JRC141261

SCIENCE FOR POLICY BRIEF

The exposure of EU inventive efforts to critical raw materials: evidence from an AI based patent indicator

HIGHLIGHTS

→ Reducing uncertainty around critical raw materials (CRM) supply is a policy priority for the EU in view of their role for advanced carbon neutral and digital technologies.

→ A new, AI based indicator is introduced to measure the exposure of inventive activities to critical raw materials, outperforming existing approaches by identifying CRM relevance even when not immediately evident.

→ High exposure sectors, such as aerospace & defense and ICT services, intensify inventive efforts in response to CRM supply risk, indicating strategic shifts towards substitution and diversification.

→ European regions differ significantly in CRM exposure: some areas (e.g. parts of France, Germany, Italy, and Scandinavia) show considerable hidden CRM based inventive activity.

→ Firms in CRM exposed sectors adapt by both increasing their inventive efforts and seeking alternative inventive routes, suggesting that innovation can mitigate supply risk vulnerabilities.

CONTEXT

Critical raw materials (CRMs) are essential materials with significant economic importance and high supply risk, such as rare earth elements, lithium, silicon metal, and strontium. In May 2024, after approval by the European Parliament and the European council, the Critical Raw Materials Act entered into force. It ensures that EU member states have secure and sustainable access to CRMs.[1] This is considered crucial for achieving the 2030 climate and digital targets, as many relevant technologies strongly depend on the availability of various CRMs. This dependence represents a clear case in which the close interplay between innovation, environmental and socio-economic aspects engenders profound transformations of local industrial structures and reconfiguration of global value chains.

The risks associated with CRMs include mineral scarcity leading to depletion, concentrated geographical distribution of deposits, political instability in producer countries, geopolitical uncertainties in global trade and low recycling rates. Collectively, these risks can potentially limit industrial

[1] The CRM Act also introduced strategic raw materials which are a subset of critical raw materials deemed vital for advanced green, digital, defence, and aerospace technologies, for which the EU has set non-binding domestic capacity targets for extraction, processing, and recycling to reduce supply risks and dependency on single third-country sources

Joint Research Centre

development and hinder the progress of modern technologies (Li et al., 2024).

The European strategy to cope with the supply risk is based mainly on import diversification and de-risking, limiting both the concentration of supplies from one single non-EU country and the proportion of consumption derived from extraction. Additionally, enhancing recycling and circularity would limit CRMs' over-exploitation in the long run.

A greater effort is required to assess the extent to which the development of new strategic technologies depends, both directly and indirectly, on the supply of CRMs. This information is crucial to define the scope of the current EU strategy with the help of science, technology and innovation policies.

In this policy brief, we provide empirical evidence of the dependence of technological activities on CRMs across European regions and of its effects on firms' innovation. We first propose an AI-based methodology for detecting inventions related to CRMs, and then we develop a measure of technological exposure at the sectoral and regional level.

IDENTIFICATION OF CRM TECHNOLOGIES AND SECTORAL DYNAMICS

Previous efforts at detecting CRM-based inventions in patent documents rely mainly on keyword search (Fifarek et al., 2008; Li et al., 2024), sometimes combined with technological classification codes (CPC or IPC classification as in Diemer et al., 2022).[2]

Two main issues with this approach can be identified: (i) it does not account for the role of specific CRMs within patented technologies as described in patent texts. For instance, a CRM might be mentioned merely as an alternative material among others or in the context of a substitution technology, even when it is not essential to the process or technology described. (ii) CRMs are often referenced indirectly in patents or in relation to specific applications, which may result in their omission from keyword-based searches. [3]

To overcome these limitations, we develop a more granular methodology, exploiting large language models (LLM) and natural language processing (NLP) aimed at two objectives: (i) to disentangle the role of the CRM mentioned in patent texts ('importance'), and (ii) to extend the identification of CRM inventions to patents that do not mention CRM explicitly ('scope').

To reach the first objective, an LLM methodology is applied to worldwide patents' abstracts extracted from the PATSTAT Global Database (Spring 2023 Edition), covering the period 1990-2019. The LLM allows us to classify patents explicitly mentioning CRMs into non-substitutable and substitutable. *Non-substitutable CRM inventions* are those in which the invention relies necessarily on the material (e.g. CRM given in percentage terms; CRM included in a chemical reaction). *Substitutable CRM inventions* are instead those inventions in which a CRM is mentioned as an example, or the abstract describes a technology aimed at replacing the CRM.

To achieve the second objective, an NLP methodology based on the BERT for patents model (Yonamine, 2020) is developed and fine-tuned for each CRM group.[4] This approach enables the identification of patents related to technologies involving CRMs, even when the materials are not explicitly mentioned (for additional details on the two methodologies see Box 1).

As a result, this procedure generates a unique global dataset of CRM-related patents. We further match the list of CRM-related patents to the EU Industrial R&D Investment Scoreboard panel data (Nindl et al., 2023), providing information on the world's top 2500 companies with the highest R&D investment. This way, we are able to assign CRM patents to their main applicant and investigate the sectoral dynamics across the four categories of CRM inventions.[5] In particular, we collected all patent information at the patent family level for all the companies in the Scoreboard[6] from 1990 to 2019, we aggregated patent counts at the sectoral level and calculated the share of CRM inventions by sectors.

[2] The Cooperative Patent Classification (CPC) and International Patent Classification (IPC) are hierarchical coding systems used to categorise patents based on their technical content.

[3] For example, the description of a high-performance magnet might omit the explicit mention of the rare earth elements involved, such as samarium or neodymium.

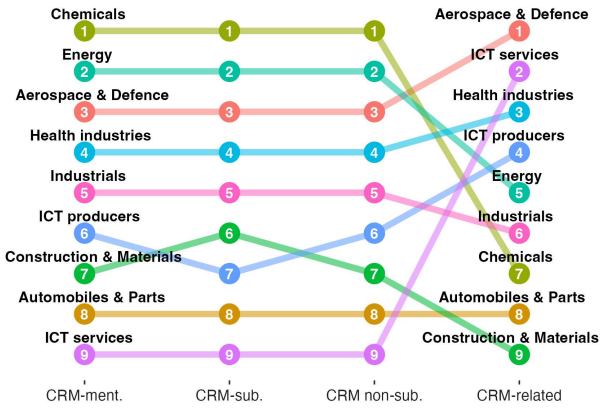
[4] A BERT for patents model is a transformer-based language model fine-tuned on patent text to better capture the domain-specific technical language and nuances for enhanced classification, semantic analysis, and retrieval tasks.

[5] For the purposes of this policy brief, companies and their patenting activity is allocated to sectors following the 3-digit ICB (International Classification Benchmark), consisting of 9 sectoral groups.

[6] We use patent families in order to avoid potential double counting of patent applications across patent offices.

Figure 1 shows the changes in the ranking of the share of CRM-based patent families across ICB sectors, based on four types of CRM identification approaches: (i) patents explicitly mentioning CRM, (ii) patents identifying the material as substitutable, (iii) patents identifying the material as non-substitutable, and (iv) CRM-related patents (patents involving CRM without explicitly mentioning the material in the patent text). In the figure, every column refers to one of the four types of CRM identification approaches, where sectors are ranked based on the value of the share of CRM-based patent families within the sector (from higher shares at the top to lower shares at the bottom). Coloured lines track the possible changes in the ranking position of each sector across CRM identification approaches.

Figure 1 – Sectoral rank in the share of CRM-related patents by type of CRM identification methodology



Source: own elaboration based on JRC Scoreboard and PATSTAT data.

Figure 1 shows that the chemicals sector has the highest share of CRM-mentioning patents, followed by the energy and aerospace & defence sectors. In contrast, the two ICT sectors rank lower, with ICT producers in sixth and ICT services in last place. The ranking remains relatively stable when considering the share of patents in CRM-substitutable technologies (CRM-sub) and CRM non-substitutable technologies (CRM non-sub). However, when focusing on CRM-related technologies, a notable reshuffling occurs in the relative importance of sectors engaging in technologies that use CRMs in their inventive activities. Specifically, the aerospace & defence sector emerges as the leader, with the highest share of CRM-related technologies, followed by ICT services, which moves from last to second place. The health industries sector gains one position, ranking third, while ICT producers rise two positions to claim fourth place in terms of CRM-related patent families.

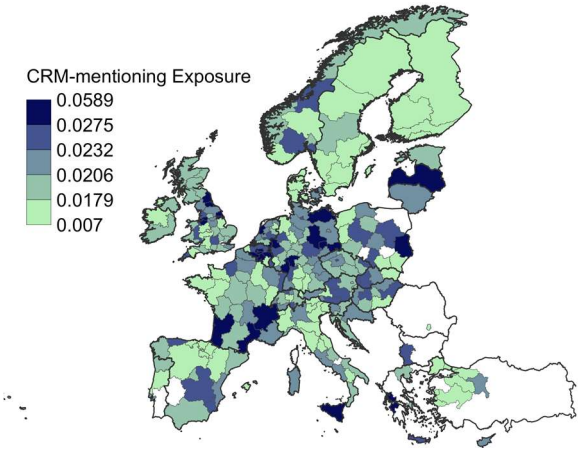
This descriptive evidence indicates that AI-based methods for identifying CRM-related technological efforts can reveal the strategic importance of CRMs in key sectors driving the twin transition. In particular, it highlights the enabling role of digitalisation, which might otherwise be underestimated.

CRM EXPOSURE IN EUROPEAN REGIONS

A novel indicator of regional exposure to CRMs has been developed by leveraging the identification of CRM-based patents in the EEA, the UK and Turkey. This indicator measures the extent to which the sectors constituting the reference market for regions' technological portfolios exhibit a high incidence of, and reliance on, CRM-based technologies.

In other words, a region's exposure to CRM is higher the higher its technological activity in a sector in which CRM-related technologies are a significant fraction of the overall technological activity. In order to highlight potential differential exposure levels stemming from different approaches to identifying CRM technologies, we build four exposure indicators, one for each of the four types of CRM identification methodology. For methodological details on the construction of the exposure indicators see Box 2.

Figure 2 – Regional exposure to CRM-mentioning technologies by NUTS2 region, average 2011-2019 (quintiles)



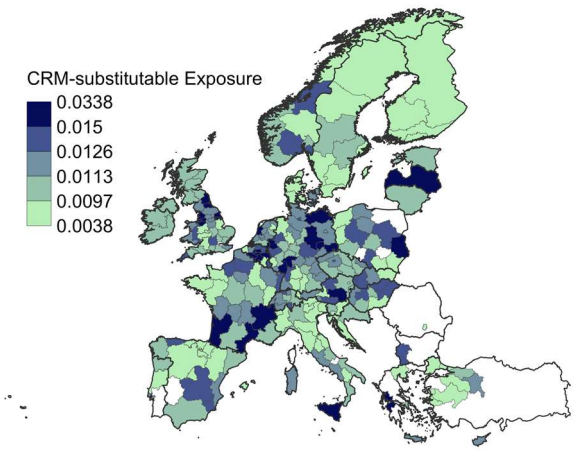
Source: own elaboration based on JRC Scoreboard and PATSTAT data.

Box 1: Identification of CRM technologies

The first research target in the granular identification of CRM inventions aims at understanding the importance of a CRM mentioned in the patent abstract, in order to classify the related CRM technology as substitutable or non-substitutable. This objective is achieved by employing the GPT-3.5-turbo-0125 model provided by OpenAI. 172,374 patent abstracts (excluding patents classified in the mining technology class) are collected and classified by each mentioned CRM. For the definition of the prompt, we adopted a chain-of-thought strategy in order to improve the classification robustness (Wei et al., 2022; Yu et al., 2023) and break down complex problems into more manageable ones. To illustrate, a patent mentioning a given CRM is classified as substitutable if: the CRM is mentioned as an example; the CRM is an option among other materials, and at least one of the other materials is not a CRM; the abstract discusses a technology to replace the CRM, and no other CRM is employed in the process. A patent mentioning a given CRM is classified as non-substitutable if: the presence of the CRM is shown in percentage terms; the CRM is included in a chemical reaction; the CRM is an option among other materials and all other materials are CRMs.

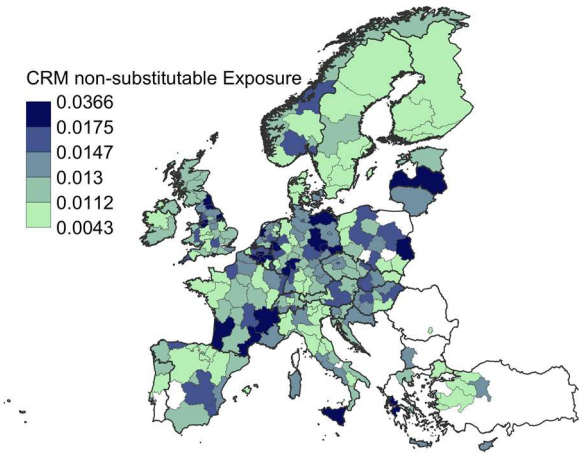
The second research target aims at identifying CRM-related patents beyond the presence of keywords. To do so, an NLP methodology is developed exploiting the BERT for Patents model (Yonamine, 2020). In particular, the sample of CRM mentioning patents is collected, and CRMs are divided into 13 groups (due to computational and data constraints); then, a BERT for Patents model is fine-tuned on each CRM group. The exposure to CRM-mentioning technologies by European NUTS-2 level regions is shown in Figure 2.[7] Regions in the southwest of France, central Germany, south Norway and southern Italy (Sicily) are characterised by higher technological exposure to CRM-mentioning inventions, while the lowest values can be mainly observed in Sweden, Finland, Northeast Italy, and central France. The distribution in the technological exposure of regions to CRM-non-substitutable (Figure 4) is very similar, as well as the distribution concerning the exposure to non-substitutable CRM technologies, though with some exceptions (Figure 3).

Figure 3 – Regional exposure to CRM-substitutable technologies by NUTS2 region, average 2011-2019 (quintiles)



Source: own elaboration based on JRC Scoreboard and PATSTAT data.

Figure 4 – Regional exposure to CRM non-substitutable technologies by NUTS2 region, average 2011-2019 (quintiles)



Source: own elaboration based on JRC Scoreboard and PATSTAT data.

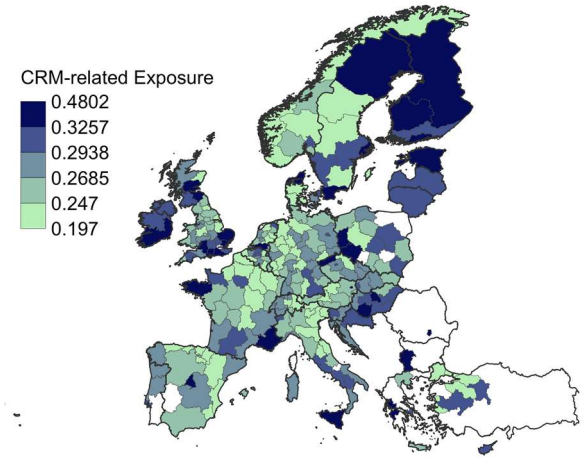
The map in Figure 5 shows notable regional differences in the technological exposure to CRM-related technologies. Regions such as Sicily (Italy), some regions in central Germany and southern France, are consistently highly exposed to CRM-based technologies, regardless of the specific type of CRM-based technologies considered. Regions in Scandinavia (particularly northern Sweden and Finland), the Baltic countries, the UK and Ireland exhibit the highest levels of technological exposure

[7] NUTS2 regions with less than 50 overall patents over the period 2001-2010 have been excluded.

to CRM-related technologies. Interestingly, substantial differences in exposure levels between CRM-mentioning and CRM-related technologies are evident in regions such as Bretagne (France), Helsinki and Pohjois (Finland), Stockholm (Sweden), Southern and Eastern Ireland, Sicily (Italy), and Île-de-France, among others. Specifically, these regions demonstrate significant gaps in exposure, with some appearing relatively less exposed when assessed using keyword-based identification. However, when broader CRM-related technological activities are considered, these regions exhibit markedly higher levels of exposure.

This descriptive evidence highlights the importance of identifying nuances in the use of CRMs in technological inventions. Specifically, it underscores the need to examine innovative efforts in the development and application of CRM-related technologies to better understand the strategic dependence of key sectors, such as carbon neutral and digital technologies, on the risks and shortages associated with CRMs.

Figure 5 – Regional exposure to CRM-related technologies by NUTS2 region, average 2011-2019 (quintiles)

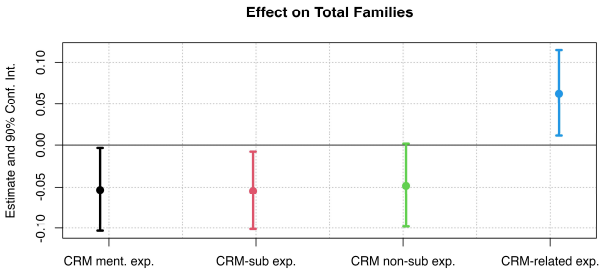


Source: own elaboration based on JRC Scoreboard and PATSTAT data.

CRM EXPOSURE AND FIRM INNOVATION

The matching of patent data with the EU Industrial R&D Investment Scoreboard also provides the opportunity to investigate the effects of CRM exposure on firms' innovation strategies in terms of the rate and direction of inventive activities.

Figure 6 – The effect of CRM exposure on firms' innovation



Source: own elaboration based on EU Industrial R&D Investment Scoreboard and PATSTAT data

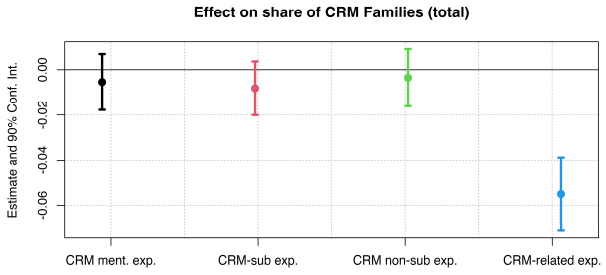
Firms in sectors more exposed to CRM are expected to leverage innovation as a strategy to respond to supply risk. Figure 6 compares the effect on firms' patenting of the four measures of sectoral CRM exposure.[8] Only the indicator based on the AI-based identification of CRM patents shows a statistically significant effect. The coefficient is positive, suggesting that the larger the exposure to CRM, the higher the supply risk, and the higher firms' inventive efforts in this direction. This finding highlights how firms in critical sectors exert inventive efforts to mitigate vulnerabilities in their supply chains, underscoring the role of targeted policies in fostering resilience through, for example, R&D support.

Figure 7 reports instead the results on the share of firms' CRM-based patents for the four measures of sectoral CRM exposure.[9]

[8] Figure 6 graphically reports the estimated coefficients, and the associated 90% confidence intervals, of the effect of CRM exposure on firm innovation, expressed in the following form: $Pat_{it} = \alpha + \beta_1 CRMexp_{it-1} + \beta_2 X_{it-1} + \phi_i + \gamma_t + \mu_{it}$. The subscript *i* denotes the firm and *t* the time-period. The dependent variable Pat_{it} is the firm total number of patent families; $CRMexp_{it-1}$ is, alternatively, the exposure of firm *i* at time *t-1*, to CRM mentioning, CRM-substitutable, CRM non-substitutable and CRM-related; X_{it-1} includes a set of firm-levels control variables (i.e., number of employees, capex and net sales). ϕ_i and γ_t, are, respectively, firm and time fixed effects; μ_{it} is the error term. The CRM exposure variables are standardised, while control variables are log-transformed. All models are estimated using fixed-effects panel regressions. We also cluster standard errors at the firm level.

[9] The coefficients reported in Figure 7 are estimated replicating the methodology described in footnote 5, where the dependent variable is the firm-level share of total CRM patent families out of the firm patent families portfolio.

Figure 7 – The effect of CRM exposure on firms' innovation



Source: own elaboration based on EU Industrial R&D Investment Scoreboard and PATSTAT data

Also in this case, the only CRM exposure indicator that yields a statistically significant coefficient is the one exploiting the AI-based identification of CRM patents. Accordingly, the larger the sectoral exposure to CRM supply risk, the higher the firms' efforts to drive their technological activities away from CRM. This result suggests that firms strategically reorient their innovation priorities to reduce dependence on CRM, indicating the importance of policies that support diversification in technological development to enhance supply chain security.

The results suggest that the AI-based CRM exposure indicator is a more reliable predictor of firm-level inventive strategies. It highlights the role of innovation as a strategic lever for addressing supply risks, particularly by redirecting research efforts towards less CRM-intensive domains or by substituting CRMs with alternative materials.

MAIN POLICY MESSAGES

European countries face a significant challenge regarding the supply risk of critical raw materials (CRMs). This issue is particularly pertinent due to the pivotal role of CRMs in advanced digital technologies that facilitate the green transition.

The results from this study underscore the urgent need to strengthen policy frameworks ensuring secure and sustainable access to CRMs and foster innovation-driven resilience in EU supply chains. In line with the Critical Raw Materials Act[10], the New European Innovation Agenda[11], the European Raw Materials Alliance (launched by the European Commission in 2020)[12], and the "Supply chain analysis and material demand forecast in strategic technologies and sectors in the EU – A foresight study" 2023 report[13], the following policy messages emerge:

→ Refine and expand CRM mapping

The novel AI-based methodology reveals that current keyword-based identification tools may underestimate the significance of CRM in technological development. Policymakers should invest in advanced data analytics and machine learning methods to regularly update inventories of CRM-dependent technologies, ensuring the Critical Raw Materials Act's implementation remains evidence-based and forward-looking.

→ Targeted innovation support for high-exposure sectors

Sectors with substantial hidden exposure to CRM-related technologies—such as aerospace & defence and ICT services—respond to supply risk by redirecting their inventive activities. Policymakers should tailor R&D incentives, Horizon Europe calls, and European Innovation Council funding instruments towards sectors facing acute supply risks, thereby strengthening technological sovereignty and aligning with the objectives of the European Raw Materials Alliance.

→ Strengthen regional innovation ecosystems and circularity in CRM-exposed areas

Our analysis identifies regions (e.g., parts of France, Germany, Italy, and Scandinavia) with elevated CRM exposure. European structural and investment funds, interregional innovation investments (I3), and smart specialisation strategies should be deployed to build local capacities in circularity and CRM-efficient design. This will boost regional resilience, advance the New European Innovation Agenda's place-based innovation focus, and accelerate the EU's green and digital transitions.

[10] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:L_202401252
[11] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52022DC0332
[12] https://erma.eu/
[13] https://data.europa.eu/doi/10.2760/386650

→ **Leverage AI-based insights for strategic foresight and policy coordination:**

The predictive power of the AI-based CRM exposure indicator suggests integrating such approaches into the policy cycle. By informing strategic foresight exercises (e.g., in the Commission's Joint Research Centre) and enhancing coherence among DGs (notably DG GROW, DG RTD, and DG ENV), policymakers can better anticipate emerging CRM bottlenecks and coordinate industrial, innovation, and trade policies to ensure sustainable and resilient value chains.

Box 2: Technological exposure to CRM

In order to calculate the index of technological exposure to CRM, we exploit a modified shift-share logic that leverages contemporaneous variation at the sectoral level and pre-sample variation at the region-sector level, where the pre-sample refers to the period 2001-2010. The index is constructed by multiplying the region's share of patents in each sector in the years 2001-2010 by the share of extended CRM patents over the total number of patents in each industry, where the extended CRM patents refer to the number of CRM patents plus the number of patents citing CRM patents in their backwards citations. Formally, the technological exposure of region r to CRM in industry j is defined as:

$$CRMexp_{rjt} = \frac{Pat_{rj}}{Pat_r} \times \frac{CRMPat_{jt} + CRMPatCit_{jt}}{Pat_{jt}}$$

where Pat_{rj} is the number of patents of region r in industry j over the period 2001-2010; Pat_r is the total number of patents of region r over the period 2001-2010; $CRMPat_{jt}$ and $CRMPatCit_{jt}$ are the number of CRM patents and the number of patents citing CRM patents, in industry j at time t, respectively; Pat_{jt} is the total number of patents in industry j at time t. The exposure to CRM of each region r at time t is thus obtained by summing the region-industry index each year for each region, as follows:

$$CRMexp_{rt} = \sum_j CRMexp_{rjt}$$

Four versions of the exposure indicators are calculated, depending on the set of CRM-identified inventions. To illustrate, the regional exposure to CRM mentioning inventions is calculated by exploiting, in the time-varying part of the indicator, only CRM mentioning patents and patents citing in their backward citations at least one CRM-mentioning invention. Similarly, the exposure to CRM-related patents only considers this last set of patents together with those citing as backward citations at least one CRM-related patent

REFERENCES

Diemer, A., Iammarino, S., Perkins, R., and Gros, A. (2022). Technology, resources and geography in a paradigm shift: the case of critical and conflict materials in ICTs. *Regional Studies*, 1-13

Fifarek, B., Veloso, F., and Davidson, C. (2008). Offshoring technology innovation: A case study of rare-earth technology. *Journal of Operations Management*, 26:222–238.

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

Nindl, E., Confraria, H., Rentocchini, F., Napolitano, L., Georgakaki, A., Ince, E., Fako, P., Tuebke, A., Gavigan, J., Hernandez Guevara, H., Pinero Mira, P., Rueda Cantuche, J., Banacloche Sanchez, S., De Prato, G. and Calza, E., The 2023 EU Industrial RandD Investment Scoreboard, Publications Office of the European Union, Luxembourg, 2023, doi:10.2760/73822, JRC135576.

Yonamine, J. S. R. (2020). Leveraging the BERT algorithm for patents with TensorFlow and big query.

DISCLAIMER OR OTHER FINAL DETAILS

The opinions expressed in this policy brief are those of the authors and cannot be attributed to the European Commission or any of its Member States.

SUGGESTED CITATION

European Commission: Joint Research Centre, Fusillo, F., Manera, M., Orsatti, G., Quatraro, F. and Rentocchini, F. (2025). The exposure of EU inventive efforts to critical raw materials: evidence from an AI based patent indicator, European Commission, Seville, 2025, JRC141261.

COPYRIGHT

CONTACT INFORMATION

francesco.rentocchini@ec.europa.eu