related CPC codes; Pseudo Poisson Fixed-Effect model.

	Excl. top 10 AI CPC codes				Excl. top 20 AI CPC codes			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
REL_DENS	0.0201***		0.0200***	0.0237***	0.0200***		0.0199***	0.0240***
	(0.00117)		(0.00119)	(0.00114)	(0.00108)		(0.00110)	(0.00110)
K_AI		0.226***	0.199***	0.205***		0.226***	0.204***	0.211***
		(0.0304)	(0.0348)	(0.0346)		(0.0305)	(0.0344)	(0.0341)
REL_DENS * K_AI				-0.00158***				-0.00175***
				(0.000612)				(0.000635)
RTA	0.820***	1.069***	0.821***	0.823***	0.828***	1.071***	0.829***	0.832***
	(0.0180)	(0.0322)	(0.0181)	(0.0178)	(0.0182)	(0.0315)	(0.0183)	(0.0180)
TECH_CLAIMS	0.0314***	0.0268***	0.0314***	0.0321***	0.0266***	0.0226***	0.0267***	0.0272***
	(0.00512)	(0.00431)	(0.00521)	(0.00520)	(0.00451)	(0.00413)	(0.00456)	(0.00452)
POP_DENS	1.251***	1.045***	1.186***	1.203***	1.173***	0.972***	1.104***	1.111***
	(0.265)	(0.204)	(0.250)	(0.255)	(0.254)	(0.207)	(0.240)	(0.244)
GDP_PC	0.768***	0.510***	0.562***	0.499***	0.707***	0.446***	0.495***	0.419***
	(0.128)	(0.108)	(0.130)	(0.126)	(0.123)	(0.108)	(0.125)	(0.121)
IND_S	1.756**	1.822***	1.939***	1.852**	1.764**	1.811**	1.956***	1.859**
	(0.822)	(0.694)	(0.747)	(0.741)	(0.823)	(0.707)	(0.745)	(0.737)
Constant	1.561***	1.462***	1.260***	1.271***	1.577***	1.480***	1.275***	1.295***
	(0.190)	(0.161)	(0.193)	(0.197)	(0.181)	(0.165)	(0.186)	(0.191)
Observations	1,451,340	1,451,340	1,451,340	1,451,340	1,429,350	1,429,350	1,429,350	1,429,350
Period FE	YES	YES	YES	YES	YES	YES	YES	YES
NUTS-2 FE	YES	YES	YES	YES	YES	YES	YES	YES
Tech. Field FE	YES	YES	YES	YES	YES	YES	YES	YES

Note: Robust Standard errors clustered at the NUTS-2 level in parenthesis. P-values are represented by: *** p<0.01, ** p<0.05, * p<0.10. The dependent variable is the count of patents in NUTS-2 regions by four-digit CPC technological classes, excluding those classified as AI. In Columns 1-4, the top ten most frequent AI CPC classes are excluded; in Columns 5–8, a stricter criterion is applied, omitting the top twenty AI CPC classes. Independent variables are centred around their mean. All regressors are lagged by one period. All models include Period, NUTS-2 and Technological Field Fixed Effects.

Table 5. Total Patents Specification including R&D intensity; Pseudo Poisson Fixed-Effect model.

	Main specification		Excl. top 10 AI CPC codes		Excl. top 20 AI CPC codes	
	(1)	(2)	(3)	(4)	(5)	(6)
REL_DENS	0.0198***	0.0249***	0.0202***	0.0246***	0.0200***	0.0245***
	(0.00123)	(0.00110)	(0.00115)	(0.00111)	(0.00107)	(0.00111)
K_AI	0.199***	0.195***	0.199***	0.198***	0.206***	0.205***
	(0.0371)	(0.0374)	(0.0377)	(0.0376)	(0.0375)	(0.0374)
REL_DENS * K_AI		-0.00226***		-0.00196***		-0.00201***
		(0.000692)		(0.000701)		(0.000722)
RTA	0.818***	0.821***	0.824***	0.827***	0.833***	0.836***
	(0.0179)	(0.0175)	(0.0186)	(0.0183)	(0.0187)	(0.0184)
TECH_CLAIMS	0.0277***	0.0286***	0.0245***	0.0251***	0.0230***	0.0235***
	(0.00587)	(0.00573)	(0.00527)	(0.00516)	(0.00482)	(0.00472)
POP_DENS	1.276***	1.316***	1.182***	1.193***	1.097***	1.095***
	(0.270)	(0.278)	(0.268)	(0.272)	(0.259)	(0.263)
GDP_PC	0.577***	0.491***	0.514***	0.431***	0.441***	0.351***
	(0.136)	(0.131)	(0.135)	(0.127)	(0.130)	(0.123)
IND_S	1.593**	1.555**	1.581**	1.536**	1.561**	1.506**
	(0.746)	(0.748)	(0.752)	(0.751)	(0.750)	(0.748)
R&D_INT	2.512	3.254	2.978	3.668*	3.238	3.956**
	(2.110)	(1.996)	(2.068)	(1.954)	(2.014)	(1.902)
Constant	1.398***	1.419***	1.372***	1.404***	1.377***	1.419***
	(0.197)	(0.203)	(0.193)	(0.196)	(0.187)	(0.191)
Observations	1,293,770	1,293,770	1,274,460	1,274,460	1,255,150	1,255,150
Period FE	YES	YES	YES	YES	YES	YES
NUTS-2 FE	YES	YES	YES	YES	YES	YES
Tech. Field FE	YES	YES	YES	YES	YES	YES

Note: Robust Standard errors clustered at the NUTS-2 level in parenthesis. P-values are represented by: *** p<0.01, ** p<0.05, * p<0.10. The regional R&D intensity is included as a control. The drop in the number of observations is due to missing values. The dependent variable is the count of patents in NUTS-2 regions by four-digit CPC technological classes, excluding those classified as AI. In Columns 1-2, results of the main specification are shown. In Columns 3-4, the top ten most frequent AI CPC classes are excluded; in Columns 5–6, a stricter criterion is applied, omitting the top twenty AI CPC classes. Independent variables are centred around their mean. All regressors are lagged by one period. All models include Period, NUTS-2 and Technological Field Fixed Effects.

Table 6. Total Patents Specification including the High-Tech employment share; Pseudo Poisson Fixed Effects model.

	Main Specification		Excl. top 10 AI CPC codes		Excl. top 20 AI CPC codes	
	(1)	(2)	(3)	(4)	(5)	(6)
REL_DENS	0.0203***	0.0260***	0.0209***	0.0257***	0.0207***	0.0253***
	(0.00115)	(0.00128)	(0.00109)	(0.00130)	(0.00103)	(0.00130)
K_AI	0.205***	0.191***	0.226***	0.216***	0.238***	0.230***
	(0.0350)	(0.0372)	(0.0371)	(0.0389)	(0.0377)	(0.0393)
REL_DENS * K_AI		-0.00286***		-0.00243***		-0.00235**
		(0.000938)		(0.000937)		(0.000948)
RTA	0.822***	0.824***	0.829***	0.831***	0.839***	0.841***
	(0.0183)	(0.0178)	(0.0189)	(0.0185)	(0.0188)	(0.0184)
TECH_CLAIMS	0.0121**	0.0134**	0.0268***	0.0280***	0.0304***	0.0315***
	(0.00543)	(0.00547)	(0.00470)	(0.00475)	(0.00446)	(0.00455)
POP_DENS	1.532***	1.541***	1.391***	1.365***	1.227***	1.188***
	(0.319)	(0.318)	(0.322)	(0.316)	(0.292)	(0.287)
GDP_PC	0.636***	0.583***	0.523***	0.466***	0.397***	0.336**
	(0.164)	(0.164)	(0.161)	(0.159)	(0.148)	(0.147)
HT_S	2.771*	2.778*	2.391	2.374	2.871*	2.845*
	(1.547)	(1.580)	(1.500)	(1.497)	(1.467)	(1.455)
Constant	1.543***	1.589***	1.508***	1.563***	1.555***	1.617***
	(0.208)	(0.208)	(0.204)	(0.201)	(0.181)	(0.179)
Observations	905,840	905,840	892,320	892,320	878,800	878,800
Period FE	YES	YES	YES	YES	YES	YES
NUTS-2 FE	YES	YES	YES	YES	YES	YES
Tech. Field FE	YES	YES	YES	YES	YES	YES

Note: Robust Standard errors clustered at the NUTS-2 level in parenthesis. P-values are represented by: *** p<0.01, ** p<0.05, * p<0.10. In this specification, we included the High-Tech employment share as a control, in place of the Industrial Share. The drop in the number of observations is due to missing values. The dependent variable is the count of patents in NUTS-2 regions by four-digit CPC technological classes, excluding those classified as AI. In Columns 1-2, results of the main specification are shown. In Columns 3-4, the top ten most frequent AI CPC classes are excluded; in Columns 5–6, a stricter criterion is applied, omitting the top twenty AI CPC classes. Independent variables are centred around their mean. All regressors are lagged by one period. All models include Period, NUTS-2 and Technological Field Fixed Effects.

Table 7. Total Patents Specification including R&D Intensity and H.T. employment share; Pseudo Poisson Fixed-Effect model.

	Main Specification		Excl. top 10 AI CPC codes		Excl. top 20 AI CPC codes	
	(1)	(2)	(3)	(4)	(5)	(6)
REL_DENS	0.0203***	0.0261***	0.0209***	0.0257***	0.0207***	0.0253***
	(0.00115)	(0.00126)	(0.00110)	(0.00128)	(0.00104)	(0.00128)
K_AI	0.199***	0.184***	0.219***	0.209***	0.231***	0.221***
	(0.0345)	(0.0367)	(0.0368)	(0.0386)	(0.0375)	(0.0391)
REL_DENS * K_AI		-0.00288***		-0.00246***		-0.00240**
		(0.000935)		(0.000935)		(0.000946)
RTA	0.822***	0.824***	0.829***	0.831***	0.839***	0.841***
	(0.0182)	(0.0178)	(0.0189)	(0.0185)	(0.0188)	(0.0184)
TECH_CLAIMS	0.0118**	0.0130**	0.0265***	0.0276***	0.0301***	0.0312***
	(0.00553)	(0.00557)	(0.00480)	(0.00486)	(0.00459)	(0.00469)
POP_DENS	1.551***	1.563***	1.408***	1.385***	1.244***	1.206***
	(0.322)	(0.322)	(0.326)	(0.322)	(0.296)	(0.292)
GDP_PC	0.611***	0.552***	0.491***	0.428***	0.359**	0.290**
	(0.166)	(0.166)	(0.164)	(0.160)	(0.149)	(0.146)
HT_S	2.902*	2.936*	2.553*	2.566*	3.066**	3.073**
	(1.548)	(1.585)	(1.511)	(1.511)	(1.470)	(1.459)
R&D_INT	2.606	3.153	2.999	3.570*	3.503*	4.114**
	(2.000)	(1.939)	(1.949)	(1.902)	(1.849)	(1.814)
Constant	1.531***	1.574***	1.497***	1.551***	1.545***	1.605***
	(0.211)	(0.213)	(0.208)	(0.207)	(0.185)	(0.184)
Observations	900,480	900,480	887,040	887,040	873,600	873,600
Period FE	YES	YES	YES	YES	YES	YES
NUTS-2 FE	YES	YES	YES	YES	YES	YES
Tech. Field FE	YES	YES	YES	YES	YES	YES

Note: Robust Standard errors clustered at the NUTS-2 level in parenthesis. P-values are represented by: *** p<0.01, ** p<0.05, * p<0.10. In this specification, the High-Tech employment share and the regional R&D intensity are included as controls. The drop in the number of observations is due to missing values. The dependent variable is the count of patents in NUTS-2 regions by four-digit CPC technological classes, excluding those classified as AI. In Columns 1-2, results of the main specification are shown. In Columns 3-4, the top ten most frequent AI CPC classes are excluded; in Columns 5–6, a stricter criterion is applied, omitting the top twenty AI CPC classes. Independent variables are centred around their mean. All regressors are lagged by one period. All models include Period, NUTS-2 and Technological Field Fixed Effects.

Table 8. Total Patents Specification including RTA and AI knowledge stock interaction; Pseudo Poisson Fixed-Effect model.

	Main Specification	Excl. top 10 AI CPC codes	Excl. top 20 AI CPC codes	Including R&D Intensity	Including High-Tech Employment share
	(1)	(2)	(3)	(4)	(5)
RTA	0.887***	0.884***	0.887***	0.897***	0.903***
	(0.0212)	(0.0216)	(0.0213)	(0.0233)	(0.0252)
K_AI	0.212***	0.210***	0.214***	0.210***	0.213***
	(0.0342)	(0.0345)	(0.0343)	(0.0369)	(0.0345)
RTA * K_AI	-0.0329***	-0.0291**	-0.0268**	-0.0378***	-0.0472***
	(0.0122)	(0.0127)	(0.0124)	(0.0136)	(0.0171)
REL_DENS	0.0198***	0.0202***	0.0201***	0.0200***	0.0206***
	(0.00129)	(0.00118)	(0.00109)	(0.00121)	(0.00111)
TECH_CLAIMS	0.0415***	0.0314***	0.0266***	0.0277***	0.0116**
	(0.00599)	(0.00516)	(0.00450)	(0.00578)	(0.00542)
POP_DENS	1.287***	1.183***	1.099***	1.282***	1.551***
	(0.253)	(0.251)	(0.240)	(0.272)	(0.321)
GDP_PC	0.610***	0.550***	0.482***	0.565***	0.605***
	(0.130)	(0.129)	(0.125)	(0.135)	(0.166)
IND_S	1.882**	1.911***	1.927***	1.572**	
	(0.739)	(0.740)	(0.738)	(0.743)	
R&D_INT				2.618	2.664
				(2.104)	(1.993)
HT_S					2.924*
					(1.551)
Constant	1.240***	1.244***	1.262***	1.375***	1.509***
	(0.198)	(0.194)	(0.187)	(0.200)	(0.213)
Observations	1,473,330	1,451,340	1,429,350	1,293,770	900,480
Year FE	YES	YES	YES	YES	YES
Nuts-2 FE	YES	YES	YES	YES	YES
Tech. Field FE	YES	YES	YES	YES	YES

Note: Robust Standard errors clustered at the NUTS-2 level in parenthesis. P-values are represented by: *** p<0.01, ** p<0.05, * p<0.10. In this specification, an interaction between the RTA indicator and the regional AI stock is included. The drop in the number of observations is due to missing values. The dependent variable is the count of patents in NUTS-2 regions by four-digit CPC technological classes, excluding those classified as AI. In Column 1, results of the main specification are shown. In Columns 2-3, the top ten and top twenty most frequent AI CPC classes are excluded; in Columns 4–5, the regional R&D intensity and the High-Tech employment share are included as controls. Independent variables are centred around their mean. All regressors are lagged by one period. All models include Period, NUTS-2 and Technological Field Fixed Effects.

Table 9. Total Patents Specification excluding ICT patents; Pseudo Poisson Fixed-Effect model.

	Main Specification		Excl. top 10 AI CPC codes		Excl. top 20 AI CPC codes		Including R&D Intensity		Including High-Tech Employment share	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
REL_DENS	0.0198***	0.0227***	0.0198***	0.0228***	0.0197***	0.0232***	0.0198***	0.0230***	0.0205***	0.0239***
	(0.00102)	(0.00102)	(0.00101)	(0.00101)	(0.00102)	(0.00103)	(0.000990)	(0.00105)	(0.000997)	(0.00119)
K_AI	0.127***	0.138***	0.121***	0.132***	0.124***	0.137***	0.125**	0.133***	0.182***	0.182***
	(0.0434)	(0.0430)	(0.0432)	(0.0429)	(0.0431)	(0.0428)	(0.0510)	(0.0510)	(0.0492)	(0.0498)
REL_DENS * K_AI		-0.00130**		-0.00132**		-0.00151**		-0.00147**		-0.00180**
		(0.000626)		(0.000626)		(0.000641)		(0.000702)		(0.000895)
K_ICT	0.0348	0.0239	0.0399	0.0292	0.0382	0.0262	0.0490	0.0330	-0.00728	-0.0210
	(0.0500)	(0.0491)	(0.0496)	(0.0487)	(0.0497)	(0.0487)	(0.0583)	(0.0574)	(0.0706)	(0.0713)
RTA	0.810***	0.812***	0.812***	0.814***	0.827***	0.830***	0.814***	0.816***	0.822***	0.824***
	(0.0183)	(0.0180)	(0.0185)	(0.0183)	(0.0189)	(0.0186)	(0.0188)	(0.0185)	(0.0192)	(0.0188)
TECH_CLAIMS	0.0302***	0.0304***	0.0259***	0.0261***	0.0240***	0.0241***	0.0286***	0.0289***	0.0320***	0.0330***
	(0.00422)	(0.00426)	(0.00420)	(0.00420)	(0.00424)	(0.00422)	(0.00470)	(0.00474)	(0.00474)	(0.00502)
POP_DENS	0.920***	0.953***	0.894***	0.926***	0.891***	0.927***	0.918***	0.949***	1.064***	1.085***
	(0.265)	(0.266)	(0.266)	(0.267)	(0.266)	(0.267)	(0.287)	(0.289)	(0.308)	(0.309)
GDP_PC	0.292*	0.259	0.279*	0.245	0.277*	0.238	0.238	0.202	0.203	0.178
	(0.166)	(0.161)	(0.167)	(0.161)	(0.167)	(0.160)	(0.168)	(0.161)	(0.186)	(0.182)
IND_S	1.563**	1.524**	1.545**	1.502**	1.548**	1.495**	1.450*	1.435*		
	(0.748)	(0.747)	(0.747)	(0.745)	(0.746)	(0.743)	(0.781)	(0.784)		
R&D_INT							0.600	1.274	1.487	2.068
							(2.181)	(2.092)	(1.963)	(1.989)
HT_S									2.717*	2.664*
									(1.587)	(1.589)
Constant	1.564***	1.566***	1.577***	1.579***	1.540***	1.542***	1.654***	1.670***	1.816***	1.837***
	(0.200)	(0.201)	(0.200)	(0.202)	(0.199)	(0.200)	(0.203)	(0.203)	(0.188)	(0.187)
Observations	1,398,564	1,398,564	1,383,171	1,383,171	1,369,977	1,369,977	1,228,116	1,228,116	854,784	854,784
Year FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
NUTS-2 FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Tech. Field FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES

Note: Robust Standard errors clustered at the NUTS-2 level in parenthesis. P-values are represented by: *** p<0.01, ** p<0.05, * p<0.10. The dependent variable is the count of patents in NUTS-2 regions by four-digit CPC technological classes, excluding those classified as AI or ICT. In this specification, in addition to the AI stock, also the ICT technological stock is included as a control. The drop in the number of observations is due to missing values. In Columns 1-2, results of the main specification are shown. In Columns 3-4, the top ten most frequent AI CPC classes are excluded; in Columns 5–6, a stricter criterion is applied, omitting the top twenty AI CPC classes. In Columns 7-8 the R&D intensity is included as a control. In Columns 9-10 the High-Tech employment share is included as a control. Independent variables are centred around their mean. All regressors are lagged by one period. All models include Period, NUTS-2 and Technological Field Fixed Effects.

Table 10. Total Patents Specification by subperiods; Pseudo Poisson Fixed-Effect model.

	Period 1986-2005					Period 2006-2021				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
REL_DENS	0.0195***	0.0195***	0.0239***	0.0228***	0.0235***	0.0215***	0.0215***	0.0338***	0.0287***	0.0270***
	(0.00151)	(0.00153)	(0.00105)	(0.00106)	(0.00138)	(0.00113)	(0.00113)	(0.00267)	(0.00226)	(0.00173)
K_AI		0.177***	0.197***	0.109**	0.0725		0.222***	0.206***	0.260***	0.233***
		(0.0341)	(0.0340)	(0.0442)	(0.0828)		(0.0591)	(0.0625)	(0.0733)	(0.0682)
REL_DENS * K_AI			-0.00249***	-0.00202	-0.00233			-0.00395***	-0.00251***	-0.00251***
			(0.000944)	(0.00126)	(0.00170)			(0.00102)	(0.000936)	(0.000939)
K_ICT				-0.00744	0.165*				-0.0641	-0.0800
				(0.0529)	(0.0996)				(0.0999)	(0.103)
RTA	0.784***	0.784***	0.786***	0.782***	0.785***	0.834***	0.834***	0.835***	0.838***	0.838***
	(0.0170)	(0.0171)	(0.0170)	(0.0178)	(0.0182)	(0.0196)	(0.0196)	(0.0192)	(0.0204)	(0.0205)
TECH_CLAIMS	0.0595***	0.0591***	0.0589***	0.0212***	0.0168***	-0.0236***	-0.0225***	-0.0222***	-0.00867	-0.00851
	(0.00567)	(0.00583)	(0.00573)	(0.00465)	(0.00601)	(0.00798)	(0.00796)	(0.00792)	(0.00770)	(0.00770)
POP_DENS	0.526	0.467	0.666*	0.203	0.274	2.364***	2.045***	2.166***	1.486***	1.330***
	(0.444)	(0.367)	(0.372)	(0.399)	(0.707)	(0.500)	(0.510)	(0.501)	(0.458)	(0.471)
GDP_PC	0.382*	0.323*	0.235	-0.0796	-0.359	0.657***	0.489**	0.453**	-0.0866	-0.0864
	(0.204)	(0.193)	(0.192)	(0.197)	(0.284)	(0.239)	(0.226)	(0.228)	(0.271)	(0.252)
IND_S	2.545***	2.858***	2.585***	2.452***		-1.517	-1.431	-1.336	-0.845	
	(0.854)	(0.733)	(0.730)	(0.791)		(1.510)	(1.522)	(1.489)	(1.400)	
R&D_INT				-1.897	4.240				1.525	1.622
				(4.688)	(9.591)				(3.048)	(2.857)
HT_S					1.040					3.079
					(2.239)					(2.062)
Constant	2.164***	1.941***	1.842***	2.432***	2.485***	1.129***	0.777**	0.773**	1.379***	1.818***
	(0.285)	(0.250)	(0.257)	(0.262)	(0.426)	(0.373)	(0.388)	(0.388)	(0.343)	(0.290)
Observations	859,048	859,048	859,048	656,824	316,500	606,544	606,544	606,544	564,894	535,096
Year FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
NUTS-2 FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Tech. Field FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES

Note: Robust Standard errors clustered at the NUTS-2 level in parenthesis. P-values are represented by: *** p<0.01, ** p<0.05, * p<0.10. The dependent variable is the count of patents in NUTS-2 regions by four-digit CPC technological classes, excluding those classified as AI. In Columns 1-3, the analysis is performed for the subperiod 1986-2005; in Columns 6–8, the focus shifts to the subperiod 2006-2021. In Columns 4-5 and 9-10, the analysis is conducted excluding ICT patents and including the ICT stock as a control for the first and second subperiods, respectively. Independent variables are centred around their mean. All regressors are lagged by one period. All models include Period, NUTS-2 and Technological Field Fixed Effects.

Table 11. High-impact Patents Specification with 5% threshold; Pseudo Poisson Fixed-Effect model.

	Main Specification		Excl. top 10 AI CPC codes		Excl. top 20 AI CPC codes		Including R&D Intensity and H.T. Emp. share	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
REL_DENS	0.0190***	0.0279***	0.0193***	0.0273***	0.0195***	0.0269***	0.0191***	0.0294***
	(0.00127)	(0.00131)	(0.00120)	(0.00130)	(0.00121)	(0.00131)	(0.00130)	(0.00214)
K_AI	0.251***	0.258***	0.266***	0.275***	0.272***	0.280***	0.361***	0.345***
	(0.0516)	(0.0530)	(0.0523)	(0.0533)	(0.0525)	(0.0535)	(0.0827)	(0.0870)
REL_DENS * K_AI		-0.00349***		-0.00314***		-0.00291***		-0.00492***
		(0.000648)		(0.000591)		(0.000563)		(0.00117)
PAT	0.000757***	0.000818***	0.000867***	0.000913***	0.000952***	0.000992***	0.000687***	0.000754***
	(0.000143)	(0.000150)	(0.000177)	(0.000185)	(0.000258)	(0.000267)	(0.000156)	(0.000158)
RTA	0.772***	0.776***	0.772***	0.775***	0.778***	0.781***	0.785***	0.785***
	(0.0197)	(0.0196)	(0.0207)	(0.0206)	(0.0203)	(0.0203)	(0.0192)	(0.0191)
TECH_CLAIMS	0.00892*	0.00898*	-0.00337	-0.00324	-0.00796	-0.00798	0.00794	0.00880
	(0.00522)	(0.00515)	(0.00485)	(0.00473)	(0.00507)	(0.00493)	(0.00676)	(0.00655)
POP_DENS	-0.177	-0.161	-0.0867	-0.0978	0.0823	0.0595	-0.488	-0.521
	(0.340)	(0.346)	(0.337)	(0.341)	(0.340)	(0.341)	(0.495)	(0.483)
GDP_PC	-0.0785	-0.190	-0.0727	-0.180	-0.0527	-0.159	-0.714**	-0.824**
	(0.216)	(0.223)	(0.225)	(0.231)	(0.230)	(0.236)	(0.333)	(0.337)
IND_S	2.082**	1.993**	1.842**	1.748*	1.755*	1.659*		
	(0.900)	(0.910)	(0.903)	(0.908)	(0.915)	(0.917)		
R&D_INT							8.722**	9.869***
							(3.809)	(3.684)
HT_S							2.737	2.872
							(2.374)	(2.360)
Constant	-0.0737	-0.0365	-0.233	-0.186	-0.413	-0.360	0.442	0.529
	(0.249)	(0.258)	(0.249)	(0.256)	(0.259)	(0.265)	(0.335)	(0.327)
Observations	1,467,970	1,467,970	1,446,060	1,446,060	1,424,150	1,424,150	899,810	899,810
Year FE	YES	YES	YES	YES	YES	YES	YES	YES
NUTS-2 FE	YES	YES	YES	YES	YES	YES	YES	YES
Tech. Field FE	YES	YES	YES	YES	YES	YES	YES	YES

Note: Robust Standard errors clustered at the NUTS-2 level in parenthesis. P-values are represented by: *** p<0.01, ** p<0.05, * p<0.10. The dependent variable is the count of breakthrough patents in NUTS-2 regions by four-digit CPC technological classes, excluding those classified as AI. In this specification, patents are defined as breakthrough if they fall within the 95th percentile of the technology–year citation distribution. The drop in the number of observations is due to missing values. In Columns 1-2, results of the main specification are shown. In Columns 3-4, the top ten most frequent AI CPC classes are excluded; in Columns 5–6, a stricter criterion is applied, omitting the top twenty AI CPC classes. In Columns 7-8 the R&D intensity and the High-Tech employment share are included as controls. Independent variables are centred around their mean. All regressors are lagged by one period. All models include Period, NUTS-2 and Technological Field Fixed Effects.

Table 12. High-impact Patents Specification with 1% threshold; Pseudo Poisson Fixed-Effect model.

	Main Specification		Excl. top 10 AI CPC codes		Excl. top 20 AI CPC codes		Including R&D Intensity and H.T. Emp. share	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
REL_DENS	0.0175***	0.0246***	0.0179***	0.0245***	0.0183***	0.0243***	0.0171***	0.0250***
	(0.00125)	(0.00131)	(0.00122)	(0.00132)	(0.00123)	(0.00131)	(0.00128)	(0.00176)
K_AI	0.225***	0.235***	0.233***	0.244***	0.240***	0.251***	0.230**	0.215**
	(0.0454)	(0.0471)	(0.0449)	(0.0463)	(0.0448)	(0.0461)	(0.0913)	(0.0949)
REL_DENS * K_AI		-0.00309***		-0.00291***		-0.00268***		-0.00403***
		(0.000559)		(0.000558)		(0.000577)		(0.000950)
PAT	0.000690***	0.000750***	0.000786***	0.000833***	0.000856***	0.000894***	0.000580***	0.000639***
	(0.000120)	(0.000129)	(0.000156)	(0.000167)	(0.000227)	(0.000237)	(0.000127)	(0.000132)
RTA	0.762***	0.767***	0.763***	0.768***	0.767***	0.773***	0.770***	0.773***
	(0.0220)	(0.0220)	(0.0223)	(0.0222)	(0.0228)	(0.0228)	(0.0228)	(0.0228)
TECH_CLAIMS	0.0193***	0.0189***	0.0125**	0.0123**	0.00960	0.00934	0.0135**	0.0135**
	(0.00616)	(0.00624)	(0.00616)	(0.00620)	(0.00606)	(0.00607)	(0.00655)	(0.00652)
POP_DENS	-0.0277	-0.0190	-0.0572	-0.0745	0.0833	0.0561	-0.107	-0.144
	(0.295)	(0.302)	(0.296)	(0.303)	(0.305)	(0.310)	(0.492)	(0.486)
GDP_PC	0.103	-0.00183	0.103	-0.000313	0.134	0.0340	-0.449	-0.542
	(0.190)	(0.194)	(0.206)	(0.211)	(0.216)	(0.220)	(0.343)	(0.344)
IND_S	1.808**	1.733**	1.671**	1.589*	1.657**	1.576*		
	(0.809)	(0.821)	(0.813)	(0.821)	(0.815)	(0.820)		
R&D_INT							6.444	7.361*
							(4.214)	(4.112)
HT_S							5.579**	5.600**
							(2.357)	(2.369)
Constant	-2.260***	-2.227***	-2.335***	-2.292***	-2.506***	-2.458***	-1.732***	-1.647***
	(0.222)	(0.231)	(0.228)	(0.235)	(0.242)	(0.248)	(0.371)	(0.367)
Observations	1,446,530	1,446,530	1,420,980	1,420,980	1,399,450	1,399,450	893,110	893,110
Year FE	YES	YES	YES	YES	YES	YES	YES	YES
NUTS-2 FE	YES	YES	YES	YES	YES	YES	YES	YES
Tech. Field FE	YES	YES	YES	YES	YES	YES	YES	YES

Note: Robust Standard errors clustered at the NUTS-2 level in parenthesis. P-values are represented by: *** p<0.01, ** p<0.05, * p<0.10. The dependent variable is the count of breakthrough patents in NUTS-2 regions by four-digit CPC technological classes, excluding those classified as AI. In this specification, patents are defined as breakthrough if they fall within the 99th percentile of the technology–year citation distribution. The drop in the number of observations is due to missing values. In Columns 1-2, results of the main specification are shown. In Columns 3-4, the top ten most frequent AI CPC classes are excluded; in Columns 5–6, a stricter criterion is applied, omitting the top twenty AI CPC classes. In Columns 7-8 the R&D intensity and the High-Tech employment share are included as controls. Independent variables are centred around their mean. All regressors are lagged by one period. All models include Period, NUTS-2 and Technological Field Fixed Effects.

on high-quality patents: we proxy invention quality by forward citations, interpreting highly cited patents as high-impact inventions with greater ex post technological impact (Verhoeven et al., 2016). This distinction is particularly relevant because breakthrough inventions contribute disproportionately to future innovation and exert strong local spillovers (Kerr, 2010). Since such inventions typically rely on deep localized knowledge and related technological capabilities (Boschma et al., 2023), AI may be especially important in relaxing these constraints. We therefore replicate the approach of Boschma et al. (2023), identifying high-quality inventions using forward family-to-family citations[24]. We classify a patent as high-quality if it belongs to the top 5 percent (95th percentile) of the forward citation distribution, conditional on its 4-digit CPC technological class and earliest filing year[25]. As a further robustness exercise, we employ a more stringent threshold and classify patents as high-quality only if they fall within the top 1 percent (99th percentile) of the technology–year citation distribution. Tables 11 and 12 report the results when the dependent variable is restricted to the count of high-quality patents, defined as those belonging to the top 5% and top 1% of the technology–year citation distribution, respectively[26]. Overall, the results confirm the baseline findings. The coefficient on relatedness remains positive and significant, indicating that even high-impact inventions rely heavily on local search, consistent with Boschma et al. (2023). Regional AI endowments are also positively associated with the production of high-quality patents, while their interaction with relatedness remains negative and significant. This suggests that AI partly compensates for limited access to related knowledge inputs even in the generation of high-impact inventions. Notably, the AI coefficient and the interaction term are larger than in the baseline estimates, indicating that AI's enabling role may be particularly pronounced for high-impact innovations. Although these findings should be interpreted with caution, they suggest that AI contributes not only to the volume but also to the quality of regional innovative activity, especially in more technologically distant domains.

5. Conclusions

This paper examined how regional technological relatedness and local AI endowments jointly shape innovative activity across technological fields. Using a novel three-way panel (670 four-digit CPC classes × 302 NUTS-2 regions × nine four-year periods, 1986–2021), we identify three main findings. First, consistent with the EEG literature, technologies that are cognitively close to a region's existing knowledge base exhibit higher patenting rates, confirming relatedness as a robust predictor of innovative activity.

---

[24] Family-to-family citations provide a more accurate measure of technological influence by reducing duplication. Rather than counting citation links between individual applications, they aggregate citations at the DOCDB family level, so that multiple applications from the same citing family referring to the same cited family are counted only once.

[25] For each four-digit CPC class–year cell, we construct the distribution of forward family-to-family citations and identify patents located in its upper tail. This within-cell normalization compares patents only with technologically and temporally similar peers, thereby accounting for differences in citation practices across fields and for the longer citation window available to older patents.

[26] In Tables 11 and 12, we additionally control for the total number of patents produced in each region–technology–period cell. Following Boschma et al. (2023), this accounts for scale effects, since greater inventive activity mechanically increases the likelihood of generating high-impact patents.

Second, local AI endowments are positively associated with patenting across technological fields, supporting the view of AI as a general-purpose enabling technology. Third, the negative and significant interaction between relatedness and AI indicates that AI reduces the dependence of invention on local technological proximity, broadening the scope for regional technological recombination. Overall, AI both raises overall inventive activity and relaxes the dependence on local relatedness.

These outcomes, validated by a broad set of robustness checks, further support the relevance of geography (Morgan, 2004). While AI broadens the scope of feasible combinations, its impact depends on local presence: regions with stronger AI endowments capture larger gains. AI does not make place irrelevant: regions that invest in AI infrastructure and technologies appear to be better positioned to both exploit proximate technological opportunities and venture into distant fields.

The paper contributes to the literature in three ways. First, it provides further evidence supporting the relatedness framework at a highly disaggregated region–technology–period level. Second, it conceptually extends existing theory by arguing that local AI endowments have broad stimulatory effects across technological domains within regions, in accordance with the view that characterizes AI as a general-purpose enabler for innovation and as an IMI. Finally, it offers new empirical evidence in support of such a theoretical framework.

In terms of policy implications, our arguments and results clearly point to the need to foster AI innovation and diffusion, especially in "trapped" regions (Diemer et al. 2022), which may be less able to attract AI investment or develop AI capabilities, unless targeted policies help them escape the trap. Policy makers should be aware that industrial and innovation policies supporting AI are not only crucial in themselves, but also because they generate two complementary indirect effects: on the one hand, the acceleration of all the other innovative activities, and, on the other hand, the broadening of the feasible directions for local innovation. To this end, it would be crucial for Europe to develop not only AI-related technological knowledge, but also solid, domain-specific computing infrastructure (Hess & Sieker, 2025; Lemke & Schneider, 2025) [27] and a resilient and sustainable supply chain of AI-enabling raw materials (Li et al., 2024)[28]. Integrating these objectives with AI and digital innovation policies would be essential to prevent new forms of technological lock-in and sustain long-term capacity for innovation.

[27] The EU is developing a two-tier sovereign AI-compute infrastructure. Its distributed layer includes 19 AI Factories and 13 Antennas, supported by AI-optimised supercomputers and offering compute, data, and technical services to SMEs and researchers, with combined European and national investment expected to reach €10 billion over 2021–2027. At the frontier scale, InvestAI aims to mobilise €20 billion for up to five AI Gigafactories equipped with more than 100,000 advanced processors each. However, implementation risks remain: AI Factories may lack the sectoral alignment of private cloud providers and flexibility needed by SMEs, while Gigafactories may face insufficient demand to ensure effective utilisation (Hess & Sieker, 2025; Lemke & Schneider, 2025).

[28] The material foundations of AI and other digital technologies are often overlooked (Li et al., 2024). The production and diffusion of AI technologies depend on a continuous supply of such materials (e.g., lithium, cobalt, tin, tungsten, or tantalum) that are vital for semiconductors, sensors, and advanced computational hardware (Diemer et al., 2025). Ensuring secure access to these inputs is thus not merely a matter of industrial competitiveness, but of technological sovereignty and strategic autonomy (Guarascio et al., 2025). The recent European Critical Raw Materials Act (Regulation (EU) 2024/1252) sets an important precedent, aiming to reduce dependency on a limited number of external suppliers and to strengthen the EU's internal capabilities in extraction, processing, and recycling.

Some caveats and limitations concerning our study should be discussed. First, as already mentioned, patents are an imperfect proxy for innovation: they omit non-patented knowledge, may be used strategically, and do not reflect the overall spectrum of innovations that are successfully commercialized or brought to the marketplace. Second, our measure of regional AI endowments is based on disclosed, patentable AI inventions; this choice may underestimate local AI capabilities that do not appear in patent records. Third, the sample ends in 2021 and thus does not capture the very recent surge of generative and increasingly agentic AI systems, whose scientific and technological impacts may differ in nature and scale from the earlier waves of AI diffusion. While these developments extend beyond the temporal scope of the present analysis, they open up important opportunities for future research to assess how the role of regional AI endowments evolves. Fourth, the focus on European NUTS-2 regions limits immediate generalizability to other institutional and market environments. Finally, the region–technology–period aggregation is powerful for detecting granular patterns but cannot by itself reveal the micro-mechanisms that generate the observed associations (e.g., complementarity effects, different task allocations in R&D departments and scientific labs, lab automation, heterogeneous human–AI workflows, etc.).

Building on these limitations suggests future avenues of research. Firstly, studies based on firm- and laboratory-level datasets (R&D inputs, hiring patterns, platform usage, product launches, experiment logs) can unpack the micro-mechanisms through which AI accelerates local innovation and surfaces distant innovative complementarities. Secondly, directly examining the effects of Generative and Agentic AI on innovation processes would be particularly valuable to assess whether, and under what circumstances, these advances amplify the effects identified in this study. Thirdly, comparative studies and policy experiments will also be essential to understand heterogeneity in outcomes and identify special regional interventions that may amplify the benefits of AI while containing risks, notably the potential widening of gaps between leading and lagging regions (Iammarino et al., 2019; Rodríguez-Pose & You, 2024).

References

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

Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A. J., … Jumper, J. M. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature, 630(8016), 493–500.

Acemoglu, D. (2025). The simple macroeconomics of AI. Economic Policy, 40(121), 13-58.

Acs, Z. J., Anselin, L., & Varga, A. (2002). Patents and innovation counts as measures of regional production of new knowledge. Research Policy, 31(7), 1069-1085.

Agrawal, A., Gans, J., & Goldfarb, A. (2022). Prediction Machines, Updated and Expanded: The Simple Economics of Artificial Intelligence. Harvard Business Press.

Agrawal, A., McHale, J., & Oettl, A. (2019). Finding needles in haystacks: Artificial intelligence and recombinant growth. In Agrawal, A., Gans, J., & Goldfarb, A. (eds.), The economics of artificial intelligence: an agenda. University of Chicago Press.

Agrawal, A., McHale, J., & Oettl, A. (2023). Superhuman science: How artificial intelligence may impact innovation. Journal of Evolutionary Economics, 33(5), 1473-1517.

Agrawal, A., McHale, J., & Oettl, A. (2024). Artificial intelligence and scientific discovery: A model of prioritized search. Research Policy, 53(5), 104989.

Alekseeva, L., Azar, J., Giné, M., Samila, S., & Taska, B. (2021). The demand for AI skills in the labor market. Labour Economics, 71, 102002.

Allison, P. D., & Waterman, R. (2002). Fixed effects negative binomial regression models. In Stolzenberg, R. M. (ed.), Sociological Methodology. Oxford: Basil Blackwell.

Audretsch, D. B., & Feldman, M. P. (1996). R&D spillovers and the geography of innovation and production. American Economic Review, 86(3), 630-640.

Babina, T., Fedyk, A., He, A., & Hodson, J. (2024). Artificial intelligence, firm growth, and product innovation. Journal of Financial Economics, 151, 103745.

Balassa, B. (1965). Trade Liberalization and Revealed Comparative Advantage. The Manchester School of Economic and Social Studies, 33, 99-123.

Balland, P. A., Boschma, R., & Frenken, K. (2015). Proximity and innovation: From statics to dynamics. Regional Studies, 49(6), 907-920.

Balland, P. A., Boschma, R., Crespo, J., & Rigby, D. L. (2019). Smart specialisation policy in the European Union: relatedness, knowledge complexity and regional diversification. Regional Studies, 53(9), 1252-1268.

Balland, P.A. (2017) Economic Geography in R: Introduction to the EconGeo Package, Papers in Evolutionary Economic Geography, 17 (09): 1-75

Bekar, C., Carlaw, K., & Lipsey, R. (2018). General purpose technologies in theory, application and controversy: a review. Journal of Evolutionary Economics, 28(5), 1005-1033.

Bianchini, S., Müller, M., & Pelletier, P. (2022). Artificial intelligence in science: An emerging general method of invention. Research Policy, 51(10), 104604.

Bloom, N., Jones, C. I., Van Reenen, J., & Webb, M. (2020). Are Ideas Getting Harder to Find?. American Economic Review, 110(4), 1104–44.

Borgonovi, F., Calvino, F., Criscuolo, C., Nania, J., Nitschke, J., O'Kane, L., Samek, L., & Seitz, H. (2023). Emerging trends in AI skill demand across 14 OECD countries. OECD Artificial Intelligence Papers, No. 2, OECD Publishing, Paris.

Boschma, R. (2005). Proximity and innovation: a critical assessment. Regional Studies, 39(1), 61-74.

Boschma, R. (2017). Relatedness as driver of regional diversification: A research agenda. Regional Studies, 51(3), 351-364.

Boschma, R. A., & Frenken, K. (2006). Why is economic geography not an evolutionary science? Towards an evolutionary economic geography. Journal of Economic Geography, 6(3), 273-302.

Boschma, R. A., & Lambooy, J. G. (1999). Evolutionary economics and economic geography. Journal of Evolutionary Economics, 9(4), 411-429.

Boschma, R., & Frenken, K. (2011). The emerging empirics of evolutionary economic geography. Journal of Economic Geography, 11(2), 295-307.

Boschma, R., & Gianelle, C. (2014). Regional branching and smart specialisation policy. S3 Policy Brief Series, 6, 2014.

Boschma, R., & Iammarino, S. (2009). Related Variety, Trade Linkages, and Regional Growth in Italy. Economic Geography, 85(3), 289–311.

Boschma, R., Balland, P. A., & Kogler, D. F. (2015). Relatedness and technological change in cities: the rise and fall of technological knowledge in US metropolitan areas from 1981 to 2010. Industrial and Corporate Change, 24(1), 223-250.

Boschma, R., Miguelez, E., Moreno, R., & Ocampo-Corrales, D. B. (2023). The role of relatedness and unrelatedness for the geography of technological breakthroughs in Europe. Economic Geography, 99(2), 117-139.

Boschma, R., Minondo, A., & Navarro, M. (2012). Related variety and regional growth in Spain. Papers in Regional Science, 91(2), 241-257.

Breschi, S. & Malerba, F. (1997). Sectoral Innovation Systems: Technological Regimes, Schumpeterian Dynamics and Spatial Boundaries. In Edquist, C. (ed.), Systems of Innovation. Technologies, Institutions and Organizations, Routledge.

Breschi, S., Lissoni, F., & Malerba, F. (2003). Knowledge-relatedness in firm technological diversification. Research Policy, 32(1), 69-87.

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

Brynjolfsson, E. (2023). The turing trap: The promise & peril of human-like artificial intelligence. In Augmented education in the global age (pp. 103-116). Routledge.

Brynjolfsson, E., & McAfee, A. (2014). The second machine age: Work, progress, and prosperity in a time of brilliant technologies. WW Norton & company.

Buarque, B. S., Davies, R. B., Hynes, R. M., & 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.

Calvino, F., Dernis, H., Samek, L., & Ughi, A. (2024). A sectoral taxonomy of AI intensity. OECD Artificial Intelligence Papers, No. 30, OECD Publishing, Paris.

Calvino, F., Samek, L., Squicciarini, M., & Morris, C. (2022). Identifying and characterising AI adopters: A novel approach based on big data. OECD Science, Technology and Industry Working Papers, No. 2022/06, OECD Publishing, Paris.

Cameron, A. C., & Trivedi, P. K. (2013). Regression analysis of count data (No. 53). Cambridge university press.

Caragliu, A., & Nijkamp, P. (2016). Space and knowledge spillovers in European regions: the impact of different forms of proximity on spatial knowledge diffusion. Journal of Economic Geography, 16(3), 749-774.

Carbonara, E., & Santarelli, E. (2023). Artificial intelligence and robots: A threat or an opportunity for SMEs and entrepreneurship?. In Carbonara, E., & Tagliaventi, M. (eds.). SMEs in the digital era. Cheltenham: Edward Elgar, pp. 104-121.

Castaldi, C., Frenken, K., & Los, B. (2015). Related Variety, Unrelated Variety and Technological Breakthroughs: An analysis of US State-Level Patenting. Regional Studies, 49(5), 767-781.

Catacutan, D. B., Alexander, J., Arnold, A., & Stokes, J. M. (2024). Machine learning in preclinical drug discovery. Nature Chemical Biology, 20(8), 960–973.

Caviggioli, F., Colombelli, A., De Marco, A., Scellato, G., & Ughetto, E. (2023). Co-evolution patterns of university patenting and technological specialization in European regions. The Journal of Technology Transfer, 48(1), 216-239.

Cicerone, G., Faggian, A., Montresor, S., & Rentocchini, F. (2023). Regional artificial intelligence and the geography of environmental technologies: does local AI knowledge help regional green-tech specialization?. Regional Studies, 57(2), 330-343.

Cockburn, I. M., Henderson, R., & Stern, S. (2019). The impact of artificial intelligence on innovation. In Agrawal, A., Gans, J., & Goldfarb, A. (eds.), The economics of artificial intelligence: an agenda. University of Chicago Press.

Cohen, W. M., & Levinthal, D. A. (1990): Absorptive capacity: A new perspective on learning and innovation. Administrative Science Quarterly, 35(1), 128–152, 2393553.

Colombelli, A., & Quatraro, F. (2018). New firm formation and regional knowledge production modes: Italian evidence. Research Policy, 47, 139–157.

Colombelli, A., D'Amico, E., & Paolucci, E. (2023). When computer science is not enough: universities knowledge specializations behind artificial intelligence startups in Italy. The Journal of Technology Transfer, 48(5), 1599-1627.

Colombelli, A., Krafft, J., & Quatraro, F. (2014). The emergence of new technology-based sectors in European regions: A proximity-based analysis of nanotechnology. Research Policy, 43(10), 1681-1696.

Correia, S., Guimarães, P., & Zylkin, T. (2020). Fast Poisson estimation with high-dimensional fixed effects. The Stata Journal, 20(1), 95-115.

D'Alessandro, F., Santarelli, E., & Vivarelli, M. (2026). The KSTE+I approach and the advent of AI technologies: Evidence from the European regions. The Journal of Technology Transfer, 1-38.

Dahlke, J., Beck, M., Kinne, J., Lenz, D., Dehghan, R., Wörter, M., & Ebersberger, B. (2024). Epidemic effects in the diffusion of emerging digital technologies: evidence from artificial intelligence adoption. Research Policy, 53(2), 104917.

Dahlke, J., Schmidt, S., Lenz, D., Kinne, J., Dehghan, R., Abbasiharofteh, M., ... & Rammer, C. (2025). The WebAI Paradigm of Innovation Research: Extracting Insight From Organizational Web Data Through AI.

Damioli, G., Van Roy, V., & Vertesy, D. (2021). The impact of artificial intelligence on labor productivity. Eurasian Business Review, 11(1), 1-25.

Damioli, G., Van Roy, V., Vértesy, D., & Vivarelli, M. (2024). Drivers of employment dynamics of AI innovators. Technological Forecasting and Social Change, 201, 123249.

Damioli, G., Van Roy, V., Vértesy, D., & Vivarelli, M. (2025). Is artificial intelligence leading to a new technological paradigm?. Structural Change and Economic Dynamics, 72, 347-359.

David, P. A. (1985). Clio and the Economics of QWERTY. The American Economic Review, 75(2), 332-337.

David, P. A. (1990). The dynamo and the computer: An historical perspective on the modern productivity paradox. The American Economic Review, 80(2), 355–361.

de Rassenfosse, G., Kozak, J., & Seliger, F. (2019). Geocoding of worldwide patent data. Scientific Data, 6(1), 260.

Dernis, H., Calvino, F., Moussiegt, L., Nawa, D., Samek, L., & Squicciarini, M. (2023). Identifying artificial intelligence actors using online data. OECD Science, Technology and Industry Working Papers, No. 2023/01, OECD Publishing, Paris.

Dernis, H., Moussiegt, L., Nawa, D., & Squicciarini, M. (2021). Who develops AI-related innovations, goods and services? A firm-level analysis. OECD Science, Technology and Industry Policy Papers, No. 121, OECD Publishing, Paris.

Diemer, A., Iammarino, S., Perkins, R., & Gros, A. (2025). Technology, resources and geography in a paradigm shift: the case of critical and conflict materials in ICTs. Regional Studies, 59(1), 2077326.

Diemer, A., Iammarino, S., Rodríguez-Pose, A., & Storper, M. (2022). The regional development trap in Europe. Economic Geography, 98(5), 487-509.

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., & Nelson, R. R. (2023). Innovation as an Evolutionary Process. In Dosi, G., The foundations of complex evolving economies: Part one: Innovation, organization, and industrial dynamics, Oxford University Press.

Dunham, J., Melot, J., & Murdick, D. (2020). Identifying the development and application of artificial intelligence in scientific text. arXiv preprint arXiv:2002.07143.

EPO (2020). Patents and the Fourth Industrial Revolution. The global technology trends enabling the data-driven economy, Munich.

Felice, G., Lamperti, F., & Piscitello, L. (2022). The employment implications of additive manufacturing. Industry and Innovation, 29(3), 333-366.

Fleming, L. (2001). Recombinant uncertainty in technological search. Management Science, 47, 117–132.

Fleming, L. & Sorenson, O. (2004). Science as a map in technological search. Strategic Management Journal 25, 909–928.

Frenken, K., Van Oort, F., & Verburg, T. (2007). Related variety, unrelated variety and regional economic growth. Regional Studies, 41(5), 685-697.

Gabbard, H., Messenger, C., Heng, I.S., & Murray-Smith, R. (2022). Bayesian parameter estimation using conditional variational autoencoders for gravitational-wave astronomy. Nature Physics, 18, 112–117.

Giczy, A. V., Pairolero, N. A., & Toole, A. A. (2022). Identifying artificial intelligence (AI) invention: A novel AI patent dataset. The Journal of Technology Transfer, 47(2), 476-505.

Goldfarb, A., Taska, B., & Teodoridis, F. (2023). Could machine learning be a general purpose technology? A comparison of emerging technologies using data from online job postings. Research Policy, 52(1), 104653.

Grashof, N., & Kopka, A. (2023). Artificial intelligence and radical innovation: an opportunity for all companies?. Small Business Economics, 61(2), 771-797.

Griliches, Z. (1957). Hybrid Corn: An Exploration in the Economics of Technological Change. Econometrica 25 (4), 501–522.

Griliches, Z. (1979). Issues in assessing the contribution of R&D to productivity growth. Bell Journal of Economics 10: 92-116.

Griliches, Z. (1998). Patent statistics as economic indicators: a survey. In R&D and productivity: the econometric evidence (pp. 287-343). University of Chicago Press.

Guarascio, D., Holzner, M., Iacobucci, D., & Meliciani, V. (2025). European competitiveness in the new global context: structural constraints, strategic dependencies and the role of the new industrial policy. Journal of Industrial and Business Economics, 1-9.

Guimaraes, P. (2008). The fixed effects negative binomial model revisited. Economics Letters, 99(1), 63-66.

Hall, B. H., Jaffe, A., & Trajtenberg, M. (2005). Market Value and Patent Citations. The RAND Journal of Economics, 36(1), 16–38.

Hess, J. C., & Sieker, F. (2025). Built for Purpose? Demand-Led Scenarios for Europe's AI Gigafactories. Policy Brief, Interface.

Hidalgo, C. A., & Hausmann, R. (2009). The building blocks of economic complexity. Proceedings of the National Academy of Sciences, 106(26), 10570-10575.

Hidalgo, C. A., Balland, P. A., Boschma, R., Delgado, M., Feldman, M., Frenken, K., ... & Zhu, S. (2018). The principle of relatedness. In International conference on complex systems (pp. 451-457). Cham: Springer International Publishing.

Hidalgo, C. A., Klinger, B., Barabási, A. L., & Hausmann, R. (2007). The product space conditions the development of nations. Science, 317(5837), 482-487.

High-level Expert Group on Artificial Intelligence (2019). Ethics guidelines for trustworthy AI. European Commission.

Iammarino, S. (2005). An evolutionary integrated view of regional systems of innovation: concepts, measures and historical perspectives. European Planning Studies, 13(4), 497-519.

Iammarino, S., Rodríguez-Pose, A., & Storper, M. (2019). Regional inequality in Europe: evidence, theory and policy implications. Journal of Economic Geography, 19(2), 273-298.

Inaba, T., & Squicciarini, M. (2017). ICT: A new taxonomy based on the international patent classification. OECD.

Jacobs, J. (1969). The economy of cities. New York: Random House.

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.

Johnson, P. C., Laurell, C., Ots, M., Sandström, C. (2022). Digital innovation and the effects of artificial intelligence on firms' research and development–automation or augmentation, exploration or exploitation? Technological Forecasting and Social Change, 179, 121636.

Kerr, W. R. (2010). Breakthrough inventions and migrating clusters of innovation. Journal of Urban Economics 67 (1): 46–60.

King, R. D., Rowland, J., Oliver, S. G., Young, M., Aubrey, W., Byrne, E., Liakata, M., Markham, M., Pir, P., Soldatova, L. N., Sparkes, A., Whelan, K. E., & Clare, A. (2009). The automation of science. Science, 324(5923), 85–89.

Kogler, D. F., Rigby, D. L., & Tucker, I. (2013). Mapping Knowledge Space and Technological Relatedness in US Cities. European Planning Studies, 21(9), 1374-1391.

Lábaj, M., Oleš, T. & Procházka, G. (2025). Impact of robots and artificial intelligence on labor and skill demand: evidence from the UK. Eurasian Business Review.

Lazzeretti, L., Innocenti, N., Nannelli, M., & Oliva, S. (2023). The emergence of artificial intelligence in the regional sciences: a literature review. European Planning Studies, 31(7), 1304-1324.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature 521, 436–444.

Lee, J., & Lee, K. (2021). Is the fourth industrial revolution a continuation of the third industrial revolution or something new under the sun? Analyzing technological regimes using US patent data. Industrial and Corporate Change, 30(1), 137-159.

Lemke, N., & Schneider, C. (2025). The European Union's AI Factories. Lessons for Public Investment in AI Infrastructure in Europe. Policy Brief, Interface.

Li, G. Y., Ascani, A., & Iammarino, S. (2024). The material basis of modern technologies. A case study on rare metals. Research Policy, 53(1), 104914.

Liu, J., Chang, H., Forrest, J. Y. L., & Yang, B. (2020). Influence of artificial intelligence on technological innovation: Evidence from the panel data of china's manufacturing sectors. Technological Forecasting and Social Change, 158, 120142.

Maraut, S., Dernis, H., Webb, C., Spiezia, V., & Guellec, D. (2008). The OECD REGPAT database: a presentation. OECD.

Martínez, C. (2010). Patent families: When do different definitions really matter?. Scientometrics, 86(1), 39-63.

Mewes, L. (2019). Scaling of atypical knowledge combinations in American metropolitan areas from 1836 to 2010. Economic Geography, 95(4), 341-361.

Mezzanotti, F., & Simcoe, T. (2023). Innovation and appropriability: Revisiting the role of intellectual property (No. w31428). National Bureau of Economic Research.

Miguelez, E., & Moreno, R. (2018). Relatedness, external linkages and regional innovation in Europe. Regional Studies, 52(5), 688-701.

Minniti, A., Prettner, K., & Venturini, F. (2025). AI innovation and the labor share in European regions. European Economic Review, 105043.

Miric, M., Jia, N., & 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.

Mokyr, J. (1990). The Lever of Riches, Oxford University Press.

Montresor, S., & Quatraro, F. (2017). Regional branching and key enabling technologies: Evidence from European patent data. Economic Geography, 93(4), 367-396.

Morgan, K. (2004). The exaggerated death of geography: learning, proximity and territorial innovation systems. Journal of Economic Geography, 4(1), 3-21.

Neffke, F., Hartog, M., Boschma, R., & Henning, M. (2018). Agents of structural change: The role of firms and entrepreneurs in regional diversification. Economic Geography, 94(1), 23-48.

Nelson, R. R., & Winter, S. G. (1985). An evolutionary theory of economic change. Harvard University Press.

OECD (2019). Artificial Intelligence in Society. OECD Publishing, Paris.

OECD (2023), Artificial Intelligence in Science: Challenges, Opportunities and the Future of Research, OECD Publishing, Paris.

Pairolero, N. A., Giczy, A. V., Torres, G., Islam Erana, T., Finlayson, M. A., & Toole, A. A. (2025). The artificial intelligence patent dataset (AIPD) 2023 update. The Journal of Technology Transfer, 1-24.

Pavitt, K. (1985). Patent statistics as indicators of innovative activities: Possibilities and problems. Scientometrics 7, 77–99.

Purificato, E., Bili, D., Jungnickel, R., Ruiz Serra, V., Fabiani, J., Abendroth Dias, K., Fernandez Llorca, D., & Gomez, E. (2025). The Role of Artificial Intelligence in Scientific Research - A Science for Policy, European Perspective, Publications Office of the European Union, Luxembourg

Qiao, Y., & Li, Y. (2025). Unpacking the role of relatedness in technological diversification in the US metropolitan statistical areas. Industry and Innovation, 1-23.

Rodríguez-Pose, A., & You, Z. (2024). Bridging the innovation gap. AI and robotics as drivers of China's urban innovation (No. 2406). Utrecht University, Department of Human Geography and Spatial Planning, Group Economic Geography.

Rosenberg, N. (1998). Chemical engineering as a general purpose technology. In Helpman, E. (Ed.), General Purpose Technologies and Economic Growth. MIT Press, Cambridge.

Russell, S. J., & Norvig, P. (2016). Artificial intelligence: a modern approach. Pearson.

Santarelli, E., Staccioli, J., & Vivarelli, M. (2023). Automation and related technologies: a mapping of the new knowledge base. Journal of Technology Transfer, 48(2), 779-813.

Schumpeter, J. A., (1939). Business cycles: A theoretical, historical and statistical analysis of the capitalist process. McGraw-Hill, New York.

Schumpeter, J. A., [1911] (1934). The Theory of Economic Development. Harvard University Press, Cambridge.

Sejnowski, T. J. (2018). The deep learning revolution. MIT press.

Solow, R. M. (1987). We'd better watch out. New York Times Book Review, 36.

Squicciarini, M. & Nachtigall, H. (2021). Demand for AI skills in jobs: Evidence from online job postings. OECD Science, Technology and Industry Working Papers, No. 2021/03, OECD Publishing, Paris.

Squicciarini, M., Dernis, H., & Criscuolo, C. (2013). Measuring patent quality: Indicators of technological and economic value. OECD Publishing.

Srebrovic, R., & Yonamine, J. (2020). Leveraging the BERT algorithm for Patents with TensorFlow and BigQuery. White paper.

Staccioli, J. & Virgillito, M.E. (2025). Will your boss be an algorithm? A patent-based analysis of artificial intelligence worker management technologies and labour exposure. Eurasian Business Review.

Tanner, A. N. (2014). Regional branching reconsidered: Emergence of the fuel cell industry in European regions. Economic Geography, 90(4), 403-427.

Tavassoli, S., & Carbonara, N. (2014). The role of knowledge variety and intensity for regional innovation. Small Business Economics, 43(2), 493-509.

Trajtenberg, M. (2019). Artificial intelligence as the next gpt. In Agrawal, A., Gans, J., & Goldfarb, A. (eds.), The economics of artificial intelligence: an agenda. University of Chicago Press.

Truong, Y., & Papagiannidis, S. (2022). Artificial intelligence as an enabler for innovation: A review and future research agenda. Technological Forecasting and Social Change, 183, 121852.

Turkina, E., Van Assche, A., & Doloreux, D. (2021). How do firms in co-located clusters interact? Evidence from Greater Montreal. Journal of Economic Geography, 21(5), 761-782.

van Eck, N. J., & Waltman, L. (2009). How to Normalize Cooccurrence Data? An Analysis of Some Well-Known Similarity Measures. Journal of the American Society for Information Science and Technology 60(8), 1635-1651.

Van Roy, V., Vertesy, D., Damioli, G. (2020). AI and Robotics Innovation. In Zimmermann, K. (ed.), Handbook of Labor, Human Resources and Population Economics. Springer.

Verhoeven, D., Bakker, J., & Veugelers, R. (2016). Measuring technological novelty with patent-based indicators. Research Policy, 45(3), 707–723.

Vivarelli, M. (1995). The economics of technology and employment: Theory and empirical evidence. In The Economics of Technology and Employment. Edward Elgar Publishing.

Vivarelli, M., & Arenas Díaz, G. (2025). Innovation and employment in the era of artificial intelligence. IZA World of Labor

Wang, H., Fu, T., Du, Y., Gao, W., Huang, K., Liu, Z., Chandak, P., Liu, S., Van Katwyk, P., Deac, A., Anandkumar, A., … & Zitnik, M. (2023). Scientific discovery in the age of artificial intelligence. Nature, 620(7972), 47–60.

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

Whittle, A., & Kogler, D. F. (2020). Related to what? Reviewing the literature on technological relatedness: Where we are now and where can we go?. Papers in Regional Science, 99(1), 97-114.

Williams, K., Bilsland, E., Sparkes, A., Aubrey, W., Young, M., Soldatova, L. N., De Grave, K., Ramon, J., de Clare, M., Sirawaraporn, W., Oliver, S. G., & King, R. D. (2015). Cheaper faster drug development validated by the repositioning of drugs against neglected tropical diseases. Journal of the Royal Society, Interface, 12(104), 20141289.

Wooldridge, J. M. (1999). Distribution-free estimation of some nonlinear panel data models. Journal of Econometrics, 90(1), 77-97.

Wooldridge, J. M. (2010). Econometric analysis of cross section and panel data. MIT press.

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

Xu, L., Fu, Z., Xu, Y., & Zhou, Z. (2025). Rivaling reinforcement and resource reallocation: How do industrial robots enhance company's innovation?. Humanities and Social Sciences Communications, 12(1), 1-15.

Online Appendix A

Identifying AI patents through the Automated Patent Landscaping and Deep Learning

A1. Background literature

The accurate delineation of AI-related inventions in patents remains a formidable challenge, owing to both the multifaceted technical disclosures characterizing patent filings and the dynamic nature of the AI realm, which in turn defies simple taxonomy and often renders nearly impossible the development of classification schemes precisely tracing AI developments across space and time. Furthermore, definitional ambiguities cause an additional hurdle (Baruffaldi et al., 2020): indeed, the boundaries of such a technological field remain rather elusive (Van Roy et al., 2020).

Over the past decade, scholars have therefore pursued a spectrum of methodologies to detect AI advances within patents, leveraging their metadata and full texts. Early efforts typically relied on rule-based heuristics (such as curated keyword searches or filtering by International Patent Classification (IPC) or Cooperative Patent Classification (CPC) codes) to flag potentially relevant documents. Examples in this field are provided by Cockburn et al. (2019), who proposed a procedure that combines USPC classes 901 and 706, with a patent title keyword search to identify AI technologies. Similarly, the World Intellectual Property Organization (WIPO, 2019) proposes a three-fold heuristic framework for delineating AI-related inventions within patent records by combining patent-classification filters with targeted keyword searches. Similar to its predecessors, Baruffaldi et al. (2020) developed a methodology that combines both patent classification codes and keywords. This work distinguishes itself by proposing a meticulously curated lexicon of 193 AI-specific terms, extracted via comprehensive text analysis of leading AI journals and conference proceedings. Van Roy et al. (2020) further contribute to this stream by proposing a curated taxonomy of AI-related keywords that covers a broad range of technologies associated with the AI macro-field.

While robust classification schemes relying only on patent classification codes and keywords can capture the state of knowledge at a fixed point in time, static measurement frameworks will inevitably lag behind and overlook nascent subdomains unless they are regularly recalibrated. In addition, classification codes may fail to fully represent the underlying technological content of patents, leading to a higher risk of Type II errors (i.e., false negatives). Finally, the overreliance on keyword-based searches introduces a different challenge, namely, an increased risk of Type I errors (i.e., false positives), particularly when simple keyword searches are performed, given their lack of contextual or semantic sensitivity. As emphasized by Abood and Feltenberger (2018), modern machine learning approaches offer a promising alternative by addressing many of these limitations. Specifically, such models: (a) learn latent patterns and semantic regularities within a curated seed set of AI-related patents; (b) improve the ability to distinguish these from unrelated (anti-seed) patents; (c) generalize effectively to previously unseen patent documents during inference; and (d) substantially reduce the burden on human experts, as the model autonomously captures complex domain-specific signals without requiring exhaustive manual labelling or deep field-specific knowledge. Consequently, machine learning, and particularly deep learning, provides a scalable and context-aware framework for AI patent identification.

Following this line of reasoning, more recent studies have also proposed machine-learning methods to detect the AI content of patents at scale. Leveraging modern machine learning methodologies and the automated patent landscaping procedure (Abood and Feltenberger, 2018), Giczy et al. (2022) employed long short-term memory (LSTM) neural networks to automatically classify patents in eight AI-related fields, using USPTO patent data. Miric et al. (2023) developed a supervised ML pipeline to detect AI-related patents employing any form of "statistical learning" methodology in the broader USPTO corpus. Mann and Püttmann (2023), focusing on the broad domain of "automation," train a Bernoulli Naïve Bayes model on a hand-labelled sample of 560 patent documents, which was then applied to more than five million USPTO-granted patents.

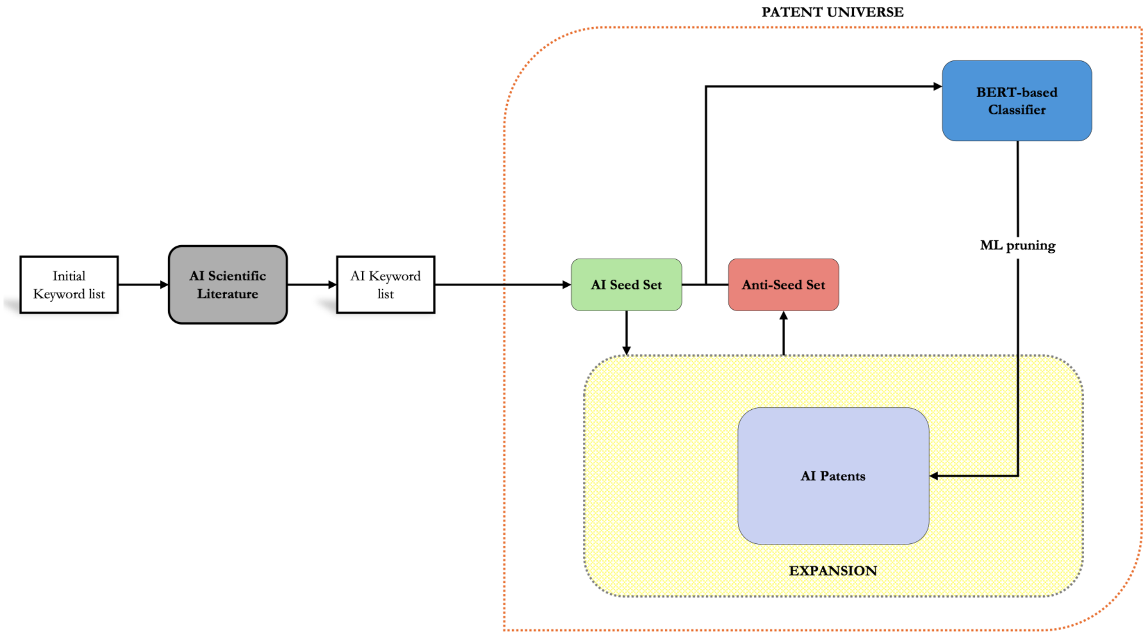
A2. The proposed methodology

Building on recent advances in automated patent landscaping and deep-learning-based text classification, we propose a methodology to identify AI-related patents at scale. Extending the framework developed by Van Roy et al. (2020), we first construct an expanded AI keyword list through a hybrid top-down and bottom-up strategy (Colombelli et al., 2023). This list is then used to retrieve high-precision AI-related seed patents from our Patent Universe[29]. Following Giczy et al. (2022) and Pairolero et al. (2025), we subsequently apply the automated patent landscaping procedure proposed by Abood and Feltenberger (2018) to expand the initial seed set and to construct an anti-seed sample. The resulting training corpus is used to fine-tune a BERT-based classifier relying on the pre-trained BERT for Patents architecture (Srebrovic & Yonamine, 2020)[30]. The classifier is then applied to the expanded patent set to prune documents erroneously retained during the landscaping phase, thereby preserving only patents with substantive AI-related content. Figure A1 summarizes the overall procedure.

[29] The Patent Universe comprises all patent families with at least one application filed at either the European Patent Office, World Intellectual Property Organization, or United States Patent and Trademark Office, and published between 1980 and 2021. The dataset was constructed using the PATSTAT database and includes approximately 27 million patent applications, corresponding to nearly 14 million DOCDB patent families. Given data availability constraints and the objective of ensuring cross-office comparability, our machine learning–based analysis relies exclusively on the information contained in patent titles and abstracts. The patent universe is used to (a) retrieve AI patents forming the seed set for training, (b) perform the patent landscaping procedure, and (c) conduct the inference on patents included in the expansion.

[30] This model is an extension of the BERT architecture (Devlin et al., 2018) trained on over 100 million patent publications from the U.S. and other countries. The model released by the authors allows immediate reuse of the resulting encoder representations for downstream tasks, such as multiple or binary classifications, without the need to retrain the core Transformer architecture.

Figure A1. Methodological diagram



We begin by leveraging a simplified version of the keyword list proposed by Van Roy et al. (2020) and Damioli et al. (2024), which serves as the initial query input in Elsevier's Scopus database[31]. Using this list, we retrieved scientific publications that contain at least one of the predefined n-grams in their abstract, title, or authors' keywords. To minimize noise and ensure thematic focus, the search was restricted to core disciplinary areas that are foundational to AI research: namely, Computer Science, Engineering, and Mathematics (Bianchini et al., 2022). The retrieved publications were then analysed through text-mining techniques to identify additional n-grams that are coherent with the scientific and technological evolution of the AI field. To improve interpretability and avoid excessive ambiguity, we discarded generic terms[32]. This procedure yields a final extended keyword list of 221 unique n-grams, displayed in Table A1. Within this list, n-grams marked with an asterisk are classified as high-specificity AI terms. These terms identify concepts that are central to the AI domain and unlikely to appear in unrelated technological contexts. They therefore serve as the basis for constructing the initial high-precision AI seed set in the patent corpus. Broader and more general terms are retained in the extended vocabulary and are used in conjunction with high-specificity AI terms to define high-confidence AI seed patents. This distinction ensures precision during the overall procedure: highly specific terms provide reliable positive examples for model training, while the broader vocabulary helps capture the evolving and heterogeneous nature of AI-related technological knowledge.

In particular, we construct the initial AI seed set by mining patent titles and abstracts in the Patent Universe using our extended list of keywords. Consistent with dictionary-based text-classification

[31] This initial list, which serves as a starting input for the top-down phase, consists of only 35 keywords.
[32] In particular, most of the unigrams were not retained and acronyms were discarded.

Table A1. Final list of keywords

action recognition	active learning	active vision system
activity recognition	adaptive boosting*	adaptive control
adaptive learning	adversarial attack	adversarial learning*
adversarial training*	apriori algorithm	artificial immune system
artificial intelligence*	association rule learning*	association rule mining
attention mechanism*	autoencoder*	automated optical inspection
automatic classification	automatic control	automatic detection
automatic generation	automatic identification	automatic recognition
automatic segmentation	automatic target detection	automatic target recognition
autonomous aerial vehicle*	autonomous car*	autonomous driving*
autonomous ground vehicle*	autonomous system	autonomous underwater vehicle*
autonomous vehicle*	backpropagation*	base classifier
base learner	batch normalization*	bayes classifier
bayesian learning	bayesian model	bayesian network
beam search	bidirectional associative memory*	bidirectional encoder representations from transformers*
bootstrap aggregation*	brain computer interface	capsule network*
central pattern generator	cerebellar model arithmetic computer*	character recognition
chatbot	closed loop control system	cluster analysis
collision avoidance	collision detection	competitive learning*
computational intelligence	computational neuroscience	computer vision*
conditional random field	connectionist temporal classification*	continual learning
contrastive learning*	convolutional layer*	convolutional network*
data mining	decision tree	deep belief network*
deep learning*	deep q network*	dilated convolution*
edge detection	emotion detection	emotion recognition
end effector	ensemble classifier	ensemble learning*
entity recognition	evolutionary algorithm*	evolutionary computation*
evolutionary programming*	expert system*	expression recognition
extended kalman filter	extreme learning machine*	face recognition
fall detection	fault classification	fault detection
feature engineering*	feature extraction	feature learning*
feature pyramid network*	feature representation	federated learning*
feedforward network*	forward kinematic*	fuzzy inference system
fuzzy logic	fuzzy system	gated recurrent unit*
gaussian mixture model	generative adversarial network*	generative model
generative pretrained transformer*	genetic algorithm*	genetic network programming*
genetic programming*	gesture recognition	gradient boosting*
gradient descent*	graph attention network*	haar cascade*
haar classifier*	hidden markov model	histogram of oriented gradient
image captioning	image classification	image recognition
image segmentation	imitation learning*	inductive logic programming*
intelligent agent	intelligent vehicle	inverse kinematic*
k means	k nearest neighbor	knowledge based system
knowledge distillation*	knowledge graph	knowledge process automation
knowledge representation	large language model*	latent dirichlet allocation*

learning vector quantisation*	long short term memory*	machine intelligence
machine learning*	machine vision	manipulator
markov random field	meta learning*	model quantisation*
motion planning	multi label classification	multi layer perceptron*
multi task learning*	multiagent system	natural language generation*
natural language processing*	neural architecture search*	neural classifier*
neural controller*	neural machine translation*	neural network*
object classification	object detection	object recognition
obstacle avoidance	opinion mining*	pattern classification
pattern recognition	pedestrian detection	pid controller
predictive model	probabilistic model	proximal policy optimization*
q learning*	quadcopter*	quadrotor*
radial basis function network*	random forest	rapidly exploring random tree
regression tree	reinforcement learning*	representation learning*
restricted boltzmann machine*	robot*	self driving*
self organising map*	semantic web	sensor data fusion
sensor fusion	sentiment analysis*	sentiment classification*
shapley additive explanation	simultaneous localization mapping	soft computing
speaker recognition	speech recognition	statistical learning*
supervised learning*	support vector machine	support vector regression
swarm intelligence*	swin transformer*	text classification
text mining	topic model*	trajectory planning
trajectory prediction	transfer learning*	transformer model*
unmanned aerial system*	unmanned aerial vehicle*	unmanned aircraft system*
unmanned aircraft vehicle*	unmanned ground vehicle*	unmanned surface vehicle*
unmanned underwater vehicle*	unmanned vehicle*	unsupervised domain adaptation*
unsupervised learning*	vehicle detection	vehicular ad hoc network
vision transformer*	visual servoing*	voice recognition
web mining	word embedding*	

strategies that start from established definitions and keyword lists (Lamperti, 2024), high-specificity AI n-grams provide the starting signal for identifying candidate patents. Accordingly, a necessary condition for inclusion in the seed set is the presence of at least one high-specificity AI n-gram in either the title or the abstract. This ensures the most central and unambiguous AI concepts anchor the dataset, as they serve as an initial marker of AI relevance. However, these terms are not used mechanically. Once this first condition is satisfied, candidate documents are further screened using the full extended vocabulary of 221 AI-related n-grams, including both high-specificity and broader terms. During this additional screening, we consider the number of distinct AI-related n-grams matched in the document, their overall frequency of occurrence, and their co-occurrence patterns within the title–abstract text. The rationale is that patents containing multiple AI-related expressions, or repeated and coherent references to AI concepts, are more likely to disclose genuine AI technological content than patents containing only isolated or incidental mentions. This procedure allows us to privilege precision and semantic relevance in the construction of the seed set, while preserving sufficient heterogeneity to cover the broad technological scope of AI. It also provides an additional consistency check, allowing us to identify and discard ambiguous or likely false-positive documents before constructing the final seed set. The seed set ultimately comprises 17,902 unique, non-duplicate patent filings, corresponding to 17,468 DOCDB families[33].

The seed set described above serves as the foundation for our AI landscape. Following Abood & Feltenberger (2018), the Automated Patent Landscaping can be understood as the process of identifying patents related to a focal technological domain by combining seed documents, patent metadata, and machine-learning classification. In our setting, the initial seed set is expanded using patent-level information on CPC classifications, family citations, and AI mentions to identify a broader set of candidate patents that are likely to be technologically related to AI. These candidate patents are subsequently pruned by a machine-learning model trained to distinguish seed patents, used as positive examples, from anti-seed patents, used as negative examples. We implement a two-level expansion procedure. The Level 1 Expansion, encompasses (1) all patents retrieved through the extended AI keyword list but excluded from the seed set because they contain only isolated or general AI-related mentions; (2) all patents belonging to a DOCDB family that either cites or is cited by patent families included in the seed set (i.e., forward and backward family citations); (3) all patents pertaining to DOCDB families assigned to relevant CPC codes[34]. The Level 2 Expansion further includes all backward and forward family citations of patent families included in the Level 1 expansion. Overall, the Level 1 expansion retrieves 765,704 patent documents, while the Level 2 expansion comprises 5,396,694 patent documents. Since patents in the two expansion levels differ in their semantic proximity to the original seed set, the machine-learning model is expected to retain a higher proportion of Level 1 patents and to prune a larger share of Level 2 candidates (Abood and Feltenberger, 2018). Patents not captured by the two-level expansion process are used as the pool from which anti-seed documents are drawn. From this pool, we randomly sample 35,000 unique, non-duplicate patent documents to construct the negative training sample. This strategy provides a large set of reasonably reliable negative examples that are expected to be semantically distant from the core AI landscape[35]. Combining seed and anti-seed documents yields a final training corpus of 52,902 examples.

---

[33] In most cases, we retain only one patent per DOCDB family in the seed set. Patent documents within the same family often share highly similar, if not identical, titles and abstracts, which increases the risk of redundancy and data leakage across data splits during the training and testing phases. As a result, we preserve a single representative application per family, except in rare instances where two or more documents within the same DOCDB family exhibit sufficiently distinct semantic structures to warrant separate inclusion.

[34] The relevance of CPC codes is evaluated within the seed set relative to the overall Patent Universe. A CPC code is considered relevant if: (i) it appears in at least 0.5% of the patent families within the corresponding seed set (as in Choi et al., 2022); and (ii) the share of patent families containing the CPC code within the seed set relative to its share in the overall patent universe is greater than 50. This criterion ensures that only technology classes disproportionately represented in the seed sets are included in the expansion.

[35] By design, the positive (seed) and negative (anti-seed) examples are intentionally drawn from opposite ends of the relevance spectrum. As such, patents with intermediate similarity (i.e., those falling in the "gray area" between clearly relevant and irrelevant) are underrepresented. While this approach leads to training a machine learning model to filter out documents that

To classify patents as AI-related or not, we rely on the BERT for Patents model developed by Srebrovic and Yonamine (2020), trained on a vast corpus of patent texts, encompassing over 100 million publications from all over the world. Transformer-based models are particularly suitable for text-classification tasks because they capture contextual relationships among words through self-attention mechanisms (Vaswani et al., 2017). BERT further exploits bidirectional representations, allowing the model to interpret terms in relation to both their preceding and following context (Devlin et al., 2018)[36]. This feature is particularly useful in patent classification, where technical terms often acquire meaning from the surrounding textual structure. BERT for Patents is especially appropriate for our task because it was pre-trained on a large corpus of patent publications and uses a tokenizer tailored to patent language[37], thereby improving the representation of technical and domain-specific expressions. This pre-training on patent texts significantly boosts downstream patent classification performance, given its exposure to the full range of patent terminology and phrasing.

We fine-tune a binary classifier head built on top of the pre-trained BERT for Patents architecture[38]. The training corpus is split into training, validation, and test subsets using a 70:15:15 ratio, following established practices (Lamperti, 2024). Patent titles and abstracts are concatenated into a single input sequence, separated by the special token [SEP], so that the model can process them as distinct but contextually connected text segments. We do not apply conventional preprocessing steps such as stemming, stop-word removal, or lemmatization, as these operations may disrupt the syntactic and semantic structure exploited by the BERT tokenizer.

Given the size of the training corpus and the objective of preserving the domain-specific knowledge already embedded in the pre-trained model, we adopt a feature-based transfer-learning strategy. Specifically, we freeze the layers of the base BERT for Patents model and train only the classification head. This reduces the risk of overfitting and catastrophic forgetting while allowing the classifier to adapt the pre-trained patent representations to our AI-identification task. The model outputs two logits, corresponding to the non-AI (negative) and AI (positive) classes, which are transformed into class probabilities through a SoftMax function. A patent is classified as AI-related when the predicted probability of the positive class exceeds 0.5; otherwise, it is classified as non-AI.

---

clearly differ from the seed set, it also has a key limitation: the binary sampling method can make the model less sensitive to borderline or ambiguous cases.

[36] As highlighted by Devlin et al. (2018), BERT model has been provided with a rich pre-training. The authors used BooksCorpus (800M words) and English Wikipedia (2,500M words) and trained the model on two unsupervised tasks: Masked Language Modelling and Next Sentence Prediction. This enables the model to leverage general language understanding, streamlining its fine-tuning for specific tasks, such as patent classification, with minimal architectural adjustments.

[37] Traditional tokenizers, trained on general corpora, often struggle with the specialized terminology and lengthy compound words found in patent filings. By creating a tokenizer attuned to patent-specific language patterns, the authors improved the model's ability to process and understand complex documents effectively. To explain this aspect, the authors refer to an example: while BERT's tokenizer would split the word "prosthesis" into <pro><thes><is> tokens, BERT for Patents tokenizer will keep <prosthesis> as a single token given its optimization. Such an improvement significantly boosts predictive accuracy.

[38] To this end, we leverage the Hugging Face Trainer API.

A3. Training and Inference

We briefly report the main validation and test statistics for the fine-tuned model, together with the results of the inference stage.

Table A2. Training and testing statistics

	T. Loss	V. Loss	Accuracy	F1 score	Precision	Recall
TRANING	0.12	0.10	0.962	0.945	0.942	0.947
TESTING	–	–	0.965	0.948	0.953	0.943

Table A2 summarizes the performance of the classifier. During fine-tuning, the training loss declined to 0.12, while the validation loss reached 0.10, indicating stable convergence. The final validation checkpoint achieved an accuracy of 0.962, an F1-score of 0.945, precision of 0.942, and recall of 0.947. Performance on the held-out test set is highly similar, with an accuracy of 0.965, an F1-score of 0.948, precision of 0.953, and recall of 0.943, alleviating concerns related to overfitting. The strong performance scores are consistent with expectations in automated patent landscaping applications. As noted by Abood and Feltenberger (2018), high classification metrics are a common feature in automated patent landscaping tasks, largely because the training data contains few borderline positive or negative examples. This design choice, favouring clearly relevant and irrelevant examples, enhances model precision and helps the classification system effectively remove patents that do not closely match the core seed topic. In addition, the F1 scores obtained in our analysis are broadly consistent with those reported by Abood and Feltenberger (2018)[39].

Table A3. Confusion matrix

	Predicted Label		
	NOT AI	**AI**	**TOTAL**
NOT AI	5,125 (97.6%)	125 (2.4%)	5250 (100%)
AI	153 (5.7%)	2533 (94.3%)	2686 (100%)

Table A3 reports the confusion matrix for the held-out test set. The model correctly identifies 2,533 true positives and 5,125 true negatives, while producing 125 false positives and 153 false negatives. The false-positive rate is therefore relatively low, indicating that the model is conservative in assigning AI

[39] The authors developed four distinct patent landscapes and trained separate classifiers for each. Their models obtain the following F1 scores: 0.987 for the “browser” topic, 0.965 for the “operating system” topic, 0.991 for the “video codec” topic, and 0.982 for the “machine learning” topic.

labels. This is a desirable feature for the present application, since downstream analyses rely on a reliable set of AI-related patents. At the same time, recall remains high, showing that this conservative behaviour does not come at the cost of missing a large share of genuinely AI-related documents.

Table A4. Results of the inference phase

	N. of documents retained	Share of documents retained	N. of patent families
Level 1 Expansion	367,364	48%	221,162
Level 2 Expansion	413,585	7.66%	288,428

We then apply the fine-tuned classifier to the patents retrieved through the two-level expansion procedure. The model is used to prune the expanded AI landscape by identifying which candidate documents contain substantive AI-related technological content. In total, the classifier is applied to 765,704 patent applications in the Level 1 expansion and 5,396,694 applications in the Level 2 expansion[40]. Table A4 reports the number and share of retained documents, together with the corresponding number of DOCDB patent families. The retention rates are consistent with the logic of the automated patent landscaping procedure. Since Level 1 patents are expected to be technologically and semantically closer to the original seed set, the classifier retains a substantially larger share of them: 367,364 documents, corresponding to 48% of the Level 1 expansion. By contrast, only 413,585 documents are retained from the Level 2 expansion, corresponding to 7.66% of candidate documents. This pattern aligns with Abood and Feltenberger (2018), as more distant expansion layers are expected to contain a larger proportion of unrelated or weakly related patents. After aggregating retained applications into DOCDB families[41], the final AI patent dataset contains 509,590 AI-related patent families with priority dates between 1980 and 2021.

A4. Empirical validation

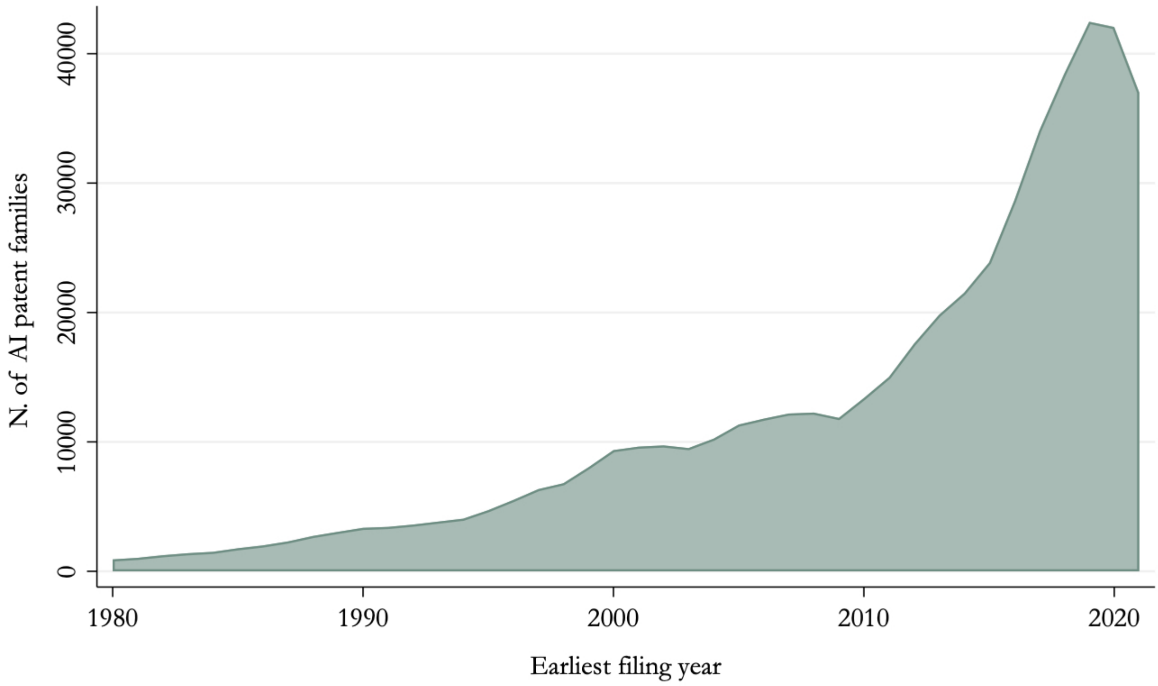
This section provides a set of empirical consistency checks for the AI patent dataset produced by the fine-tuned classifier. To avoid inflationary effects arising from multiple applications protecting the same invention, the analysis is conducted at the DOCDB patent family level.

We first examine the evolution of AI patenting over time. Figure A2 presents the number of AI-related patent families by earliest filing year. The series displays a gradual increase from the 1980s to the

[40] Inference is restricted to the two-level expansion, which defines the relevant candidate landscape around the AI seed set. Applying the model to the entire Patent Universe would substantially increase computational costs while likely adding only marginal recall gains.
[41] A DOCDB family is considered AI-related if at least one of its patent applications is classified as AI-related by the machine-learning model.

early 2000s, followed by a marked acceleration after 2010. This pattern is consistent with the evolution of the AI field and with the technological shift associated with the deep learning revolution (Sejnowski, 2018; Souza et al., 2024). It also closely mirrors the trends documented in previous patent-based studies of AI technologies (Alderucci et al., 2020; Dernis et al., 2021; Miric et al., 2023; Damioli et al., 2025)[42].

Figure A2. AI-related patenting activity over time

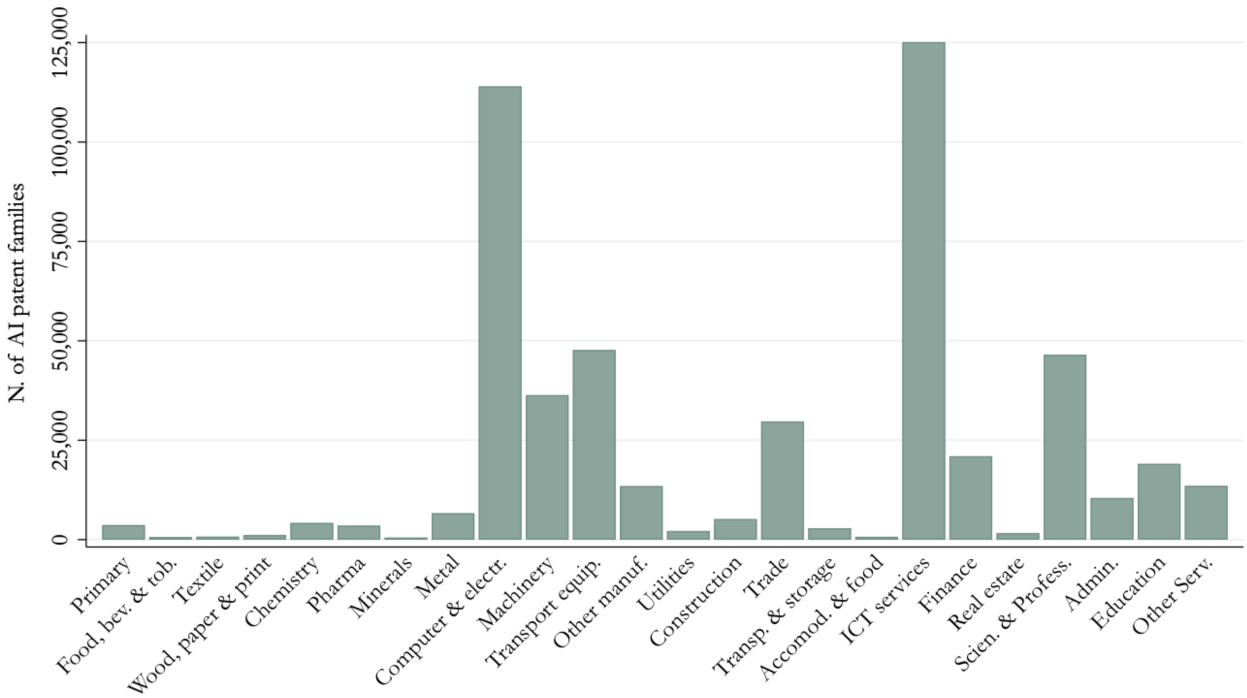


We then analyse the sectoral distribution of AI patenting. Following previous studies (Van Roy et al., 2020; Santarelli et al., 2023; Damioli et al., 2024, 2025; D'Alessandro et al., 2025), we link AI-related DOCDB families to their applicants and, through them, to economic sectors. This matching is performed using Moody's Orbis and Orbis IP, which provide firm-level information on applicants' core economic activity. Overall, we successfully assigned a two-digit NACE Rev. 2 sector to 461,585 AI patent families, corresponding to 90.5% of the AI corpus. We then aggregate NACE codes into broader sectoral classes following the grouping scheme proposed by Damioli et al. (2025). Figure A3 reports the distribution of AI-related patent families across these aggregated sectors. Consistent with prior evidence, ICT manufacturing and ICT services account for a substantial share of AI patenting activity (Van Roy et al., 2020; Dernis et al., 2021; Santarelli et al., 2023; Damioli et al., 2024, 2025; Calvino et al., 2024; D'Alessandro et al., 2025). This confirms the strong connection between AI and the broader ICT technological paradigm, reflecting the cumulative and path-dependent nature of AI knowledge production (Dosi, 1982 and 1988; Breschi & Malerba, 1997). At the same time, non-ICT sectors such as

[42] The apparent decline in the most recent years should be interpreted with caution, as it is likely driven primarily by publication lags and administrative delays in patent data rather than by a genuine contraction of AI inventive activity (Dass et al., 2017).

machinery, transport equipment, trade, and scientific and professional services also contribute significantly to AI inventive activity. This pattern is consistent with the increasing diffusion of AI technologies beyond their original ICT base and with their growing role as enabling technologies across different parts of the economy.

Figure A3. AI patenting activity across sectors



Note: Sectoral classes are composed by the following NACE codes. Primary: 1 to 9; Food, bev. & tob.: 10 to 12; Textile: 13 to 15; Wood, paper & print: 13 to 15; Chemistry: 19, 20 and 22; Pharma: 21; Minerals: 23; Metal: 24 and 25; Computer & electr.: 26 and 27; Machinery: 28; Transport equip.: 29 and 30; Other manuf.: 31 to 33; Utilities: 35 to 39; Construction: 41 to 43; Trade: 45 to 47; Transp. & storage: 49 to 53; Accomod. & food: 55 and 56; ICT services: 58 to 63; Finance: 64 to 66; Real estate: 68; Scien. & Profess.: 69 to 75; Admin.: 77 to 82; Education: 85; Other Serv.: 84 and 86 to 99.

Furthermore, we rank the most prolific AI applicants, providing an additional face-validity check of the classification procedure. Table A5 reports the leading applicants in terms of AI patent families. IBM leads the ranking, which also includes firms active across computing, electronics, software, semiconductors, telecommunications, industrial automation, automotive technologies, aerospace, and advanced manufacturing. This composition is consistent with the broad definition of AI adopted in this study (Van Roy et al., 2020; Damioli et al., 2024, 2025). Overall, the applicant ranking is consistent with evidence from prior studies and supports the view that the classifier retrieves a technologically heterogeneous AI landscape (UKIPO, 2014; Montobbio et al., 2022; Miric et al., 2023). Taken together, the sectoral and applicant-level evidence suggests that the identified patent set captures both the core computational components of AI (such as data processing, machine learning, computer vision, and platform-based software systems) and related advances in control systems, autonomous mobility, robotics, and advanced industrial technologies.

Table A5. AI patent families by Applicant

Firm	Pat. Families	Firm	Pat. Families
International Business Machines Corp.	20,189	Ford Global Technologies LLC	2,189
Samsung Electronics Co., LTD.	7,645	Amazon.com, Inc.	2,155
Microsoft Corporation	5,616	Hewlett-Packard Development Co LP	2,152
Microsoft Technology Licensing, LLC	5,346	Hyundai Motor Company	2,151
Google LLC	5,136	Ford Global Technologies Inc.	2,083
Siemens AG	4,723	Qualcomm Incorporated	2,073
Koninklijke Philips N.V.	3,974	GM Global Technology Operations LLC	1,960
Hitachi LTD	3,833	SZ DJI Technology Co., Ltd.	1,932
Intel Corp.	3,518	E.T.R.I.	1,904
Fanuc Corporation	3,418	Xerox Corporation	1,861
Robert Bosch GmbH	3,342	Apple Inc.	1,767
Fujitsu Limited	3,290	Honeywell International Inc.	1,718
Huawei Technologies Co., LTD.	3,195	Tencent Technology Co., LTD.	1,692
General Electric Company	3,192	Adobe Inc.	1,654
Mitsubishi Electric Corporation	3,138	Baidu Netcom S&T Co., LTD.	1,649
Sony Group Corporation	3,056	Meta Platforms, Inc.	1,562
NEC Corporation	2,937	SAP SE	1,558
Toyota Motor Corporation	2,875	Honeywell Int.	1,548
The Boeing Company	2,850	Denso Corporation	1,430
Honda Motor Co., LTD.	2,837	Kia Corporation	1,402
Toshiba Corporation	2,824	AT&T Inc.	1,301
Panasonic Holdings Corporation	2,815	Alibaba Group Holding Limited	1,216
LG Electronics Inc.	2,586	Seiko Epson Corporation	1,199
Canon Inc.	2,574	Bank Of America Corporation	1,160
Ping An Technology Co., LTD.	2,427	Korea Electronics & Telecom Co., LTD	1,147

Finally, we examine the CPC codes associated with the identified AI patent families to describe the main technological building blocks of the resulting AI landscape. Table A6 reports the top 20 four-digit CPC codes associated with AI patent families. The resulting technological profile is again consistent with the comprehensive definition of AI adopted in this study, in line with Van Roy et al. (2020). On the one hand, the landscape is strongly anchored in software- and data-intensive technologies, as shown by the prominence of G06F, G06N, G06T, G06V, G06Q, H04L, and H04N. These classes capture the learning, symbolic, data-processing, and perception-oriented components of AI, which are central to machine learning, computer vision, and information-processing methods. On the other hand, the table also reveals the embodied and cyber-physical dimension of AI. Several relevant CPC classes relate to control and autonomous systems, robotics, navigation, and sensing, including G05B, G05D, B25J, and B60W. Finally, the presence of medical, healthcare, and material-analysis classes, such as A61B, G16H, and G01N, also points to the diffusion of AI techniques across a broader set of application domains.

Table A6. Relevant four-digit AI CPC codes.

CPC 4-digit	Description	Share
G06F	Electric digital data processing	30.3%
G06N	Computing arrangements based on specific computational models	14.6%
G06T	Image data processing or generation, in general	13.6%
G06V	Image or video recognition or understanding	12.3%
G06Q	Information and communication technology [ICT] specially adapted for administrative, commercial, financial, managerial or supervisory purposes; systems or methods specially adapted for administrative, commercial, financial, managerial or supervisory purposes, not otherwise provided for	11.5%
A61B	Diagnosis; surgery; identification	7.7%
H04L	Transmission of digital information, e.g. telegraphic communication	7.3%
G05B	Control or regulating systems in general; functional elements of such systems; monitoring or testing arrangements for such systems or elements	6.6%
H04N	Pictorial communication, e.g. television	6%
G05D	Systems for controlling or regulating non-electric variables	5.3%
B25J	Manipulators; chambers provided with manipulation devices	4.9%
G01S	Radio direction-finding; radio navigation; determining distance or velocity by use of radio waves; locating or presence- detecting by use of the reflection or reradiation of radio waves; analogous arrangements using other waves	4.5%
G16H	Healthcare informatics, i.e. information and communication technology [ICT] specially adapted for the handling or processing of medical or healthcare data	4.2%
G10L	Speech analysis techniques or speech synthesis; speech recognition; speech or voice processing techniques; speech or audio coding or decoding	4.2%
B60W	Conjoint control of vehicle sub-units of different type or different function; control systems specially adapted for hybrid vehicles; road vehicle drive control systems for purposes not related to the control of a particular sub-unit	4%
G01C	Measuring distances, levels or bearings; surveying; navigation; gyroscopic instruments; photogrammetry or videogrammetry	3.8%
G08G	Traffic control systems	3.7%
H04W	Wireless communication networks	3.3%
Y10S	Technical subjects covered by former USPC cross-reference art collections and digests	2.6%
G01N	Investigating or analysing materials by determining their chemical or physical properties	2.3%

Note: The table displays the twenty most relevant 4-digit CPC codes within the sample of AI patent families resulting from the deep learning identification procedure. The share is computed as the ratio between the number of AI patent families being assigned to the focal CPC code and the total number of AI patent families.

References

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

Alderucci, D., Branstetter, L., Hovy, E., Runge, A., & Zolas, N. (2020). Quantifying the impact of AI on productivity and labor demand: Evidence from US census microdata. In Allied social science associations—ASSA 2020 annual meeting.

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(5), 1-68.

Bianchini, S., Müller, M., & Pelletier, P. (2022). Artificial intelligence in science: An emerging general method of invention. Research Policy, 51(10)

Breschi, S. & Malerba, F. (1997). Sectoral Innovation Systems: Technological Regimes, Schumpeterian Dynamics and Spatial Boundaries. In Edquist, C. (ed.), Systems of Innovation. Technologies, Institutions and Organizations, Routledge.

Calvino, F., Dernis, H., Samek, L., & Ughi, A. (2024). A sectoral taxonomy of AI intensity. OECD Artificial Intelligence Papers, No. 30, OECD Publishing, Paris.

Choi, S., Lee, H., Park, E., & Choi, S. (2022). Deep learning for patent landscaping using transformer and graph embedding. Technological Forecasting and Social Change, 175, 121413.

Cockburn, I. M., Henderson, R., & Stern, S. (2019). The impact of artificial intelligence on innovation. In Agrawal, A., Gans, J., & Goldfarb, A. (eds.), The economics of artificial intelligence: an agenda. University of Chicago Press.

Colombelli, A., D'Amico, E., & Paolucci, E. (2023). When computer science is not enough: universities knowledge specializations behind artificial intelligence startups in Italy. The Journal of Technology Transfer, 48(5), 1599-1627.

D'Alessandro, M., Santarelli, E., & Vivarelli, M. (2025). The KSTE + I approach and the advent of AI technologies: Evidence from European regions. Journal of Technology Transfer.

Damioli, G., Van Roy, V., Vértesy, D., & Vivarelli, M. (2024). Drivers of employment dynamics of AI innovators. Technological Forecasting and Social Change, 201, 123249.

Damioli, G., Van Roy, V., Vertesy, D., & Vivarelli, M. (2025). Is artificial intelligence leading to a new technological paradigm?. Structural Change and Economic Dynamics, 72, 347-359.

Dass, N., Nanda, V., & Xiao, S. C. (2017). Truncation bias corrections in patent data: Implications for recent research on innovation. Journal of Corporate Finance, 44, 353-374.

Dernis, H., Calvino, F., Moussiegt, L., Nawa, D., Samek, L., & Squicciarini, M. (2023). Identifying artificial intelligence actors using online data. OECD Science, Technology and Industry Working Papers, No. 2023/01, OECD Publishing, Paris.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv preprint arXiv:1810.04805.

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.

Giczy, A. V., Pairolero, N. A., & Toole, A. A. (2022). Identifying artificial intelligence (AI) invention: A novel AI patent dataset. The Journal of Technology Transfer, 47(2), 476-505.

Lamperti, F. (2024). Unlocking machine learning for social sciences: The case for identifying Industry 4.0 adoption across business restructuring events. Technological Forecasting and Social Change, 207, 123627.

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

Miric, M., Jia, N., & 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.

Montobbio, F., Staccioli, J., Virgillito, M. E., & Vivarelli, M. (2022). Robots and the origin of their labour-saving impact. Technological Forecasting and Social Change, 174, 121122.

Pairolero, N. A., Giczy, A. V., Torres, G., Islam Erana, T., Finlayson, M. A., & Toole, A. A. (2025). The artificial intelligence patent dataset (AIPD) 2023 update. The Journal of Technology Transfer, 1-24.

Santarelli, E., Staccioli, J., & Vivarelli, M. (2023). Automation and related technologies: a mapping of the new knowledge base. The Journal of Technology Transfer, 48(2), 779-813.

Sejnowski, T. J. (2018). The deep learning revolution. MIT press.

Souza D., Geuna A., & Rodríguez R. (2025). How Small is Big Enough? Open Labeled Datasets and the Development of Deep Learning. Industrial and Corporate Change, 34(6).

Srebrovic, R., & Yonamine, J. (2020). Leveraging the BERT algorithm for Patents with TensorFlow and BigQuery. White paper.

United Kingdom Intellectual Property Organization (UKIPO) (2014). Eight Great Technologies - Robotics and Autonomous Systems: A Patent Overview. London: UK Intellectual Property Office.

Van Roy, V., Vertesy, D., Damioli, G. (2020). AI and Robotics Innovation. In Zimmermann, K. (ed.), Handbook of Labor, Human Resources and Population Economics. Springer.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. Advances in neural information processing systems, 30.

World Intellectual Property Organization (WIPO) (2019). WIPO technology trends 2019–Artificial intelligence.