Investment Institute

Working Paper 151 | December 2023

# Answering Clean Tech Questions with Large Language Models



Trust must be earned

Amundi ASSET MANAGEMENT

# Answering Clean Tech Questions with Large Language Models

**Lauren STAGNOL**
*Amundi Investment Institute*
*lauren.stagnol@amundi.com*

**Amina CHERIEF**
*Amundi Investment Institute*
*amina.cherief@amundi.com*

**Zakaria FARAH**
*Amundi Investment Institute*
*zakaria.farah@amundi.com*

**Théo LE GUENEDAL**
*Amundi Technology*
*theo.leguenedal-ext@amundi.com*

**Sofia SAKOUT**
*Amundi Technology*
*sofia.sakout@amundi.com*

**Takaya SEKINE**
*Amundi Investment Institute*
*takaya.sekine@amundi.com*

## Abstract

The success of the "net zero transition" relies on the acceleration of the clean technology development to increase renewable energy capacity and low-emission solutions, but also to improve energy efficiency and enable carbon capture. Tracking such technologies and their mineral requirements is becoming increasingly important, but has traditionally required expert knowledge. In this paper, we propose a framework using Large Language Models and question-answering tasks to monitor the novelty within the clean tech industry, but also the minerals on which these technologies rely. It demonstrates the benefits of using artificial intelligence, and more specifically NLP techniques, to reconstruct expert knowledge and track rapidly changing markets.

**Keywords:** Net zero, clean technologies, critical raw materials, critical minerals, natural language processing, NLP, large language models, LLMs, information virality.

**JEL classification:** C8, O3, Q42.

**Acknowledgement**

*The authors are very grateful to Leila Bennani, Louise Duquenne, Aaron Mcdougall, Thierry Roncalli, Edmond Lezmi, Jiali Xu, Raphaël Semet, Tom Calamai and Jean-Marie Dumas for their helpful comments. They are also very grateful to Florian Vincent, Jeremy Jouvance, Frejus Atohoun, Rami Mery, and Matthieu Keip for their continued technical support. The opinions expressed in this research are those of the authors and are not meant to represent the opinions or official positions of Amundi Asset Management.*

## About the authors

### Lauren STAGNOL

Lauren Stagnol, PhD is an Equity Quantitative Researcher at Amundi Investment Institute. She conducts research projects and develops tools and strategies closely linked to portfolio management platforms. She joined Amundi in 2014 as a Fixed-Income Quantitative Researcher for three years. From 2017 to 2020, she was a Bond Portfolio Manager at Candriam in Bruxelles managing euro aggregate and smart beta portfolios. In 2020, Lauren re-joined Amundi as a Quantitative Researcher. Her areas of research cover smart beta, factor investing, sustainable investing, big data based economic and geopolitical narratives in both fixed income and equities.

Lauren completed her Ph.D. thesis entitled “Accounting for risk in the design of fixed-income benchmarks” in Economics at the Université Paris-Nanterre in June 2017, within the doctoral school “Economie, Organisation et Société”, co-accredited with les Mines Paris Tech. She also holds a Master of Science in Applied Economics and Econometrics from Université Paris-Nanterre and a Bachelor of Science in Business Economics from Cardiff University (UK).

### Amina CHERIEF

Amina Cherief is a Fixed Income Quant Researcher at Amundi Investment Institute. She conducts research projects closely linked to portfolio management platforms, the risk department and Amundi Intermediation.

Amina joined Amundi’s Quantitative Research Team in April 2017 to work on the development of a multi-factor risk and performance analysis tool. She worked from 2017 to 2018 as a Financial Engineer at Natixis AM and from 2018 to 2019 as a Cross-Asset Strategist at Natixis CIB where she was in charge of support and research (equity & commodity strategies, new portfolio allocation) for the QIS team. In 2019, Amina joined SG CIB in New York as a Cross-Asset Financial Engineer in the QIS team for two years; she was in charge of research and portfolio construction of systematic equity and commodity strategies. She developed tools for the sales and traders. Amina re-joined Amundi as a Quantitative Researcher in December 2020. Her areas of research are factor investing, sustainable investing and AI-ML in both fixed-income and equities. Recent advanced topics covered by Amina have been the integration of machine learning algorithms in the investment process of a fixed income team or the creation of innovative allocation in portfolios.

Amina holds a Master’s Degree (with honors) in Risk and Asset Management from Paris-Saclay University.

## Zakaria FARAH

Zakaria Farah joined Amundi Investment Institute in April 2023 in the Quant Portfolio Strategy team. He worked on Natural Language Processing techniques to identify and monitor clean tech related events. Zakaria is currently student at the University of Paris Diderot in the Master of Random modelling, Finance and Data science (M2MO, formerly known as DEA Laure Elie). Zakaria holds an engineering Degree from Ecole Centrale de Lille.

## Théo LE GUENEDAL

Théo Le Guenedal, PhD joined Amundi's Quantitative Research team in December 2018, to work on the performance of ESG investing in the equity market. Since then, he has been involved in an extensive research project on the incorporation of ESG factors and climate risks into asset allocation strategies. More recently, Théo has been focusing on the integration of advanced climate metrics, stress tests, and analytics in investment tools at Amundi Technology's Innovation Lab.

Théo completed his Ph.D. thesis entitled "Financial Modeling of Climate-related Risks" in Applied Mathematics at the Institut Polytechnique in December 2023. This research covers both transition and physical risks. A segment of his work on transition was recognized with the GRASFI Best Paper Prize for Research on Climate Finance, a distinguished honor sponsored by Imperial College London, in 2020. On the subject of physical risks, Théo also made significant contributions to the academic domain of physical risk assessment by creating the Tropical Cyclone Generation Algorithm.

## Sofia SAKOUT

Sofia Sakout, PhD is Lead Data Scientist in Natural Language Processing and Generative AI within the Innovation Lab of Amundi Technology. Sofia plays a central role in harnessing open source text data and in the development of NLP pipelines, with a particular focus on question-answering (QA) systems. Her focus at Amundi Technology is to channel innovation, data science, artificial intelligence toward practical applications in Amundi.

Sofia holds an engineering degree from École Mohammadia d'Ingénieurs, a Master's degree from École des Ponts ParisTech, and a PhD from the Sciences, Ingénierie et Environnement Doctoral School in partnership with École Polytechnique.

### Takaya SEKINE

Takaya Sekine, CFA is the Deputy Head of Quant Portfolio Strategy within Amundi Investment Institute (formerly known as the Quantitative Research Team of Amundi). In this role, he works on the practical implementation of quant research and alternative data for investment strategies.

He joined Amundi in 2000 and is in his current position since July 2018. Prior to that, he was Deputy CIO at Amundi Japan (between 2011 and 2018) with a focus on global quantitative strategies, Head of Index and Multi-Strategies at Amundi Japan (between 2010 and 2011), Fund Manager (between 2007 and 2010) and Financial Engineer (between 2001 and 2007). He has been involved in macro and policy related investment strategies for both retail and institutional clients. Takaya began his career as an IT Manager at Amundi Japan's predecessor company (between 2000 and 2001).

Takaya is a CFA charterholder since 2005 and an Associate member of the Association of Certified Fraud Examiners since 2010. He received the Ingénieur Civil des Mines degree from Ecole des Mines de Nancy in 2000.

# 1 Introduction

The past few years have seen a growing number of promising net zero commitments from both governments and companies. But there may be some bumps on the road toward a smooth transition to net zero. Indeed, turning words into action has raised questions about the capacity of our economies to make such structural changes, which in turn depend on non-infinite supplies of capital, labour and technological progress. The development of clean energy is a prerequisite for achieving such ambitious goals by reducing greenhouse gas emissions (IPCC, 2023), alongside low emission fuels, carbon capture or zero-emission technologies such as nuclear fusion, as John Kerry proposed during COP28[1]. Indeed, today's efforts are not enough as annual global $CO_{2eq}$ emissions have continued to rise, reaching a new high of 36.8 Gt in 2022 (IEA, 2023a). But clean energy development also requires financial incentives and policy makers' actions (Rasoulinezhad & Taghizadeh-Hesary, 2022). On this front, the Biden-Harris administration enacted the Inflation Reduction Act to promote clean energy (Rudolph *et al.*, 2022), while the European Commission for its part, proposed the REPowerEU plan (Deng *et al.*, 2022). Still, inter-governmental collaboration is also needed to achieve climate targets (IEA, 2023d). For example, the "Sunnylands Statement on Enhancing Cooperation to Address the Climate Crisis" issued between the US and China ahead of the COP28, provides a positive outlook for the global development of clean energy and could contribute to the effective "phasing out [of] all unabated fossil fuels" (UNFCCC, 2023) or a "transition away", as agreed during the event.

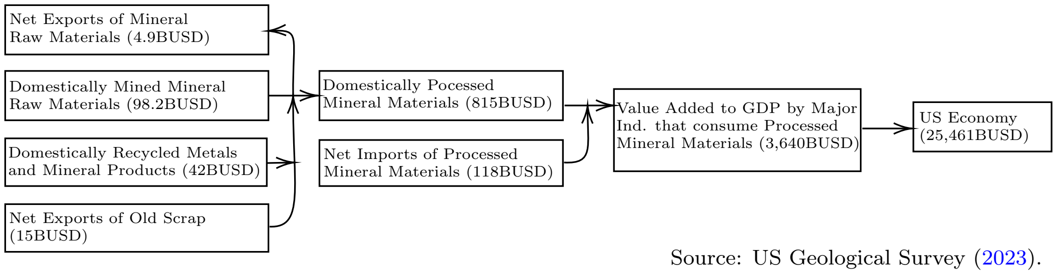
Besides, the International Energy Agency (IEA)'s scenarios (both Announced Pledges and Net Zero scenarios) assume a rapid growth in the energy supply from clean technologies - notably sourcing from modern bioenergy, wind or solar (IEA, 2023d) - a topic that the IPCC has been scrutinizing for more than a decade now (IPCC, 2011). However, the IEA also warns of the resulting tensions in the critical mineral markets. More precisely, demand is forecast to double by 2030, putting pressure on lithium, cobalt, nickel, and rare earth elements (REE) (IEA, 2023b). At the same time, questions are being raised about the environmental and health impacts of the entire life cycle of derived products, from exploration to recycling (e-waste), particularly for REE (Balaram, 2019). Doubts are also being cast on the supply side, with the focus on the ability to keep pace with such rapid demand growth. The burning issues of sources' diversification and sustainability of supply must also be addressed (IEA, 2022b). In this context, clean tech development has brought the focus of policymakers towards critical elements and minerals, already in the late 2000s/early 2010s (Eggert, 2011). This is a sizeable market: processed mineral materials (deriving from mineral raw materials transformation) represented 933 billion USD in 2022 according to US Geological Survey (2023), and their value added to the US economy is estimated to 3.64 trillion USD, representing 14% of the US GDP in 2022, as shown in Figure 1. The US is particularly vulnerable to imbalances in these markets, in the sense that demand and prices have been rising while import reliance is high: among the latest list, 31 critical minerals have net import reliance greater than 50%, and 15 at 100%.

Understanding the mineral requirements for clean technologies requires specialist knowledge and generic taxonomies may not be sufficient. To illustrate, as of the time of writing of this paper, the World Bank Group's topical taxonomy[2] has a "metal ore mining" category that includes the narrower concepts of copper, gold, industrial minerals, iron, nickel, silver, tantalum, tin, tungsten, and zinc. But it does not cover lithium for example, yet critical for clean tech (Haddad *et al.*, 2023). Furthermore, tracking mineral requirements for a net zero

[1] https://www.reuters.com/business/energy/us-envoy-kerry-launches-international-nuclear-fusion-plan-cop28-2023-12-05/.

[2] https://vocabulary.worldbank.org/taxonomy.html.

Figure 1: The Role of Nonfuel Mineral Commodities in the US Economy
(Estimated values in 2022)




Source: US Geological Survey (2023).

transition requires a given set of clean technologies. However, the latter cannot be bounded, as innovation may lead to the emergence of new clean technologies. Similarly, today's niche technologies could be industrialized on a global scale tomorrow, as demonstrated by the rapid increase in photovoltaic capacities over the last decade. In this context, Natural Language Processing (NLP) approaches seem particularly relevant. These methods are being increasingly used, notably in the finance industry (Kim *et al.*, 2023a, 2023b). They are fuelled by the emergence of generative AI offerings such as ChatGPT (Brown *et al.*, 2020), powered by Large Language Models (LLMs). Although researchers have questioned the performance trend of the most emblematic generative AI (Chen *et al.*, 2023), the financial industry has developed innovative tools. Li *et al.* (2023) propose a framework for the use of LLMs in finance, while on the sustainability aspects, Vaghefi *et al.* (2023) provide access to the expertise of the IPCC authors through conversational AI. We believe that NLP techniques - and in particular question-answering, could allow us to efficiently monitor clean technologies and thus demand for minerals.

In this paper, we aim to illustrate the benefits and pitfalls of using language models for understanding clean techs. Monitoring the development of clean techs and their demand for minerals with such techniques, without expert knowledge, would be valuable from the point of view of policy makers, but also of investors. We provide several tests and introduce a question-answering framework to monitor clean techs. Our results show it is possible for our industry to dramatically increase productivity through AI. This paper is structured as follows. Section 2 presents our benchmark analysis, testing NLP techniques and LLMs to replicate an IEA synthesis table of the critical minerals needs for clean techs. Section 3 introduces a "novelty detection pipeline" for the monitoring of novelty in clean tech from the Global Database of Events, Languages and Tone (GDELT). Finally, Section 4 offers some concluding remarks.

## 2 Identifying critical mineral needs for clean techs in the era of LLMs

Our objective is to measure the ability of modern NLP tools, such as naive extraction algorithms, to replicate the knowledge of analysts without a priori] knowledge in the clean tech domain. We take the particular case of the IEA's synthesis on the relevant connections between clean technologies and raw materials (IEA, 2022b). For this purpose, we create a collection of topical documents (CTD). When the IEA mentions the "list of critical minerals"[3], the explicit reference is to section 7002 on mineral security of the Energy Act of 2020 (US Senate Committee on Energy and Natural Ressources, 2021). A brief history of the IEA can be found in Appendix A.3. Considering its member states (see Figure 11 in the same subsection), the IEA is exposed to both the EU list of critical raw materials and the

[3] https://www.iea.org/policies/15271-final-list-of-critical-minerals-2022.

US list of critical materials. They are not the same. The US Geological Survey relies on the methodology of Nassar *et al.* (2020) to evaluate the supply risk of the US manufacturing sector. They measure the supply risk as a harmonic mean of foreign supply disruption potential, trade exposure, and economic vulnerability. The European European Commission (2020) has also been tracking "critical raw materials" (CRMs) and produced an initial list of 14 CRMs in 2011. Following the latest report of the European Commission (2023), a list of 34 CRMs has been established in the European Critical Raw Materials Act[4], based on economical importance and supply risk. We list the combination of the latest US critical minerals and EU critical materials in Table 9 in Appendix A.4.

## 2.1 Named-entity recognition

Named-entity recognition (NER) is often employed for identification purposes on a given collection of topical documents (Sang & De Meulder, 2003). In our case, an efficient NER model would require training to recognize the ten clean energy technologies (solar photovoltaic, wind, hydro, concentrated solar power, bioenergy, geothermal, nuclear, electricity networks, electric vehicles & battery storage and hydrogen) and nine associated critical minerals (copper, cobalt, nickel, lithium, rare earth elements, chromium, zinc, platinum group metals, and aluminum) identified by IEA (2022b). NER models typically involve a two-step process that first, detects named entities and then classifies them into the following categories: "localization", "organization", "person" or "misc" (miscellaneous). Given these categories, we deduce that a classical NER model could recognize clean energy technologies and critical minerals and classify them in the "misc" category. We test the performance of benchmark NER models: CamemBERT (Martin *et al.*, 2020) and BERT-large, BERT-base, roBERTa-large-NER which are fine-tuned on the English version of the CoNLL-2003 NER dataset. RoBERTa-large-NER (Conneau *et al.*, 2019) is based on Facebook's RoBERTa model. Liu *et al.* (2019) from Meta (Facebook) retrained the original BERT (Devlin *et al.*, 2019). We develop a custom benchmark dataset - Sample-CTD-NER - for NER consisting of 200 texts extracted from our CTD. We select 100 articles containing at least one critical mineral and 100 articles containing at least one clean tech from our aforementioned critical minerals and clean energy technologies respectively.

Table 1: Performance of NER models (in %)

(a) NER models - critical minerals

| NER Model | Precision | Recall | F1-score |
|---|---|---|---|
| camemBERT* | 0.75 | 0.34 | 0.46 |
| BERT-large** | 0.87 | 0.20 | 0.32 |
| BERT-base*** | 0.46 | 0.09 | 0.16 |
| roBERTa**** | 0.50 | 0.04 | 0.07 |
| deBERTa***** | 0.42 | 0.03 | 0.06 |

(b) NER models - clean tech

| NER Model | Precision | Recall | F1-score |
|---|---|---|---|
| camemBERT* | 0.93 | 0.41 | 0.57 |
| BERT-large** | 0.60 | 0.07 | 0.13 |
| BERT-base*** | 0.31 | 0.04 | 0.07 |
| roBERTa**** | 0.67 | 0.08 | 0.14 |
| deBERTa***** | 0.65 | 0.09 | 0.15 |

Note: *: Jean-Baptiste/camemBERT-ner, **: dslim/BERT-large-NER, ***: dslim/BERT-base-NER, ****: 51la5/roBERTa-large-NER, *****: Gladiator/microsoft-deBERTa-v3-large_ner_conll2003

Source: Amundi Investment Institute.

The results are presented in Table 1. The aim of this exercise is to find the exact number of critical minerals or clean technologies per text that should be classified in the "misc" category. An article may contain one or more clean technologies or one or more critical materials. Our performance measure is calculated from the sum of true positives, false positives, and false negatives found in each text. In our case, true positives correspond to the number of critical minerals and clean energy technologies detected in the text that fall

[4] https://ec.europa.eu/commission/presscorner/detail/en/ip_23_1661.

into the “misc” category. False positives translate the recognition of critical minerals and clean energy technologies in the fields of “localization”, “organizations” or “person”. Finally, if no critical minerals and clean energy technologies have been detected, it accounts for a false negative. In line with common practice in information retrieval (Goutte & Gaussier, 2005), we call precision the ability to minimize false positives and recall the ability to maximize true positives. The F1-score is the harmonic mean of Precision and Recall.

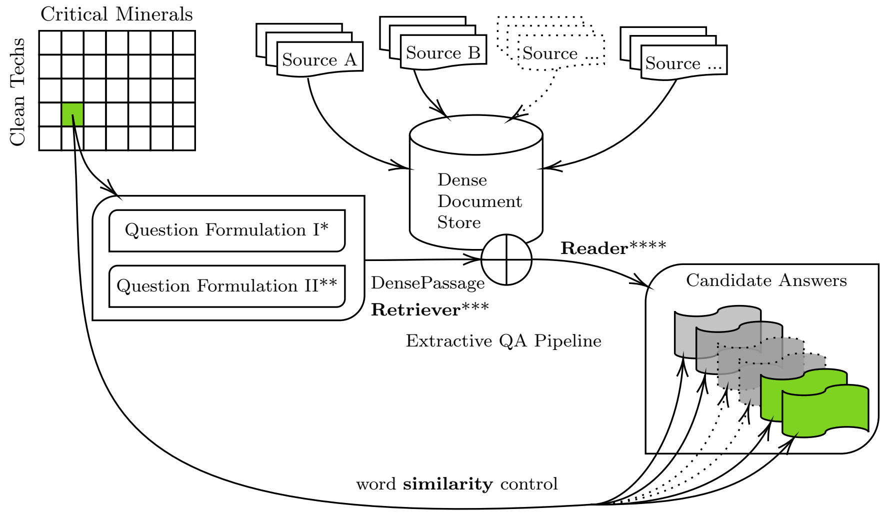
Thereby, Table 1a shows the F1-score of the selected NER models for the 100 articles containing at least one critical mineral per text. We find that the model with the highest F1-score is the CamemBERT model with a score of 46.49%. Accordingly, many critical minerals are correctly detected in “misc”, particularly lithium, otherwise the others are false positives (detected as “organization” or “localization”). This model also recognizes elements such as hydrogen or boron, or minerals such as beryllium 26 ores, sodium or helium. CamemBERT performs well on English texts since it focuses on words that are identical in French and English (such as lithium), hence correctly assigned to the “misc” category. However, as CamemBERT is a French language model it is legitimate to translate the texts from Sample-CTD-NER to analyze whether the quality of named entity recognition remains interesting. But translating the text into French means that the model has to classify and focus on a larger number of words, and therefore sometimes fails to assign minerals or materials to the “misc” category. In addition, the remaining NER models (BERT, roBERTa and deBERTa) exhibit low F1-scores ranging from 31.95% to 6.80% for the latter. For these models, lithium is the critical mineral which is the most frequently detected as a “misc”, resulting in high precision scores.

In the second step, we seek one or more clean energy technologies per text in 100 articles, using the aforementioned NER models. The performance scores of the models are presented in Table 1b and the best-performing model remains the CamemBERT model. In fact, the F1-score is even higher than in the previous example, reaching 56.83% with an accuracy of 92.86%. However, after translating the texts of the Sample-CTD-NER into French, the F1-score falls to 12.86%. For the remaining models, this time, we find that the F1-scores of the reworked models are lower and inversely ranked compared to the previous exercise, where the best-performing model after CamemBERT is deBERTa with a score of 15.49% and the worst-performing is BERT with a score of 7.35%, proving the instability of the models' performance across different text topics. Therefore we believe that a custom NER model, trained exclusively on documents in the clean tech lexical domain, is required. Achieving this goal, however, requires the construction of a sufficiently large text dataset.

## 2.2 Question-answering

Instead of the NER method, we propose an information extraction process based on question-answering, as shown in Figure 2. We do not simply ask our CTD which minerals are required for clean technologies in general with a question-answering trained model. We also query the CTD for each combination of critical minerals × clean energy technology from Table 2a. Critical minerals are derived from section 7002 of the Energy Act of 2020 (US Senate Committee on Energy and Natural Ressources, 2021). We therefore ask the questions at the level of each material, rather than at the level of REE and platinum group metals (PGMs), before recombining the output. For example, we can establish from Table 2c that the relevant material for REEs is neodymium (Nd), which is used for permanent magnets. Unlike in the previous NER case, we perform the analysis on the full CTD, not an extract. We apply two formulations to our question, to remove any bias in the order of appearance of mineral or clean tech in our question and run our information extraction pipeline in the Haystack open source framework (Pietsch *et al.*, 2019).

Figure 2: Information extraction pipeline




Note: $^\star$: Is **mineral** necessary for **techno** ?, $^{\star\star}$: Does **techno** require **mineral** ?, $^{\star\star\star}$: test retrievers, $^{\star\star\star\star}$: test readers for QA task.

Source: Amundi Investment Institute.

Table 2: Critical mineral needs for clean tech

(a) Information detection pipeline

| | Cu | Co | Ni | Li | REEs | Cr | Zi | PGMs | Al |
|---|---|---|---|---|---|---|---|---|---|
| Solar PV | ● | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ● |
| Wind | ● | ○ | ○ | ○ | ● | ○ | ● | ○ | ○ |
| Hydro | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| CSP *** | ● | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ● |
| Bioenergy | ● | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| Geothermal | ○ | ○ | ● | ○ | ○ | ● | ○ | ○ | ○ |
| Nuclear | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| Elect Net* | ● | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ● |
| EVs / BS** | ● | ● | ● | ● | ● | ○ | ● | ○ | ● |
| Hydrogen | ● | ○ | ● | ● | ○ | ○ | ○ | ● | ○ |

Source: Authors' calculations

(b) IEA (2022b)

| | Cu | Co | Ni | Li | REEs | Cr | Zi | PGMs | Al |
|---|---|---|---|---|---|---|---|---|---|
| Solar PV | ● | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ● |
| Wind | ● | ○ | ◐ | ○ | ● | ◐ | ● | ○ | ◐ |
| Hydro | ◐ | ○ | ○ | ○ | ○ | ◐ | ◐ | ○ | ◐ |
| CSP | ◐ | ○ | ◐ | ○ | ○ | ● | ◐ | ○ | ● |
| Bioenergy | ● | ○ | ○ | ○ | ○ | ○ | ◐ | ○ | ◐ |
| Geothermal | ○ | ○ | ● | ○ | ○ | ● | ○ | ○ | ○ |
| Nuclear | ◐ | ○ | ◐ | ○ | ○ | ◐ | ○ | ○ | ○ |
| Elect Net* | ● | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ● |
| EVs / BS** | ● | ● | ● | ● | ● | ○ | ○ | ○ | ● |
| Hydrogen | ○ | ○ | ● | ○ | ◐ | ○ | ○ | ● | ◐ |

Source: IEA

(c) Information extraction pipeline (continued)

| | Heavy Rare Earth Elements | | | | | | | | | | Light Rare Earth Elements | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $Dy^m$ | Er | Eu | $Gd^m$ | Ho | Lu | $Tb^m$ | Tm | Yb | Y | $Ce^m$ | La | $Nd^m$ | $Pr^m$ | $Sm^m$ |
| Solar PV | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| Wind | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ● | ○ | ○ |
| Hydro | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| CSP *** | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| Bioenergy | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| Geothermal | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| Nuclear | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| Elect Net* | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| EVs / BS** | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ● | ○ | ○ |
| Hydrogen | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ |

Source: Authors' calculations

Note: for Subtable 2b: ● = high, ◐ = moderate, ○ = low;
$^\star$: Electricity Networks; $^{\star\star}$: Electric Vehicles and battery storage; $^{\star\star\star}$: Concentrated Solar Power (CSP not being recognized as a proper abbreviation in our pipeline).

Zhu *et al.* (2021) conduct a survey on question-answering and introduce the retriever-reader architecture where the retriever extracts documents with probable answers and the reader focuses on predicting the start and end position of the answer from the retrieved documents as illustrated in Figure 10 in Appendix A.2. For a given reader (roBERTa-base-squad2), we test the performance of different retrievers, namely the Facebook dense retriever, mpnet and sentence transformers, and present the results in Table 8 in Appendix A.1. Since the Facebook dense retriever (Karpukhin *et al.*, 2020) performed better in this test, we use it as our reference retriever in the rest of our analysis. We then test several readers, namely roBERTa, electra, xlm and alBERT and present the results in Table 3.

Table 3: Performance of information extraction with different reader models

| Reader Model | Precision | Recall | F1-score | Execution time of the model |
|---|---|---|---|---|
| roBERTa⋆ | 0.76 | 0.80 | 0.78 | 98 minutes and 27 seconds |
| electra⋆⋆ | 0.74 | 0.70 | 0.72 | 144 minutes and 6 seconds |
| xlm⋆⋆⋆ | 0.76 | 0.65 | 0.70 | 146 minutes and 54 seconds |
| alBERT⋆⋆⋆⋆ | 0.83 | 0.95 | 0.88 | 369 minutes and 44 seconds |

Note: ⋆: roBERTa-base-squad2, ⋆⋆: electra_large_discriminator_squad2_512, ⋆⋆⋆: xlm-roBERTa-large-squad2, ⋆⋆⋆⋆: alBERT_xxlargev1_squad2_512

Source: Author's calculations, Amundi Investment Institute.

To complement our information extraction pipeline, we introduce a Term-Frequency Inverse Document Frequency (TF-IDF) measure (Ramos *et al.*, 2003) and we also control for cosine similarity. It measures the similarity of two n-dimensional vectors by finding the cosine of their angle, between the mineral and the answer context, as well as, on the clean energy technology and the answer context. In this way, we filter out inconsistencies from the raw results to compensate for the fact that alBERT or the other models that we have used are not specifically trained to detect clean techs or minerals. This step is important as our retention rates, shown in Tables 4a and 4b, are in general below 30%. This retention rate is presented in the bottom-right box of Figure 2. In fact, the retriever and the reader provide candidate answers, of which only a minority are "confirmed" by the world similarity control, which acts as a "safety net". We give an example of the lack of specific training for minerals/materials and clean energy technologies in Table 5 for the alBERT model.

Table 4: Filtering from word similarity control

(a) Question Formulation I

| reader model | critical material | tech |
|---|---|---|
| roBERTa⋆ | 0.204 | 0.248 |
| electra⋆⋆ | 0.196 | 0.229 |
| xlm⋆⋆⋆ | 0.196 | 0.242 |
| alBERT⋆⋆⋆⋆ | 0.216 | 0.245 |

(b) Question Formulation II

| reader model | critical material | tech |
|---|---|---|
| roBERTa⋆ | 0.162 | 0.301 |
| electra⋆⋆ | 0.155 | 0.306 |
| xlm⋆⋆⋆ | 0.157 | 0.271 |
| alBERT⋆⋆⋆⋆ | 0.164 | 0.318 |

Source: Authors' calculations, Amundi Investment Institute

Note: ⋆: roBERTa-base-squad2, ⋆⋆: electra_large_discriminator_squad2_512, ⋆⋆⋆: xlm-roBERTa-large-squad2, ⋆⋆⋆⋆: alBERT_xxlargev1_squad2_512

As shown in Table 3, using alBERT[5] model we manage to emulate the IEA Table 2b with an F1-score of 88%. The real significance of this number is that while we lack the subtlety of the IEA experts to identify a moderate need (◯ in Table 2b) for critical minerals for clean energy technologies, we can emulate the IEA experts' know-how at 88% with transformer-based language models and a systematic question-answering approach.

Table 5: Word similarity control in application

| critical material | clean energy tech | context | alBERT | material to context | clean energy tech to context | Final |
|---|---|---|---|---|---|---|
| Aluminium | Solar PV | **Aluminium** is an important input to the clean energy transition, with the production of several clean technologies, including **solar PV** installations and EVs, requiring significant amounts | ✓ | ✓ | ✓ | ✓ |
| Iridium | EVs and battery storage | **EVs and battery storage** grows nearly tenfold in the STEPS and around 30 times in the SDS over the period to 2040. By weight, **mineral** demand in 2040 i | ✓ | ✗ | ✓ | ✗ |

Source: Authors' calculations, IEA (2022b, 2023c), Amundi Investment Institute

In the first example, alBERT returns a context where we can identify both the critical material (Aluminium) and the clean tech (Solar PV). However, in the second example, alBERT returns a context with the clean tech (EVs and battery storage) but is not precise enough for the critical material. It identified the concept of mineral but not precisely Iridium. In practice, alBERT's F1-scores are the best of the models we tested. However, if we consider both the execution time and the F1-score, RoBERTa appears to be the best trade-off, as shown in Table 3. Hence, we will pursue our analysis with this model. We highlight that the use of the very precise alBERT model yields powerful results for emulating expert knowledge, although we choose to use the less dense roBERTa model for computational reasons. The context extraction feature from the question-answering task enables us to establish a word similarity control. This purely syntactic layer - with no LLM involved - acts as a very efficient "safety net" for the outputs of the alBERT model. Indeed, in our information extraction study, we removed about 70% of the false positives from the candidate responses generated by alBERT. This double contribution, with the extraction of context using the language model, and the syntactic check, is a proposal that emulates "human control" from the context.

# 3 Monitoring of novelty in clean tech

## 3.1 Event database

In our previous work, we built *economic, societal* and *geopolitical* narratives (Blanqué *et al.*, 2022), based on news information using the big data framework proposed by Leetaru and Schrodt (2013). In this second version of the Global Database of Events Language and Tone (GDELT 2.0), the authors link their dataset to the Conflict and Mediation Event Observations (CAMEO) framework. We find that the translation of news from over 100 languages into English is a significant advantage of GDELT 2.0 for news and event-based analysis. This version of the database processes metadata that includes the nature of the events, the people or entities involved, the location, and codes to describe the event using

[5] albert xxlarge version 1 language model fine-tuned on SQuAD2.0.

multiple taxonomies and the dictionary-based tonality of the news. In our previous approaches, we grouped the components of these taxonomies according to their relationship to our *narratives*' themes. In terms of tonality, Leippold (2023) confirms the vulnerability of keyword-based approaches, (for example, the one employed by Loughran and McDonald (2011)) to adversarial attacks. The author uses GPT-3 and synonyms to transform the keyword-based sentiment measure of texts from negative to neutral "while respecting the context, the meaning, and the grammar". Kurakin *et al.* (2017) and Papernot *et al.* (2015) present adversarial attacks as attempts by malicious adversaries to intentionally exploit the "loopholes" of the deep learning systems to achieve misclassifications. An early example of an adversarial attack is described by Dalvi *et al.* (2004) with an analysis of spam filtering.

Leetaru (2021) introduces GDELT's Web News NGrams 3.0 Dataset (GDELT 3.0). To illustrate that this new dataset brings context, he uses the example of the term **delta**. This approach enables the differentiation between "**delta** airlines" or the "**delta** Covid-19 variant". An advantage of GDELT 3.0 for news and events analysis is that we can focus our study on areas that are not *a priori* dependent on the depth of the third-party taxonomies available through GDELT 2.0. Specifically in Blanqué *et al.* (2022) for our "innovation" theme within the environment narrative, we identify GDELT 2.0 identifiers: ENV Carboncapture, WB 1853 Hydrofluorocarbons, WB 1851 Biocarbon, WB 2003 Sanitation Technologies, WB 2639 Climate Efficient Industries, WB 2673 Jobs and Climate Change, WB 2674 Green Jobs, WB 399 Innovation for Green Growth, WB 400 Innovation Driven Inclusive Growth, WB 408 Green Buildings, WB 568 Climate Services, WB 571 Climate Science, WB 573 Climate Risk Management. The identifiers in GDELT 2.0 with the prefix "WB" refer to the aforementioned World Bank Group Topical Taxonomy. GDELT 3.0 offers greater flexibility through its N-grams approach, compared to the constrained set of identifiers available in GDELT 2.0 through its event taxonomies. As our objective is to create a novelty monitor on clean techs, we select the GDELT 3.0 dataset over GDELT 2.0 and we manage to recompose the text of approximately 95% of the articles covered in GDELT 3.0 on a daily basis. This exercise focuses on news in English as we do not have the 100+ language translation feature of GDELT 2.0.

As new technologies emerge, we cannot guarantee that the components of the pre-defined taxonomies will adequately cover them, even if extensive studies can be found (IEA, 2023c). To anticipate potential novelties, we could go upstream in the innovation cycle and enrich our taxonomy with the information contained in patents (IEA, 2021). However, Artificial Intelligence is known to have gone through "AI winters" caused in particular by overly high expectations of expert systems (Hendler, 2008) and the underlying knowledge bases. In Figure 12 in Appendix A.5, we present a knowledge graph describing the relationships between minerals and clean techs, but also their interdependencies. We think that designing a robust approach to capture novelty is a better option than just expanding our knowledge base. Accordingly, in subsection 3.2, we present our novelty detection pipeline for clean techs. A preliminary check consisted of a robustness test of our process. We focus on the news of December $12^{th}$, 2022, when a scientific breakthrough in **nuclear fusion**[6] was widely reported. We run the novelty detection pipeline on both the raw GDELT 3.0 data for this given date (Sample-GDELT-Fusion), and on an equivalent dataset (Sample-GDELT-Bananas) where we replace all occurrences of **nuclear fusion** with the imaginary concept of **green bananas**. The results are presented in Figure 3. Our novelty detection pipeline was able to identify **green bananas** as a novel technology, demonstrating the robustness of our approach.

[6] https://www.nationalgeographic.com/science/article/scientists-achieve-breakthrough-nuclear-fusion.

Figure 3: **Green bananas** detection test on 12/12/2022

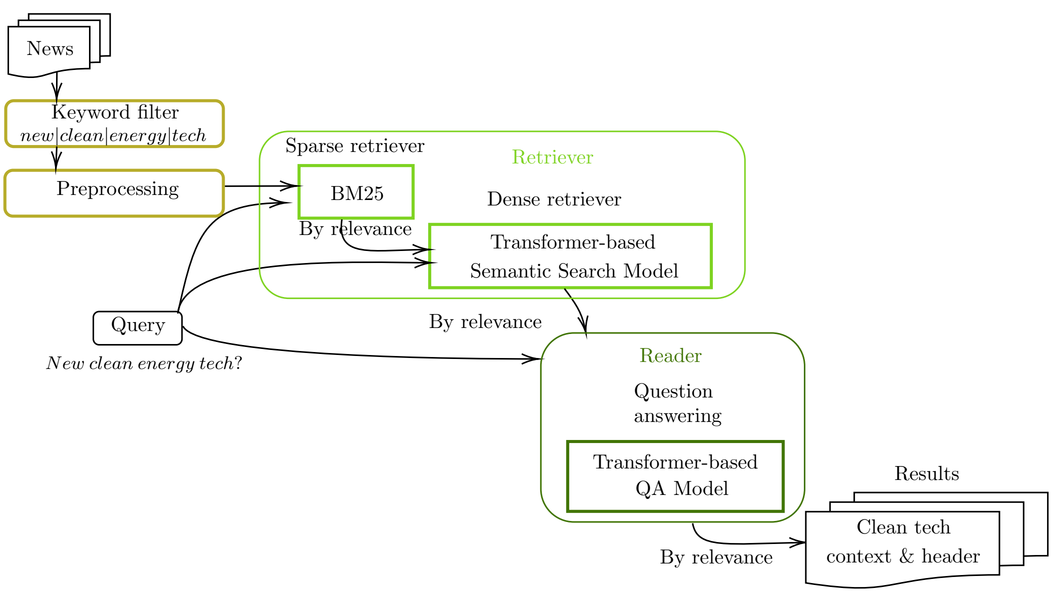

(a) raw GDELT 3.0 (b) **green bananas** in-lieu of **nuclear fusion**

Source: Authors' calculations, Amundi Investment Institute.

## 3.2 Filter, retriever, and reader

To identify novelty in clean tech from global events through GDELT 3.0, we propose a novelty detection pipeline. We have an upstream process with the filtering, retriever, and reader steps shown in Figure 4. This section detects relevant news. This is followed by a downstream process where we classify the detected clean tech news along large families of clean techs. Our goal is to monitor the daily sentiment distribution for each clean tech class. These distributions allow us to measure novelty for each clean tech class.

Figure 4: Upstream novelty detection pipeline: filter, retriever & reader

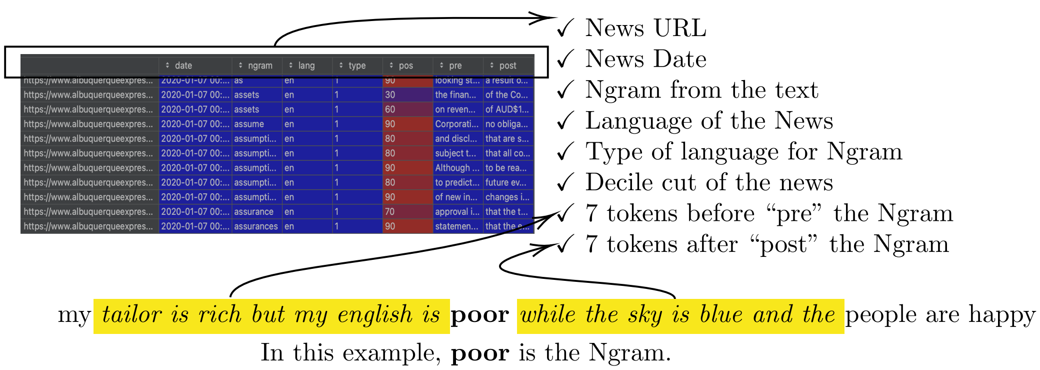



Source: Amundi Investment Institute.

Given that, on the one hand, the ability of NER models to detect both critical minerals and clean tech in texts is limited, as exhibited in Section 2, and that on the other hand, word embedding approaches generally have longer execution times than lexical search methods (Brunila & LaViolette, 2022), we decided to use the same question-answering approach as described in Section 2. This decision also takes into account our non-infinite computational resources. In the spirit of Syed *et al.* (2021), we build our question-answering detection pipeline with Haystack.

We receive news items daily in the N-gram format from GDELT V3 as described in Figure 5. We apply a simple filter to the N-grams to retain only articles that contain at least one of the following four words: "new", "clean", "technology" or "energy". This simple novelty filter for clean tech allows us to eliminate articles discussing unrelated topics and thus to reconstruct a smaller number of targeted articles. We rebuild full articles from the N-grams. This step is essential to provide consistent contextual text for transformer-based models. In fact, as displayed in Table 5, the extracted context can be superior to 15 tokens (7 "pre" tokens + 1 NGram + 7 "post" tokens). In the pre-processing, we clean up our articles by removing blank lines, redundant spaces, and special characters. Articles are divided into sub-documents of no more than 200 words. Thus, each article consists of several sub-documents of fixed size (200 words/tokens), which standardizes our subsequent searches. Such a "chunk" approach also ensures that we do not exceed the input size that the language models can accept.

Figure 5: GDELT's Web News NGrams 3.0 Dataset




Source: https://blog.gdeltproject.org/announcing-the-new-web-news-ngrams-3-0-dataset/, Amundi Investment Institute.

The next step in our detection pipeline is retrieval, one of the main components of the question-answering mechanism, which allows us to classify and select the potential documents that answer our question. Our query is simply "New clean energy tech ?". To perform this type of search, we have two options: either we use sparse methods based only on the number of occurrences, or we use methods where we create embeddings of both the query and the texts and calculate cosine similarity scores between the two vectors. The latter is more powerful and accurate because, unlike sparse methods, it takes the context into account when vectorizing. This semantic search concept is described in the SentenceTransformers Reimers and Gurevych (2019) documentation as a way "to improve search accuracy by understanding the content of the search query". Semantic search has many advantages, but also some limitations, most notably the execution time since vectorizing the query and the texts using a transformer-based model is time-consuming. Therefore, to be efficient in terms of computation time, we use two layers of retrieval. In the first layer, we use "Okapi BM25" after the preprocessing step to handle sparse retrieval. This battle-tested statistical term-based model aims to provide accurate and relevant search results by ranking documents based on term frequency and document length (Robertson, Zaragoza, *et al.*, 2009). After calculating this score (within [0:1]) for all our document segments, we retain the most relevant documents (with the BM25 score above 0.5). Our second layer is a dense retriever. We compute the embeddings of all the retained news segments, as well as those from our query, using the Facebook transformer-based model, because of its solid results on our IEA benchmark. We then compute the cosine similarity between the two vectors. We select the most relevant scores (scores above the average of a representative sample).

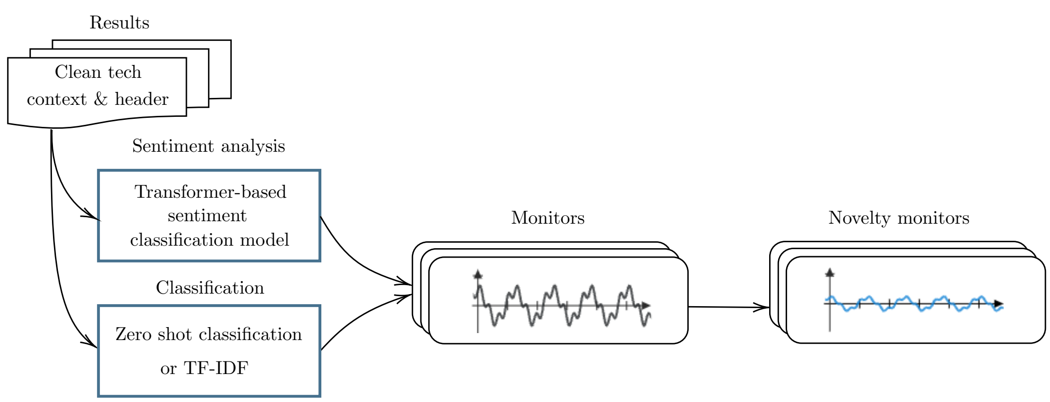
Following the two retrieval layers, as illustrated in Figure 4, we apply the reader which is the second core component of the question-answering process. The reader extracts potential answers from the retrieved chunks of news, relying on the transformer-based architecture. The reader assigns a score to the extracted answer based on its understanding of the text and the query and provides the context in which the answer was derived. The model chosen for the reader is roBERTa-base, as it provides the best compromise between extraction accuracy and execution time (see Table 3 in Section 2).

At the end of the information extraction pipeline, we obtain the most relevant answers to our query "New clean energy tech ?" along with the associated contexts. The next step involves mapping these answers back to their corresponding segments within the document and, then, to the entire article.

## 3.3 Sentiment scoring and classification

Following the upstream novelty detection pipeline (described in Figure 4), which identifies new clean technologies and their context, we implement the downstream part of the pipeline, presented in Figure 6 to evaluate sentiment.

Figure 6: Downstream novelty detection pipeline: sentiment and classification




Source: Amundi Investment Institute.

As described in Van Dijk (1985), the headline and the first sentence of a news article can be considered as indicators of the most important information (for unbiased news). Since GDELT v3 decomposes article content into deciles, we pay special attention to the first decile of the article. We assume that this segment generally contains the headline and the lead statement of the news article. Therefore to construct the final sentiment score for a GDELT v3 news article, we average the score of the first decile and the identified contexts related to clean tech novelty. To decrease vulnerability to adversarial attacks (Leippold, 2023), we use sentiment classification based on transformer models that have been explicitly fine-tuned to measure tonality.

As described by Potts *et al.* (2020), we can use several benchmark datasets mostly based on posts from X - the platform formerly known as Twitter - or movie reviews to evaluate sentiment classification models (Zimbra *et al.*, 2018). Our case is specific as we extract top components of news and contexts that significantly differ from the typical benchmark phrases commonly used in sentiment analysis. Thereby, we chose to refine the evaluation of sentiment models and developed a custom benchmark dataset consisting of approximately 150 paragraphs and news headers discussing clean technologies: the Sample-CTD-TONE. These elements were then manually assigned to a positive, negative, or neutral sentiment by four investment professionals, based on their perception of the text. The final sentiment was derived from the average of the scores given by all four professionals. To select our model, we performed sentiment analysis on our dataset using three different classification models and calculated their corresponding scores. The results are displayed in Table 6.

Table 6: Performance of sentiment analysis models on our benchmark dataset

| Sentiment Analysis models | Accuracy | Precision (Avg) | Recall (Avg) | F1-score (Avg) | Pos/Neg Gap |
|---|---|---|---|---|---|
| sentiment_tweet_roBERTa | 0.63 | 0.66 | 0.61 | 0.63 | 2.7% |
| sentiment_tweet_xlm_roBERTa | 0.57 | 0.60 | 0.58 | 0.58 | 3.4% |
| sentiment_finBERT | 0.52 | 0.56 | 0.50 | 0.50 | 5.4% |

Source: Authors' calculations, Amundi Investment Institute.

These models show the probability that the news segment analyzed is either positive, negative or neutral. We project these probabilities back into a sentiment score between -100 and 100 using the following methodology:

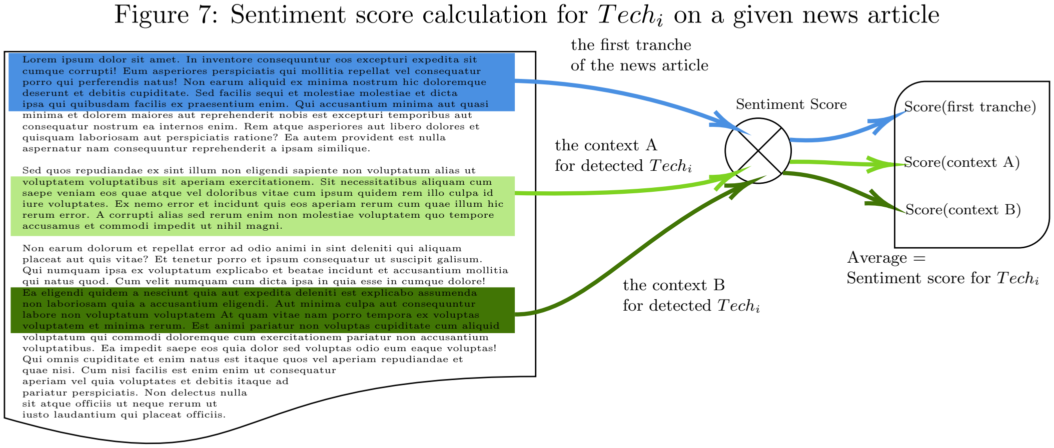
$$Score(N) = \frac{100}{3}((1 + 2p_{N\in C_1})\delta_1 - (1 + 2p_{N\in C_2})\delta_2 + p_{N\in C_3} sign(p_{N\in C_1} - p_{N\in C_2})\delta_3) \quad (1)$$

Where :

- $N$: the news segment from GDELT v3 on which we are evaluating the sentiment score.
- $C_1$, $C_2$ and $C_3$: represent the positive, negative and neutral class respectively.
- $\delta_i$: is equal to 1 if class i has the highest probability and 0 otherwise.

We choose the fine-tuned roBERTa model for tweet sentiment analysis (Barbieri *et al.*, 2020) because of its superior performance on our test dataset. We assign a sentiment score to each technology detected by our system. These results detail the names of the technologies, the context in which they were extracted, the first tranche of the news article, and the sentiment score associated with each technology in the news. We illustrate this step in Figure 7.

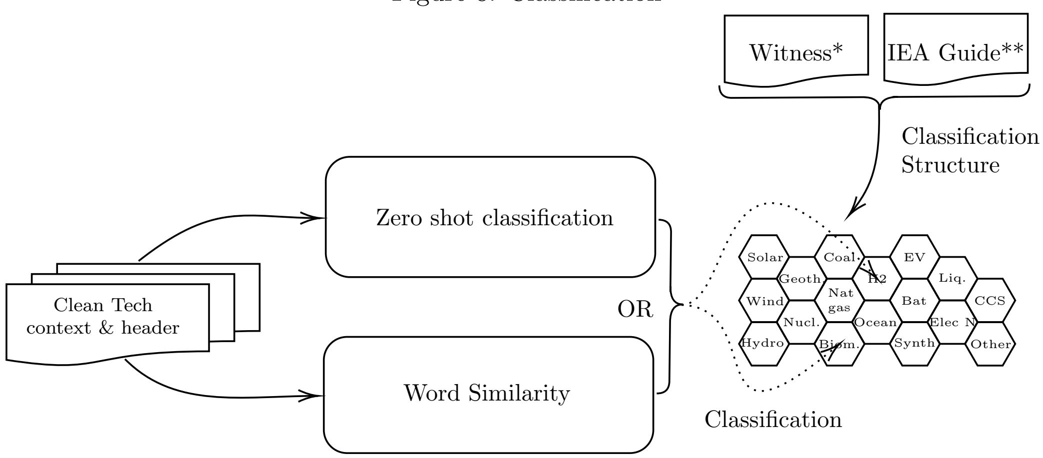
Figure 7: Sentiment score calculation for $Tech_i$ on a given news article




Source: Amundi Investment Institute.

We then create clean tech classes. Defining categories produces denser signals than an approach per individual clean tech. We use the technology mapping provided in the IEA's clean energy technology guide (IEA, 2022a) as well as the energy models of the "Witness" energy models (Gauthier *et al.*, 2022) to create 16 clean tech categories. These categories are: "solar","wind", "hydropower", "geothermal", "nuclear", "coal", "natural gas", "biomass", "hydrogen", "ocean and tidal", "electric vehicles", "battery and storage", "synthesis gas", "liquid fuels", "electricity networks", "carbon capture and storage CCS". As we explicitly intend to capture novelty, we add an "other" class to capture the emergence of a new technology that would not fit into the existing categories.

Figure 8: Classification



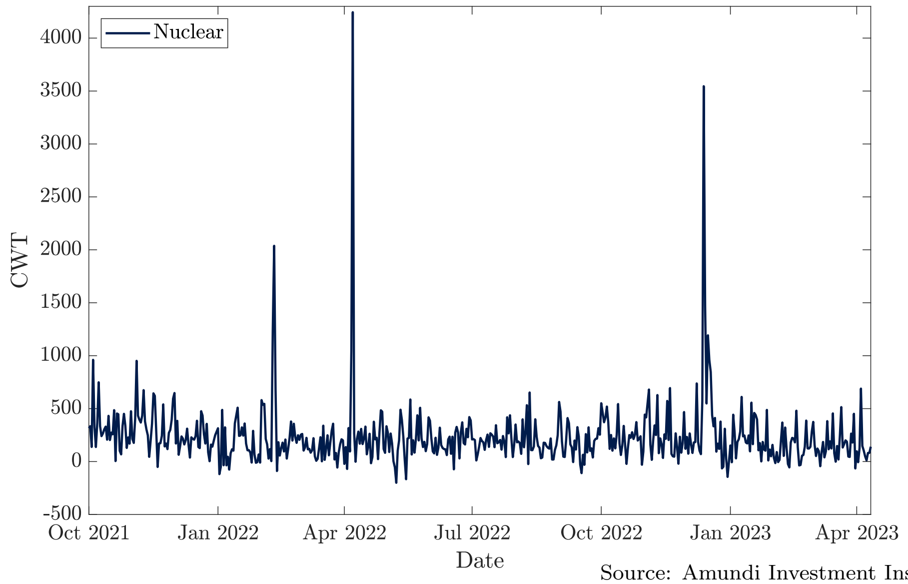

Note: *: Gauthier *et al.* (2022), **:IEA (2022a)
Source: Amundi Investment Institute.

Once technologies are detected, we use two methods to classify them. The first approach involves a multi-label classification of the context and name of the detected clean tech across the 16 previously defined categories. We use Facebook's bart-large-mnli zero-shot classification model (Yin *et al.*, 2019). We could have trained a Support Vector Machine type supervised classification model (Joachims, 1998), but the lack of labeled and multi-labeled data in our case study led us to choose zero-shot classification models. These models take classes as input and compute the probability of belonging to each class without requiring any specific training. For our second method, we follow the world similarity approach described in Figure 8. Word similarity allows us to separately classify the well-identified technologies into their respective categories. By combining both methods (with a union rather than an intersect approach), we can increase the likelihood of accurately classifying new technologies into their categories and reclassifying the known technologies into their relevant categories.

## 3.4 Monitoring and novelty detection

The final step in our exercise is to monitor clean technology developments and potential innovation (or novelty) in the field. In the spirit of Blanqué *et al.* (2022), who measure the Count-Weighted Tone for different narratives, we propose a similar metric but here calculated on a narrower - detected - set of news, related to clean technologies. We therefore introduce the Detected Count Weighted Tone(t) (DCWT(t)). This metric is the sentiment $\tau$(t) associated with the news articles identified by our novelty detection pipeline weighted by the number of identified news articles $v(t)$. An illustration of the DCWT for the "nuclear" clean tech class, for the period October 2021 - April 2023, is provided in Figure 9.

Figure 9: Detected Count-Weighted Tone - Nuclear




Source: Amundi Investment Institute.

Monitoring the DCWT for the different technologies could allow us to track pressures in minerals markets. For instance, wind power (deriving from wind turbines) and the manufacturing of electric vehicles (through permanent magnet motors for instance) are highly dependent on REEs, such as neodymium.

On top of monitoring, our framework provides novelty detection. In order to robustly identify local peaks in our Detected Count-Weighted Tone time series, we propose to use two metrics. On the one hand, we simply measure whether the most recent DCWT(t) is higher than the 15-day moving average and a multiple (namely 3) of the standard deviation. On the other hand, we use the Kullback and Leibler (1951) (KL) divergence, which measures the novelty of information for each previously defined clean tech class using the following formula:

$$D_{KL}(p||q) = \sum_{bins} p(x) \log \left( \frac{p(x)}{q(x)} \right) dx \tag{2}$$

The KL divergence is a metric that reflects the dissimilarity between two probability distributions. For each date $t$, and for each clean tech class, we compute a distribution $p(x)$ corresponding to the distribution of DCWT for a 15-day window. The distribution $q(x)$ is $p(x)_{t-1}$, or the 15-day distribution $p(x)$ for the same clean tech class on the previous day. To be comparable, the two density distributions are built on the same bin edges. The KL divergence provides a metric per bin width and is, by definition, sensitive to zero values. We then sum the KL divergences per bin.

The results for the nuclear clean tech class over our sample period are shown in Table 7. To distinguish the monitoring from the true innovation (novelty) we follow a two-step approach. First and for robustness reasons, we assume that to qualify as a true spike, a date must be identified as such by the two distance metrics we propose ($3\sigma$ and KL). We then retrieve the detected clean tech for the given date to check if it contains novelty. We consider that the novelty in a clean tech class is viral if it was detected as a novelty on the spike date and consecutive subsequent days. In our case, if the most viral clean tech is simply the generic term "nuclear", we disregard this spike since as it does not reflect true novelty. April

Table 7: Novelty detection - candidate spikes - nuclear

| Date | $3\sigma$ | KL | Virality | Novelty |
|---|---|---|---|---|
| 2021-11-04 | | ✓ | nuclear hydrogen | ✓ |
| 2022-02-09 | | ✓ | nuclear fusion technology | ✓ |
| 2022-02-10 | ✓ | ✓ | nuclear fusion | ✓ |
| 2022-04-06 | ✓ | ✓ | nuclear | |
| 2022-04-07 | ✓ | ✓ | nuclear | |
| 2022-11-19 | | ✓ | nuclear | |
| 2022-12-08 | | ✓ | hydrogen and advanced nuclear | ✓ |
| 2022-12-13 | ✓ | ✓ | nuclear fusion | ✓ |
| 2023-04-04 | | ✓ | nuclear | ✓ |

Source: Author's calculations, Amundi Investment Institute.

2022 is a good example of false novelty detection, when a spike in the series was not associated with a specific innovation, but with a generic term. Following this two-step process, two episodes qualify as true novelty detection. The first episode occurred in early February 2022. It is interesting to note that nuclear fusion technology had already been identified by the KL divergence the day before. In fact, the first mention of this episode can be traced back to the $5^{th}$ of February from our pipeline, but it did not turn into a jump in our monitoring, because the distance was too weak compared to historical values. This episode is certainly an echo of the progress made by the Joint European Torus (JET) laboratory on nuclear fusion, doubling the amount of energy produced from its 1997's record[7]. From our monitoring pipeline, we witnessed that nuclear fusion technology was prevalent for more than 10 days after this event. The second episode took place in December 2022, when the US Department of Energy announced another major breakthrough in fusion technology. Scientists at the Lawrence Livermore National Laboratory (LLNL) have succeeded in producing more energy from fusion reactions than was needed to start the process[8]. The impact of this event was greater, propelling nuclear fusion as the top viral technology for the next 19 days.

Compared to our previous analysis (Blanqué *et al.*, 2022; Cherief *et al.*, 2022), where we quantified the prominence of narratives and topics based solely on their volume and sentiment within the full spectrum of news, in this paper we disregard such "mass effect" by directly filtering the sources used upstream in our novelty detection pipeline. In fact, such a process reveals that clean tech novelty news does not yet attract significant virality within the mainstream media. We believe this may represent a compelling opportunity for investors in the short term, before it cascades into generalist news. This novelty detection building block provides insightful information on innovative technologies, in addition to our framework that monitor more general attention to clean tech developments.

# 4 Conclusion

Monitoring the development of clean technologies requires technical knowledge in many domains, such as electrical, electronic and mechanical engineering, signal processing, heat and fluid flow processing, robotics or control and systems engineering. The mainstream financial industry is typically not equipped with such expertise. In this paper, we test whether AI, and more specifically LLMs, can provide investors with insights into the transformation of this industry. Our first round of analysis focuses on benchmarking our information extrac-

[7] https://ccfe.ukaea.uk/fusion-energy-record-demonstrates-powerplant-future/.
[8] https://lasers.llnl.gov/news/nif-fusion-ignition-shot-hailed-as-historic-scientific-feat.

tion pipeline based on question-answering. We emphasize how NER models could apply to our research question if trained on the appropriate documents. The context extraction feature of the question-answering task allows us to set up a word similarity check. This purely syntactic layer - without complex LLMs involved - acts as a very effective "safety net" for the output of the alBERT model.The pipelines we propose allow us to efficiently replicate specific clean tech dependencies on minerals, as outlined by the IEA. We then test the accuracy of our framework in a two-step novelty identification exercise based on the GDELT dataset. Our application to the nuclear technology showcases the detection of major innovations in the field in 2022 - namely "nuclear fusion" - following major scientific breakthroughs. It also highlights the ability to track more general developments, related to the broader "nuclear" dimension or other clean tech, and therefore their demand for minerals. Our approach has proven to be very efficient in capturing niche topics from a mass media dataset such as GDELT, and transforming them into an insightful signal. Finally, the process we propose for clean technologies could be applied to other topics or industries. More generally, tools that synthesize expert knowledge and identify novelties within a given domain could be very valuable for both the financial industry and policymakers. Another direction could be to adapt our framework to gather information from a large number of documents, which could be very useful for analysts, for example.

## References


Balaram, V. (2019). *Rare earth elements: A review of applications, occurrence, exploration, analysis, recycling, and environmental impact. Geoscience Frontiers*, *10*(4), pp. pp. 1285–1303. https://doi.org/10.1016/j.gsf.2018.12.005.

Barbieri, F., Camacho-Collados, J., Espinosa Anke, L., & Neves, L. (2020). *TweetEval: Unified Benchmark and Comparative Evaluation for Tweet Classification. Findings of the Association for Computational Linguistics: EMNLP 2020*, pp. 1644–1650. https://doi.org/10.18653/v1/2020.findings-emnlp.148

Blanqué, P., Slimane, M. B., Cherief, A., Le Guenedal, T., Sekine, T., & Stagnol, L. (2022). *The Benefit of Narratives for Prediction of the S&P 500 Index. The Journal of Financial Data Science*, *4*(4), pp. 72–94. https://doi.org/10.3905/jfds.2022.1.107

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., et al. (2020). *Language models are few-shot learners. Advances in Neural Information Processing Systems*, *33*, pp.1877–1901.

Brunila, M., & LaViolette, J. (2022). *What Company do Words Keep? Revisiting the Distributional Semantics of J.R. Firth Zellig Harris.*

Bulatovic, S. M. (2015). *Handbook of Flotation Reagents: Chemistry, Theory and Practice Volume 3, Flotation of Industrial Minerals.*

Chen, L., Zaharia, M., & Zou, J. (2023). *How is ChatGPT's Behavior Changing over Time?*

Cherief, A., Sekine, T., & Stagnol, L. (2022). *The Market Effect of Acute Biodiversity Risk: the Case of Corporate Bonds. Available at SSRN 4288552.*

Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., Grave, E., Ott, M., Zettlemoyer, L., & Stoyanov, V. (2019). *Unsupervised Cross-lingual Representation Learning at Scale. arXiv preprint arXiv:1911.02116.*

Dalvi, N. N., Domingos, P. M., Mausam, Sanghai, S. K., & Verma, D. (2004). *Adversarial classification. Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* https://api.semanticscholar.org/CorpusID:14134514

Deng, M., Leippold, M., Wagner, A. F., & Wang, Q. (2022). *The Net-Zero Transition and Firm Value: Insights from the Russia-Ukraine War, REPowerEU, and the US Inflation Reduction Act. CEPR Discussion Papers.*

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv preprint arXiv:1810.04805.*

Eggert, R. G. (2011). *Minerals Go Critical. Nature Chemistry*, *3*(9), pp. 688–691.

European Commission. (2020). *Study on the EUs list of Critical Raw Materials - Final Report.*

European Commission. (2023). *Study on the Critical Raw Materials for the EU.*

Gauthier, B., Morere, M., Meaux, M., Barjhoux, P.-J., & Chevalier, T. (2022). *Complex System of Systems Simulation for Energy Transition.* https://www.witness4climate.org/witness-presentations/

Goutte, C., & Gaussier, E. (2005). *A Probabilistic Interpretation of Precision, Recall and F-score, with Implication for Evaluation. European Conference on Information Retrieval*, pp. 345–359.

Haddad, A. Z., Hackl, L., Akuzum, B., Pohlman, G., Magnan, J.-F., & Kostecki, R. (2023). *How to Make Lithium Extraction Cleaner, Faster and Cheaperin Six Steps. Nature*, *616*(7956), pp. 245–248. https://doi.org/10.1038/d41586-023-00978-2

Hendler, J. (2008). *Avoiding Another AI Winter. IEEE Intelligent Systems*, *23*(02), pp. 2–4.

IEA. (2021). *Patents and the Energy Transition.*

IEA. (2022a). *Energy Technology Perspectives - Clean Energy Technology Guide.*

IEA. (2022b). *The Role of Critical Minerals in Clean Energy Transitions, March 2022 Revised Version.* https://www.iea.org/reports/the-role-of-critical-minerals-in-clean-energy-transitions

IEA. (2023a). *CO2 Emissions in 2022.*

IEA. (2023b). *Critical Minerals Market Review 2023.*

IEA. (2023c). *Energy Technology Perspectives 2023.* https://www.iea.org/reports/energy-technology-perspectives-2023

IEA. (2023d). *Net Zero Roadmap A Global Pathway to Keep the 1.5C Goal in Reach.* https://www.iea.org/reports/net-zero-roadmap-a-global-pathway-to-keep-the-15-0c-goal-in-reach

IEA & Scott, R. (2004). *Origins of the IEA.* https://doi.org/{10.1787/9789264020931-3-en}

IPCC. (2011). *IPCC Special Report on Renewable Energy Sources and Climate Change Mitigation. Renewable Energy*, *20*(11).

IPCC. (2023). *Climate Change 2023: Synthesis Report. A Report of the Intergovernmental Panel on Climate Change. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change.*

Joachims, T. (1998). *Text Categorization with Support Vector Machines: Learning with Many Relevant Features. European Conference on Machine Learning*, pp. 137–142.

Karpukhin, V., OÄuz, B., Min, S., Lewis, P., Wu, L., Edunov, S., Chen, D., & Yih, W.-t. (2020). *Dense Passage Retrieval for Open-Domain Question Answering. arXiv preprint arXiv:2004.04906*.

Kim, A., Muhn, M., & Nikolaev, V. (2023a). *Bloated Disclosures: Can ChatGPT Help Investors Process Financial Information? arXiv preprint arXiv:2306.10224*.

Kim, A., Muhn, M., & Nikolaev, V. (2023b). *From Transcripts to Insights: Uncovering Corporate Risks Using Generative AI. arXiv preprint arXiv:2310.17721*.

Kullback, S., & Leibler, R. A. (1951). *On Information and Sufficiency. The Annals of Mathematical Statistics*, *22*(1), pp. 79–86.

Kurakin, A., Goodfellow, I., & Bengio, S. (2017). *Adversarial Machine Learning at Scale*.

Leetaru, K. (2021). *the Gdelt Project Blog*. https://blog.gdeltproject.org/realtime-global-massively-multilingual-horizon-scanning-using-gdelts-new-web-news-ngrams-3-0-dataset/

Leetaru, K., & Schrodt, P. A. (2013). *Gdelt: Global Data on Events, Location, and Tone, 1979–2012. ISA Annual Convention*, *2*(4), pp. 1–49.

Leippold, M. (2023). *Sentiment Spin: Attacking Financial Sentiment with GPT-3. Available at SSRN 4337182*.

Li, Y., Wang, S., Ding, H., & Chen, H. (2023). *Large Language Models in Finance: A Survey. Proceedings of the Fourth ACM International Conference on AI in Finance*, pp. 374–382.

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., & Stoyanov, V. (2019). *RoBERTa: A Robustly Optimized BERT Pretraining Approach. arXiv preprint arXiv:1907.11692*.

Loughran, T., & McDonald, B. (2011). *When is a Liability not a Liability? Textual Analysis, Dictionaries, and 10-Ks. The Journal of Finance*, *66*(1), pp. 35–65.

Martin, L., Muller, B., Ortiz Suàrez, P. J., Dupont, Y., Romary, L., de la Clergerie, E., Seddah, D., & Sagot, B. (2020). *CamemBERT: a Tasty French Language Model. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. https://doi.org/10.18653/v1/2020.acl-main.645

Moniz, E. J. (2015). *IEA Ministerial Meeting, Summary of the Chair*.

Nassar, N. T., Brainard, J., Gulley, A., Manley, R., Matos, G., Lederer, G., Bird, L. R., Pineault, D., Alonso, E., Gambogi, J., et al. (2020). *Evaluating the Min-*

*eral Commodity Supply Risk of the US Manufacturing Sector. Science Advances*, *6*(8). https://doi.org/10.1126/sciadv.aay8647

Papernot, N., McDaniel, P., Jha, S., Fredrikson, M., Celik, Z. B., & Swami, A. (2015). *The Limitations of Deep Learning in Adversarial Settings. arXiv e-prints*, arXiv–1511.

Pietsch, M., Möller, T., Kostic, B., Risch, J., Pippi, M., Jobanputra, M., Zanzottera, S., Cerza, S., Blagojevic, V., Stadelmann, T., Soni, T., & Lee, S. (2019). *Haystack: the End-to-end NLP Framework for Pragmatic Builders*. https://github.com/deepset-ai/haystack

Potts, C., Wu, Z., Geiger, A., & Kiela, D. (2020). *DynaSent: A Dynamic Benchmark for Sentiment Analysis. arXiv preprint arXiv:2012.15349*.

Ramos, J., et al. (2003). *Using Tf-idf to Determine Word Relevance in Document Queries. Proceedings of the First Instructional Conference on Machine Learning*, *242*(1), pp. 29–48.

Rasoulinezhad, E., & Taghizadeh-Hesary, F. (2022). *Role of Green Finance in Improving Energy Efficiency and Renewable Energy Development. Energy Efficiency*, *15*(2), pp. 14. https://doi.org/10.1007/s12053-022-10021-4

Reimers, N., & Gurevych, I. (2019). *Sentence-bert: Sentence Embeddings Using Siamese Bert-networks. arXiv preprint arXiv:1908.10084*.

Robertson, S., Zaragoza, H., et al. (2009). *The Probabilistic Relevance Framework: BM25 and Beyond. Foundations and Trends® in Information Retrieval*, *3*(4), pp. 333–389.

Rudolph, L., Beyeler, N., & Patel, L. (2022). *The Inflation Reduction Act - a Historic Piece of Climate and Health Legislation*.

Samreen, S., & Kausar, S. (2019). *Phosphorus Fertilizer: The Original and Commercial Sources*. In *Phosphorus - Recovery and Recycling*. IntechOpen. https://doi.org/10.5772/intechopen.82240

Sang, E. F., & De Meulder, F. (2003). *Introduction to the CoNLL-2003 Shared Task: Language-Independent Named Entity Recognition. Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL 2003*, 142–147. https://www.aclweb.org/anthology/W03-0419

Syed, Z. H., Trabelsi, A., Helbert, E., Bailleau, V., & Muths, C. (2021). *Question Answering Chatbot for Troubleshooting Queries based on Transfer Learning. Procedia Computer Science*, *192*, pp. 941–950.

UNFCCC, S. (2023). *Technical Dialogue of the First Global Stocktake. Synthesis Report by the co-facilitators on the Technical Dialogue. Synthesis reports - Conference UN Climate Change Conference - United Arab Emirates Nov/Dec 2023*.

US Geological Survey. (2023). *Mineral Commodity Summaries 2023*.

US Senate Committee on Energy and Natural Ressources. (2021). *Energy Act of 2020, Section-by-Section.*

Vaghefi, S., Wang, Q., Muccione, V., Ni, J., Kraus, M., Bingler, J., Schimanski, T., Colesanti Senni, C., Webersinke, N., Huggel, C., & Leippold, M. (2023). *ChatClimate: Grounding Conversational AI in Climate Science. Available at SSRN 4414628.*

Van Dijk, T. A. (1985). *Structures of News in the Press. Discourse and Communication: New Approaches to the Analysis of Mass Media Discourse and Communication*, *10*, pp. 69.

Yin, W., Hay, J., & Roth, D. (2019). *Benchmarking Zero-shot Text Classification: Datasets, Evaluation and Entailment Approach.*

Zhu, F., Lei, W., Wang, C., Zheng, J., Poria, S., & Chua, T.-S. (2021). *Retrieving and Reading: A Comprehensive Survey on Open-domain Question Answering.*

Zimbra, D., Abbasi, A., Zeng, D., & Chen, H. (2018). *The state-of-the-art in Twitter Sentiment Analysis: A Review and Benchmark Evaluation. ACM Transactions on Management Information Systems (TMIS)*, *9*(2), pp. 1–29.

# A Appendix

## A.1 Retriever models

Table 8: Performance of information extraction with different retriever models

| Retriever Model | Precision | Recall | F1-score | Execution time of the model |
|---|---|---|---|---|
| Facebook$^{\star}$ | 0.76 | 0.80 | 0.78 | 98 minutes and 27 seconds |
| mpnet$^{\star\star}$ | 0.59 | 0.95 | 0.73 | 98 minutes and 16 seconds |
| sentence transformers$^{\star\star\star}$ | 0.65 | 0.85 | 0.74 | 113 minutes and 47 seconds |

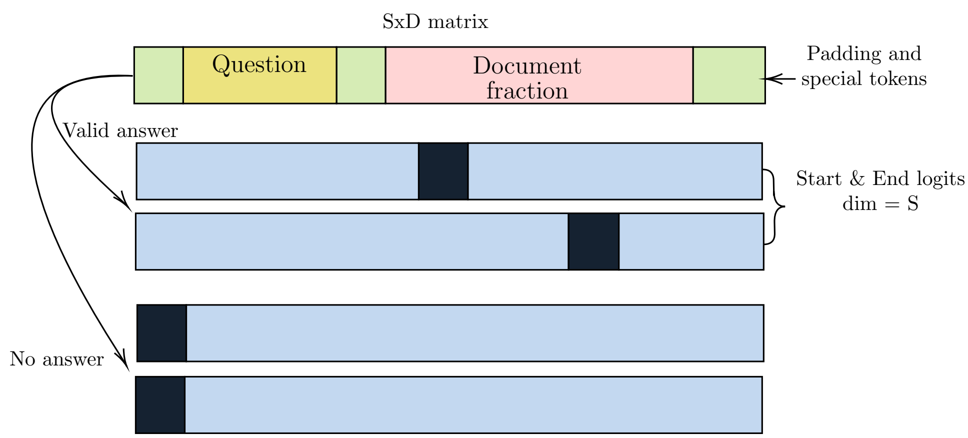
Source: Author's calculations, Amundi Investment Institute.

Note: $\star$: query_embedding_model="Facebook/dpr-question_encoder-single-nq-base", passage_embedding_model="Facebook/dpr-ctx_encoder-single-nq-base", $\star\star$: multi-qa-mpnet-base-dot-v1 , $\star\star\star$: sentence-transformers/LaBSE. The F1-score is the harmonic mean of Precision and Recall with Precision and Recall defined as :

$$Precision = \frac{True\,Positive}{True\,Positive + False\,Positive},\ Recall = \frac{True\,Positive}{True\,Positive + False\,Negative}$$

## A.2 Question-Answering

Figure 10: Answer extraction

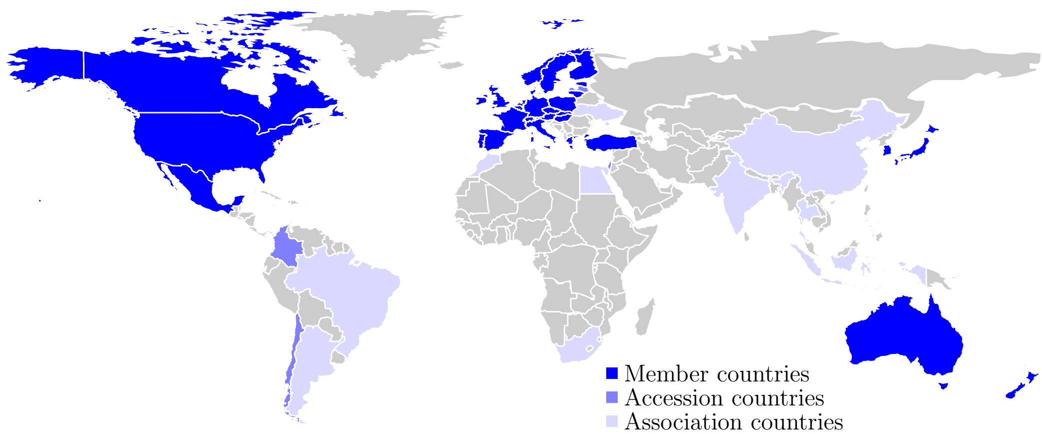



Source: https://www.deepset.ai/blog/modern-question-answering-systems-explained, Amundi Investment Institute.

## A.3 IEA history, critical minerals and materials

The IEA was founded in 1974 as an autonomous agency within the framework of the OECD as a means for "industrialised" countries to organize energy security in the context of the emerging political power of Middle Eastern oil producers. Membership to IEA is conditioned to membership to the OECD. In the narration of the first twenty years of the history of IEA, IEA and Scott (2004) also indicates the collaboration with non-member countries (NMC). This collaboration evolved in 2015 ahead of the COP21 together with the awareness of the "close relationship between energy and climate change" (Moniz, 2015). In the new reality of contributing to the fight against climate change, the IEA affirmed its modernisation objectives with three pillars: "enhanced engagement with major emerging economies, strengthened and broadened commitment to energy security, and greater focus on clean energy technology, including energy efficiency". To highlight this change relative to NMC, the IEA created the Association framework. A. At the time of writing this paper[9], the IEA has 31 Member countries, 4 Accession countries and 11 Association countries as illustrated in Figure 11.

Figure 11: IEA countries




Source: IEA, https://gitlab.com/conradolandia/WorldMap-Tikz, Amundi Institute.

[9]retrieved from https://www.iea.org/about/membership on June $14^{th}$, 2023.

## A.4 Critical minerals and materials

Table 9: Critical Raw Materials and Minerals

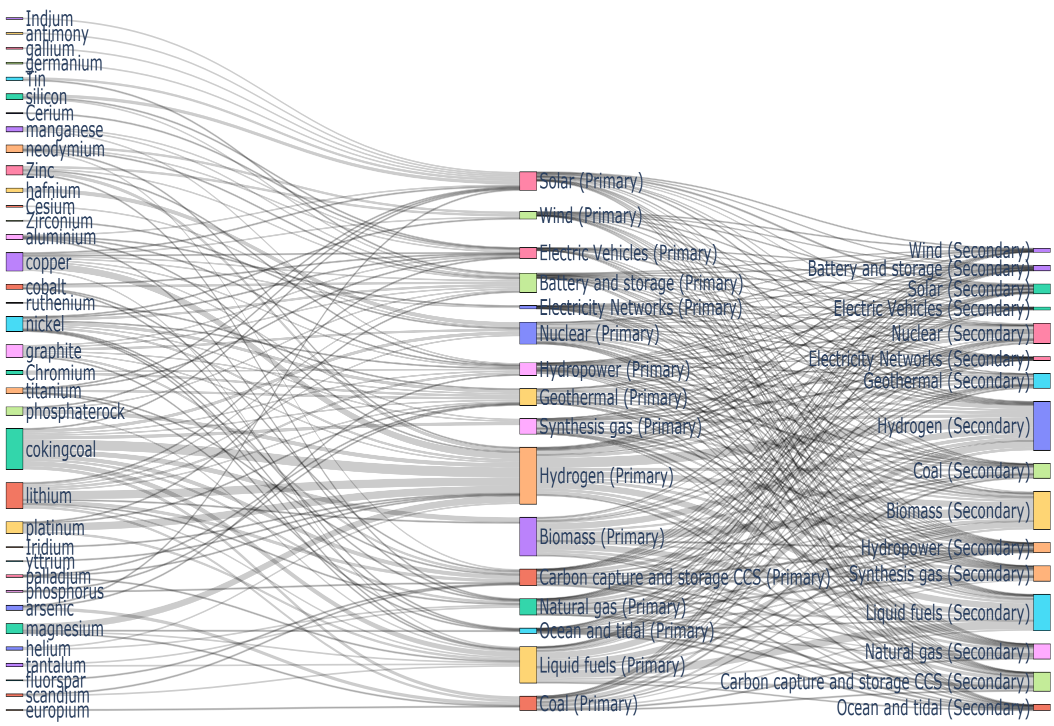
| | Critical Raw Materials for the EU | Critical Minerals for the US | Group desc | Symbol | Atomic number (Comment) |
|---|---|---|---|---|---|
| Aluminium/Bauxite | ✓ | ✓ | Other non-ferrous metals | Al | 13 |
| Antimony | ✓ | ✓ | Other non-ferrous metals | Sb | 51 |
| Arsenic | ✓ | ✓ | Other non-ferrous metals | As | 33 |
| Baryte | ✓ | ✓ | Industrial and construction minerals | BaSO4 | - (Barium sulfate) |
| Beryllium | ✓ | ✓ | Other non-ferrous metals | Be | 4 |
| Bismuth | ✓ | ✓ | Other non-ferrous metals | Bi | 83 |
| Boron/Borate | ✓ | | Industrial and construction minerals | B | 5 |
| Cerium | ✓ | ✓ | Light Rare Earth Elements | Ce | 58 |
| Cesium | | ✓ | | Cs | 55 |
| Chromium | | ✓ | Iron and ferro-alloy metals | Cr | 24 |
| Cobalt | ✓ | ✓ | Iron and ferro-alloy metals | Co | 27 |
| Coking Coal | ✓ | | Bio and other materials | Bituminous rank<br>(C content in 76-86% range) | - |
| Copper | (✓) | | Strategic (not Critical) Raw Materials<br>Other non-ferrous metals | Cu | 29 |
| Dysprosium | ✓ | ✓ | Heavy Rare Earth Elements | Dy | 66 |
| Erbium | ✓ | ✓ | Heavy Rare Earth Elements | Er | 68 |
| Europium | ✓ | ✓ | Heavy Rare Earth Elements | Eu | 63 |
| Feldspar | ✓ | | Industrial and construction minerals | Albite: ($Na_2O$, $Al_2O_3$, $6SiO_2$)<br>Microcline: ($K_2O$, $Al_2O_3$, $6SiO_2$) | - (Solid solution) |
| Fluorspar | ✓ | ✓ | Industrial and construction minerals | CaF2 | - |
| Gadolinium | ✓ | ✓ | Heavy Rare Earth Elements | Gd | 64 |
| Gallium | ✓ | ✓ | Other non-ferrous metals | Ga | 31 |
| Germanium | ✓ | ✓ | Other non-ferrous metals | Ge | 32 |
| Graphite (Natural) | ✓ | ✓ | Industrial and construction minerals | C | 6 |
| Hafnium | ✓ | ✓ | Other non-ferrous metals | Hf | 72 |
| Helium | ✓ | | Bio and other materials | He | 2 |
| Holmium | ✓ | ✓ | Heavy Rare Earth Elements | Ho | 67 |
| Indium | | ✓ | Other non-ferrous metals | In | 49 |
| Iridium | ✓ | ✓ | Platinum Group Metals | Ir | 77 |
| Lanthanum | ✓ | ✓ | Light Rare Earth Elements | La | 57 |
| Lithium | ✓ | ✓ | Other non-ferrous metals | Li | 3 |
| Lutetium | ✓ | ✓ | Heavy Rare Earth Elements | Lu | 71 |
| Magnesium | ✓ | ✓ | Other non-ferrous metals | Mg | 12 |
| Manganese | ✓ | ✓ | Iron and ferro-alloy metals | Mn | 25 |
| Neodymium | ✓ | ✓ | Light Rare Earth Elements | Nd | 60 |
| Nickel | (✓) | ✓ | Iron and ferro-alloy metals<br>Strategic (not Critical) Raw Materials | Ni | 28 |
| Niobium | ✓ | ✓ | Iron and ferro-alloy metals | Nb | 41 |
| Palladium | ✓ | ✓ | Platinum Group Metals | Pd | 46 |
| Phosphate Rock | ✓ | | Industrial and construction minerals | $Ca_1 0(PO_4)6(X)_2$<br>where X is $F_â$, $OH_â$ or $Cl_â$ | - |
| Phosphorus | ✓ | | Industrial and construction minerals | P | 15 |
| Platinum | ✓ | ✓ | Platinum Group Metals | Pt | 78 |
| Praseodymium | ✓ | ✓ | Light Rare Earth Elements | Pr | 59 |
| Rhodium | ✓ | ✓ | Platinum Group Metals | Rh | 45 |
| Rubidium | | ✓ | | Rb | 37 |
| Ruthenium | ✓ | ✓ | Platinum Group Metals | Ru | 44 |
| Samarium | ✓ | ✓ | Light Rare Earth Elements | Sm | 62 |
| Scandium | ✓ | ✓ | not considered REE | Sc | 21 |
| Silicon (metal) | ✓ | | Other non-ferrous metals | Si | 14 |
| Strontium | ✓ | | Other non-ferrous metals | Sr | 38 |
| Tantalum | ✓ | ✓ | Iron and ferro-alloy metals | Ta | 73 |
| Tellurium | | ✓ | Other non-ferrous metals | Te | 52 |
| Terbium | ✓ | ✓ | Heavy Rare Earth Elements | Tb | 65 |
| Thulium | ✓ | ✓ | Heavy Rare Earth Elements | Tm | 69 |
| Tin | | ✓ | Other non-ferrous metals | Sn | 50 |
| Titanium | ✓ | ✓ | Iron and ferro-alloy metals | Ti | 22 |
| Tungsten | ✓ | ✓ | Iron and ferro-alloy metals | W | 74 |
| Vanadium | ✓ | ✓ | Iron and ferro-alloy metals | V | 23 |
| Ytterbium | ✓ | ✓ | Heavy Rare Earth Elements | Yb | 70 |
| Yttrium | ✓ | ✓ | Heavy Rare Earth Elements | Y | 39 |
| Zinc | | ✓ | Other non-ferrous metals | Zn | 30 |
| Zirconium | | ✓ | Other non-ferrous metals | Zr | 40 |

Source: European Commission (2023), US Senate Committee on Energy and Natural Ressources (2021), Bulatovic (2015), Samreen and Kausar (2019), https://energyeducation.ca/encyclopedia/Bituminous_coal.

## A.5 Knowledge graph: Clean tech classes vs critical raw materials

In this subsection, following the same structure as Table 2, we construct a knowledge graph of clean technologies and their raw material requirements, with a focus on critical materials. We constructed our knowledge graph by first taking the list of clean technologies from Witness. This list has already been manually classified into our own categories of clean technologies. There are clean technologies that depend on other clean technologies. To visualize this interdependence in the clean tech landscape, we reapplied our discovery method and formulated questions to confront these technologies with each other.

Figure 12: Minerals to clean tech




Source: GDELT, Amundi Investment Institute.

## Important Information



Date of first use:  13 December 2023.

Document issued by Amundi Asset Management, "société par actions simplifiée"- SAS with a capital of €1,143,615,555 - Portfolio manager regulated by the AMF under number GP04000036 – Head office: 91-93 boulevard Pasteur – 75015 Paris – France – 437 574 452 RCS Paris – www.amundi.com

Photo credit: iStock by Getty Images - monsitj