Investigating fossil-fuel industry funded climate research

Scientometric analysis of industry funded research as compared with independent research

Thesis Project

Lukas Ciunaitis

Delft University of Technology

Investigating fossil-fuel industry funded climate research

Scientometric analysis of industry funded research as compared with independent research

by

Lukas Ciunaitis

Student Name	Student Number
Lukas Ciunaitis	4662431

Thesis Chairman:	Prof. dr. ir. Genserik Reniers
First Supervisor:	Dr. Oscar Oviedo-Trespalacios
Second Supervisor:	Dr. Nihit Goyal
Project Duration:	March, 2024 - July, 2024
Faculty:	Faculty of Technology & Policy Management, Delft

Cover:	Oil Rig Free Domain image from RawPixel
Style:	TU Delft Report Style, with modifications by Daan Zwaneveld

TU Delft

Preface

This thesis represents the culmination of my research at the Faculty of Technology & Policy Management at Delft University of Technology. The motivation behind this work stems from a deep concern for the integrity of climate science and the urgent need to address the biases introduced by fossil fuel industry funding. Understanding how industry lobbying influences scientific research is crucial for ensuring that policy decisions are based on unbiased and accurate information.

I would like to extend my sincere gratitude to my thesis chairman, Prof. Dr. Ir. Genserik Reniers, to my first supervisor, Dr. Oscar Oviedo-Trespalacios, and my second supervisor, Dr. Nihit Goyal, whose guidance and support were invaluable throughout this project. Their insightful feedback and encouragement made this work fruitful and meaningful throughout this entire process.

This research would not have been possible without the support of my family and friends, who provided me with unwavering support and motivation. I am also grateful to my colleagues and peers at Delft University of Technology for their constructive discussions and camaraderie.

I hope this thesis contributes to the broader conversation about the role of industry in scientific research and inspires further inquiry into ensuring the integrity and transparency of climate science.

Lukas Ciunaitis
Delft, September 2024

Summary

The intersection of fossil fuel industry funding and climate change research is a contentious area, raising significant ethical and practical concerns. Fossil fuel companies have historically played a substantial role in funding research, which has sparked debates about potential biases in climate research due to financial ties. Understanding the extent and impact of this funding is crucial for ensuring unbiased and accurate information that forms the basis of climate policies.

This thesis investigates the terminology and methodology in climate research funded by the fossil fuel industry through a comprehensive scientometric analysis, as compared to general research and a pool of research from universities with ties to fossil fuel industry. The study aims to identify patterns of industry influence on scientific outcomes and the dissemination of climate-related information. Data is collected by extracting research paper abstracts and other relevant information from Scopus database over the period of 2014-2024, divided by subject areas and funding sources, leading to over a 100,000 papers analysed with 80% relevance score word maps produced in VOSViewer. Thus, employing methodologies such as keyword-based data collection, Natural Language Processing (NLP), and topic modeling, the research identifies significant patterns in studies funded by fossil fuel companies, such as lack of focus on ESG terms, increased focus on carbon capture and hydrogen technologies, and more technical analysis methods. To ensure the robustness of these findings, the study employs validation techniques, including cross-referencing with independent literature and applying advanced statistical methods to control for potential confounding factors via topic modelling. This multifaceted approach enhances the credibility of the results and provides a clearer understanding of the extent and nature of industry-induced biases in climate research.

The analysis spans multiple disciplines, including energy, environmental science, earth and planetary science, and engineering, revealing that industry funded research outcomes are often tied directly to fossil fuel sector sphere of interests, while finding no significant influence patterns in university research. The findings of lack of ESG terms and increased focus on specific climate-mitigation technologies underscore the need for increased ESG interest and diversity in the industry funding of climate research to ensure its integrity and objectivity.

The thesis concludes with a discussion of the implications of these findings for policy and practice. It emphasizes the need for greater transparency in the disclosure of research funding sources and calls for the implementation of stricter guidelines to mitigate bias in scientific research, or for fossil fuel industry to ensure a larger diversity in the research funding. Additionally, it offers recommendations for future research, highlighting the importance of independent funding to preserve the objectivity and integrity of climate science.

By shedding light on the intersection of industry interests and scientific research, this thesis contributes to the ongoing debate about the ethical implications of industry-funded research. It underscores the critical importance of unbiased, evidence-based decision-making in addressing global climate change and supports the development of policies that ensure the reliability of scientific knowledge.

Contents

Nomenclature

List of common abbreviations and symbols occurring in the report further or common across the results.

Abbreviations and Symbols

Abbreviation	Definition
NLP	Natural Language Processing
RCP	Representative Concentration Pathway
LDA	Latent Dirichlet Allocation
ESG	Environmental Social Governance
LNG	Liquified Natural Gas
CCS	Carbon Capture & Storage
CCA	Climate Change Agreement
COP	Conference of Parties (referring to climate summits)
CPA	Carbon Pricing Act
GHG	Greenhouse Gas
SCI	Science Citation Index
CO2	Carbon Dioxide
N2O	Nitrous Oxide
CH4	Methane
CNZ	Carbon Net Zero
SSP	Shared Socioeconomic Pathway
.csv	comma-spaced value (reffering to file format)

1

Introduction

If there is one problem that unites every country in the world, every stakeholder and every person on the planet, it is global warming, and the associated climate change - the grand challenge of our day. This well-known, talked about and alarming phenomenon represents one of the most pressing challenges facing our world today, manifesting through rising global temperatures, increasing frequency of extreme weather events, and severe disruptions to ecosystems and human livelihoods [39] [1]. The scientific consensus is clear: human activities, particularly the burning of fossil fuels, are the primary drivers of this phenomenon [10]. As the impacts of climate change become more pronounced, the urgency to mitigate its effects and adapt to new environmental realities intensifies. This necessitates robust, evidence-based policies and innovative solutions grounded in rigorous scientific research. However, the integrity of such research is crucial, as it forms the foundation for global and national strategies aimed at combating climate change.

When integrity of climate research is mentioned, one of the main issues arising is the question of influence by industries with interest in research results, namely the fossil fuel industry, which has been allocating funds towards research in many universities worldwide [33]. The intersection of fossil fuel industry funding and climate change research is a contentious area, raising significant ethical and practical concerns. Fossil fuel companies, which have a vested interest in the continuation of fossil fuel extraction and consumption, have historically played a substantial role in funding research, either directly by funding specific studies, or indirectly by funding universities [7]. While the exact reasoning behind such funding is speculative, it remains straightforward that industries tend to fund research in fields that affect their business operations. This relationship has sparked a debate about the potential biases introduced into climate research due to financial ties to the fossil fuel industry.

Critics of this relationship, both activists and those in research community, argue that such funding could lead to skewed research outcomes that downplay the severity of climate change or promote solutions favorable to the fossil fuel industry. Instances of these concerns include documented cases of misinformation campaigns, lobbying against climate policies, and attempts to influence public and scientific discourse [38]. The need to scrutinize the extent and impact of fossil fuel industry involvement in climate change research is vital, as it holds significant implications for the integrity of the scientific endeavor and the effectiveness of policy responses to climate change [20]. This thesis seeks to delve into this complex issue, examining whether research funded by fossil fuel interests diverges in its topics, findings and recommendations from independently funded studies, thus shedding light on the broader debate about the role of industry in shaping scientific knowledge and public policy.

While funding of research by fossil fuel industry may raise questions, it is a more nuanced and complex issue than a simple ”no funding” or ”lots of funding”. There exists a significant dispute in regards to the relationship between academia and the fossil fuel companies, which often fund the research [25], as various sides disagree on whether the phenomenon of such funding benefits the wider scientific community or not. Critics argue that it is an obvious case of lobbying to preserve the interests of those who fund the research, while the proponents welcome any funding towards research and hail it as a communication tool to link academia and the industry decision-makers[24]. Despite that, fossil fuel industry funding is present in academia, often for valid reasons, ranging from industry's interest in scientific development linked to their field, to lack of alternate funding sources for some researchers. Cases diverge from funding for specific researchers, to less transparent funding of universities, where largest fossil fuel industry companies spend millions in donations to universities around the globe [34]. Whether such funding goes to addressing the issue of climate change, via research into its causes and effects, or is merely used to research technical questions related to fossil fuel extraction, or other industry-related problems, remains unanswered.

While this debate continues, multiple higher education institutions, including the ones in the Netherlands, have already distanced themselves from the fossil fuel industry in conducting said research [45] [44], although it is not yet clear that any wrongdoing has been going on. This decision stems from the proven unethical practices of the industry, which have included misleading information, lobbying efforts in order to influence the decision-making processes, watering down or derailing of climate initiatives, unethical business practices for PR purposes, as well as outright concealment of vital information [30]. Despite this, little is known about whether funding of research by fossil fuel companies leads to shifted conclusions, and what effect this may have on policy decision-making. This propels an evidence-lacking debate of whether fossil fuel company involvement in research is beneficial, or hindering the process [24]. A great latest example of it is the media attention to COP28 summit in Dubai, headed by the president of a local oil extracting corporation, with significant efforts to prevent the phase-out of oil from happening. Similar choice of presiding bodies is due to happen at COP29 in Baku [26]. A lot of media attention focused on whether there are any malicious intentions behind such diplomacy, with little evidence to suggest how exactly does lobbying affect policy-making [19]. Such examples do however show a struggle of fossil-fuel lobby hindering the UN Sustainability Goals, and the problem of faulty research that can be used as a basis for policy decision-making process.

The following thesis aims to contribute to this debate by taking a look at research funded by the fossil fuel industry, comparing its differences, if any, to independently funded literature. Inspired by research on road safety differences in different countries [23], this paper aims to similarly uncover whether papers funded by specific agents differ in terminology or outright ignore a certain aspect of the problem. The methodology employed involves scientometrics and textual data analysis of records within academic databases, as well as an investigation into whether research is affected not only via direct funding, but also via indirect funding, for instance, to universities. The primary focus of the analysis is to identify embedded values, thematic elements, connections to sustainable development goals, and potential societal impact. Data for the analysis is sourced from keywords found in the acknowledgments of published manuscripts. The project utilizes a sociotechnical perspective on risk management as a framework for assessing the breadth of the literature funded by the fossil fuel industry.

The outcomes of this project have the potential to contribute to the debate on research funding from parties with potential conflict of interest, by finding out whether such funding produces different results from independent research, thus establishing a risk of bias. While a lot of research has been done to prove that industry does try to affect research by funding [47], little understanding is present about what methods and terminology are used in research funded by such corporations. Current research mostly states that corporate funding does affect the research, but contrary to explaining how exactly, it simply mentions suppressing unfavorable research. [36]. Moreover, most of that research focuses on biomedical industry influence, with little to no investigations performed on research funded by fossil fuel industry.

1.1. Literature Synthesis

It sounds very logical to claim that fossil fuel industry funding, like any industry funding research will produce biases that are positive for the entity paying for the studies. While scientific evidence seems to confirm it, studies relate mostly to areas different than fossil fuel industry. Over the last decade, various forms of research were performed to assess industry effects on research in their respective related fields, ranging from tobacco [8] or medical to energy related fields [17]. Analysis of the studies related to company funding on research in a diverse field of industries, ultimately has shown that such funding leads to either focusing on commercial applications or, in many cases, research that supports the sponsor's policy and/or efforts. The idea that research funded by fossil fuel companies will produce bias is thus well founded, though no specific study focused on fossil-fuel related bias.

While there were no specific studies on effects of fossil fuel funding on research, industry ties with academia are well-known and studied. Sharmina [40] analyzes the academia's ties to the energy sector, finding that academia is more likely to have financial ties to the fossil fuel industry than to the renewable energy sector. She also identifies that studies on the effect of such industry funding on research have so far focused on medicine and life's sciences, rather than on energy or climate related science fields. While circumstantial, the risks stemming from such fossil fuel industry involvement have also been assessed from various perspectives. Bonds [6] has analyzed the perspective of climate-related violence growth, and that fossil fuel industry's involvement in research of this field tends to ignore sociotechnical aspects, such as risk of violence as a consequence of climate change, though he established that no consensus is reached on how such involvement relates to the risks. Most research, however, analyzes the incentives of fossil-fuel industry to lobby any climate-related research and decision-making, analyzing, for instance, the stock price volatility of the industry, which results in increased payoffs of lobbying in recent years [30]. Such effort shows that fossil fuel industry has a monetary incentive in funding research, which does not relate directly to technical improvements of their processes, but rather of political decision-making. Nevertheless, once again it is clear that research on fossil-fuel ties with academia does not address the results of such ties on the academia's research, but rather focuses on incentives and risks that are posed by such involvement. Studies on how exactly this funding alters the results of academia, and therefore, benefits the industry, have so far been absent, as such details are elusive in proving.

Analysis of how funding affects the results is difficult, since it involves thousands of research papers, funded by various companies, in various aspects, and performed by various researchers. Nevertheless, tools have become available for such analysis in recent years. In 2015, Justin Farrell first used large-scale textual analysis to look into dynamics of dissemination of false or biased information, finding that large corporations are more likely to engage in such kinds of activities [16]. While he further recommended studies into this, in 2019 Farrell expanded his research by using Natural Language Processing tools to look into misinformation ties with private funding, largely focusing on the political dynamic in the United States [43]. This research was closely tied to climate change, as the latter became a hotly contested political topic in the last few years in the context of US elections. This research did not specifically address fossil fuel industry or academia, yet it was the first using NLP tools to gain insights into the dynamics of information spread in the field of climate research.

The idea of using NLP and scientometric analysis has shown remarkable results in other fields, such as road safety [23]. Similar to this thesis, scientometric analysis employed so called ”word-maps” and looked into differences between research in high- and low-income countries related to road safety, finding significant differences in a variety of terms, problems that are discussed and the related solutions. For instance, low-income countries tended to have more road accidents tied to terms like “alcohol”, and in general seemed to put more emphasis on user behavior, rather than quality of infrastructure and related incentives. This method has not been used on climate-related research before, as shown by Li et. al., compiling an overview of topics the method was employed on before [31]. For this reason, scientometric analysis was deemed an interesting approach to employ in this study, with sufficient knowledge gap present.

As such, literature synthesis shows a lack of focus on the results of fossil fuel ties with academia, and provides a potential way to look at the problem, using tools that have been used to analyze problems in other fields, but also in a closely related climate change misinformation problem. The papers mentioned thus far, and many more, thus serve as a basis for the further investigation as part of the proposed research.

1.2. Knowledge Gap

The knowledge gap in effects of climate research funding by the fossil fuel industry lies in the need for a comprehensive understanding of the specific shortcomings or potential distortions of research narrative that may arise from such funding. While there is growing awareness of the ethical concerns associated with industry-funded research, that in many other industries has lead to bias and skewed results [40], there is still a lack of in-depth exploration into the nuances of biases within this body of literature, especially within the climate industry [17]. Moreover, this lack of understanding leads to backlash, arguing that fossil-fuel funding is beneficial to the academia, rather than an issue, due to increased dialogue between industry and academia, as well as increased availability of funds for research [24]. Existing discussions emphasize the general need to distance academia from the fossil fuel sector to ensure research impartiality [42]. However, there is limited systematic analysis of the actual reasons to distance the academia from industry finding, be it potential biases embedded in climate research, the ways in which funding sources may shape research outcomes, and the extent to which these biases may impact policy decisions and public perception. There exists a risk of a systematic problem with certain aspects of research or consequences being omitted or overstated, as happened in pharmaceutical industry research [40]. Not only does this pose a risk of flawed policy-making, but it also leaves the dilemma of accepting fossil fuel funds unanswered, leaving a lack of clarity on whether to reject or adversely, encourage industry funding for research in this field. Any idea of establishing clear rules of funding, with redistribution and overseeing panels, would also require rigorous research and proof of any systematic errors occurring.

As such, the following report aims to address the knowledge gap of effects of industry funding on climate-related research, by looking into keywords and methodology employed in papers funded by such research, as well as taking a look into indirect funding via universities, attempting to see whether any differences are discernible on that level as well. The exact reasoning and methodology are presented further in Chapters 1.4 and 2.

1.3. Relevance to EPA Program

The topic of this paper aims to provide evidence towards a societal problem related to the grand challenge of climate change tackling decision-making. As mentioned in the introduction, there exists a debate on whether or not funding by the fossil fuel industry is harmful to research and decision-making related to climate change, yet there is a significant knowledge gap in understanding how exactly that funding affects research, if at all. Consequently, it is not known whether such funding should be encouraged, or discarded completely, which in turn has significant implications on interpretation and monetary funding of research in this vital field. The research involves technical analysis using Natural Language Processing (NLP) tools and scientometric analysis, and is intended to provide a large audience with evidence-based input to an ethical dilemma of fossil fuel industry involvement in any global warming mitigation decision-making, arguably one of the grand challenges of modern society. Research objectives are to contribute to this debate by establishing what differences exist, if any, in research funded by fossil fuel companies, as opposed to simply claiming that the funding itself is problematic. Consequently, the findings of this research can be used as policy advice on whether to discard, or encourage fossil fuel industry, or energy industry in general, to contribute to climate change related research. As such, the project is tied to the program's goal of using technology to address various socio-technical issues and the grand challenges of our day.

1.4. Research Question

This research aims to shed light on the various dynamics between fossil fuel industry funding and research methodologies and results in the realm of energy and environmental studies. The central research question seeks to identify whether significant patterns and differences introduced by such funding exist, prompting an exploration into the nuanced ways in which financial support from the fossil fuel industry may influence the choices made in conducting research. Using scientometric analysis, networks of key words and methods can be constructed, indicating possible differences between independent and funded data, which could be further analyzed to find any potential patterns occurring from such differences. For instance, gaps in a sense of missing terms can be identified, or the opposite, terms that occur in funded research much more frequently, whether it relates to purely technical terms, or sociotechnical ones (such as terms related to policy or economics). Following the example of [23], such word maps will be able to show if significant differences in different clusters of data exist, and as such, possibly reveal the missing or overrepresented data, indicating whether funded research aligns with non-funded counterpart. Initially, the analysis focuses on direct funding, which includes papers that have sponsor affiliation from one of the fossil fuel industry companies, followed by an analysis in indirect funding, via research from universities with ties to such industry.

Subsequently, the sub-questions delve deeper into the impact of direct industry funding on the selection of research methods and whether discernible differences exist in terminology between studies funded by the fossil fuel industry and independent research endeavors. For that, a large pool of paper abstracts will be analyzed with VOSviewer to construct clusters of key terms, revealing both methodology used and key problem identifications. Such knowledge structures are able to point out whether a certain group of papers omit certain nuances of a problem, or look into a problem from only a certain perspective. By addressing these questions, this study aims to contribute valuable insights into the complex interplay between funding sources and research practices in the critical domains of energy and the environment.

Having established differences in research funded by fossil fuel industry, an attempt is made to identify research differences in "indirect funding" as well, by creating a pile of research that is not directly funded by fossil fuel companies, but comes from institutions that have links to such funding, as established via previously mentioned whistle-blowing organizations such as OpenSecrets, as well as other journalistic investigations listed in Chapter 2. This third pile of data is then analyzed similarly to the previous two, with key term networks established to explain whether significant differences occur. This allows for comparison between direct and indirect funding, and any differences between them can be identified.

Finally, having analyzed differences in methodology and terminology of all pools of data, an attempt is made to identify any patterns occurring within differences, if those are proven to exist. In such case, patterns are presented via data and analysis as mentioned above, followed by a discussion on potential reasons for such differences, and suggestions for further studies and insights in order to identify those reasons with clarity.

It is important to note, that while this thesis is exploring potential patterns and differences occurring between various pools of research, it assumes no causality and no patterns existing unless proven otherwise. As such, reader should not perceive this research as means to attack or defame an industry, but rather an investigation into what subjects are of interest to those who decide where the research funding goes.

The main research question:

- What are the potential differences between fossil fuel industry funded university research on climate-change, and research without such funding, over the last decade?

The sub-questions:

- 2.1 Does funding from the fossil fuel industry impact the choice of research methods employed in studies related to energy and climate-related research, and how?
- 2.2 To what extent does the terminology used in studies funded by the fossil fuel industry differ from independent research?
- 2.3 To what extent does research with indirect funding via donations to universities differ compared to research with direct funding?
- 2.4 Can any patterns be identified, if they exist, occurring from potential differences in research methods and terminology?

2 Methodology

The following Chapter outlines the methods, tools and strategies used to perform research and analysis of the topic of this paper. First, choices of research pools to address research questions are discussed, followed by tools utilised to analyse the research pools.

In recent years, the intersection of scientific research and industry funding has garnered significant attention, particularly within the realm of climate change and environmental studies. The fossil fuel industry, a major contributor to greenhouse gas emissions and global warming, has been scrutinized for its influence on scientific research [25]. This influence raises important questions about potential discrepancies and the integrity of research funded by such companies. The overarching goal of this research is to delve into how fossil fuel industry-funded research compares to other pools of research, particularly in the areas related to climate science.

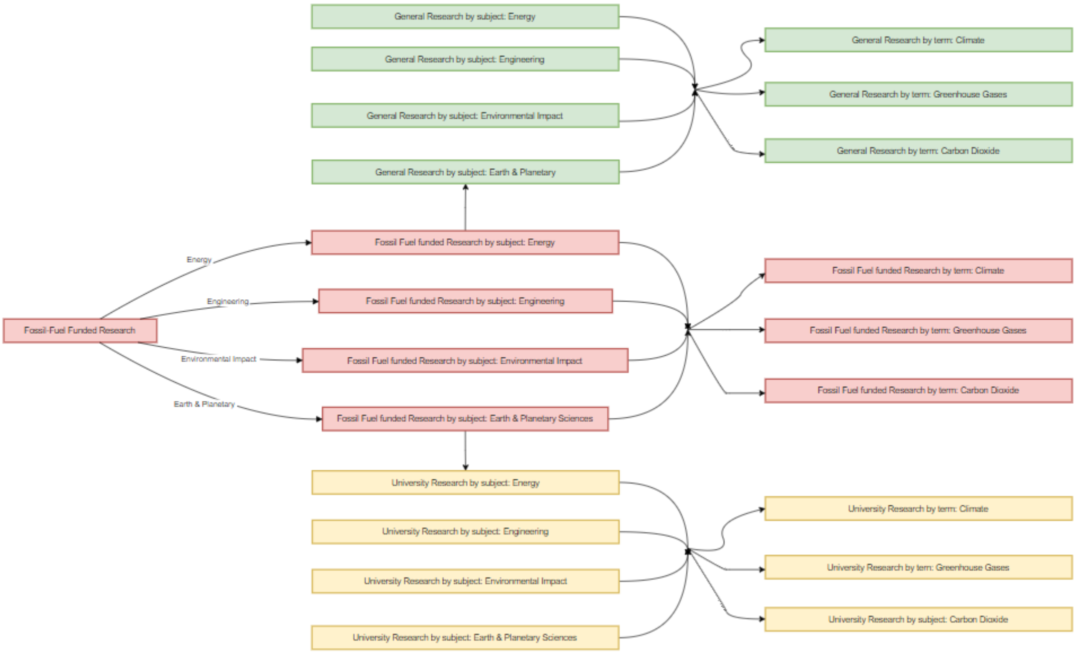
By comparing the research outputs funded by the fossil fuel industry with those from general academic sources and leading universities, this study aims to uncover any discernible differences or patterns that may exist. Such comparisons are essential to understand whether industry funding skews the focus, outcomes, or interpretations of scientific studies in ways that could impact public policy, regulatory decisions, and public perception of climate change.

To achieve this, a comprehensive methodology has been devised, encompassing the selection of data pools, specific research questions, and analytical tools. This includes identifying the most prominent fossil fuel companies and their funding patterns, selecting top universities known for their receipt of fossil fuel funding, and utilizing advanced tools like VOSViewer for scientometric mapping and analysis. The study also incorporates a temporal analysis to observe changes over time and employs validation strategies to ensure the robustness of the findings.

Figure 2.1 below presents a flowchart relating to the pools of data that were analysed as part of this research. At first, all papers that were funded by selected largest fossil fuel industry companies are analysed, in order to see most popular terms and topics, and to provide a general overview. Then, in order to compare fossil fuel industry funded research with other pools, research from specific four most popular among industry funding subject areas is selected: Energy, Engineering, Earth & Planetary Sciences and Environmental Sciences. Finally, in order to double check that research is indeed related to same topics, these pools have extra key terms added, in order to filter out any irrelevant research. These terms are: ”Climate change”, ”Greenhouse Gases” and ”Carbon Dioxide”.

Exact choices of companies, universities, terms and other nuances are presented in this chapter in sections below. It is important to note, that many of such selections are limiting the scope of research, and inclusion of dozens of other companies, universities and terms may result in more accurate results. However, most limitations are discussed in Chapter 5.

Figure 2.1: Flowchart of Research Pools Analysed as Part of this research. Note that all presented pools were in 2014-2024 time interval, and were further divided by two for temporal analysis.


General Research by subject: Energy
General Research by subject: Engineering
General Research by subject: Environmental Impact
General Research by subject: Earth & Planetary
General Research by term: Climate
General Research by term: Greenhouse Gases
General Research by term: Carbon Dioxide
Fossil-Fuel Funded Research
Energy
Engineering
Environmental Impact
Earth & Planetary
Fossil Fuel funded Research by subject: Energy
Fossil Fuel funded Research by subject: Engineering
Fossil Fuel funded Research by subject: Environmental Impact
Fossil Fuel funded Research by subject: Earth & Planetary Sciences
Fossil Fuel funded Research by term: Climate
Fossil Fuel funded Research by term: Greenhouse Gases
Fossil Fuel funded Research by term: Carbon Dioxide
University Research by subject: Energy
University Research by subject: Engineering
University Research by subject: Environmental Impact
University Research by subject: Earth & Planetary Sciences
University Research by term: Climate
University Research by term: Greenhouse Gases
University Research by subject: Carbon Dioxide

2.1. Data Collection

Due to availability of ”Funding Affiliation” information, research papers were retrieved from Scopus database, as such information was much more difficult to extract from Dimensions or any similar databases. As such, it was a straightforward process to single out research that had funding affiliation to specific companies or their affiliates, and it was possible to check which companies have sponsored the largest amount of papers.

The data, with keywords specified in sections below, was then extracted from Scopus with abstract and key information about authors, citations and paper name into a .csv file, which was then used in VOSViewer software to produce word maps, temporal analysis and density visualization maps, as expanded upon in chapter 3. Due to a very large amount of various words present, only those with 10+ occurrences were selected, and further filtered to have at least an 80 % relevance score in connection to other keywords. This way, the map had reduced the number of outliers visible and reduced the time needed to produce each map from almost two hours per map to around 30 minutes.

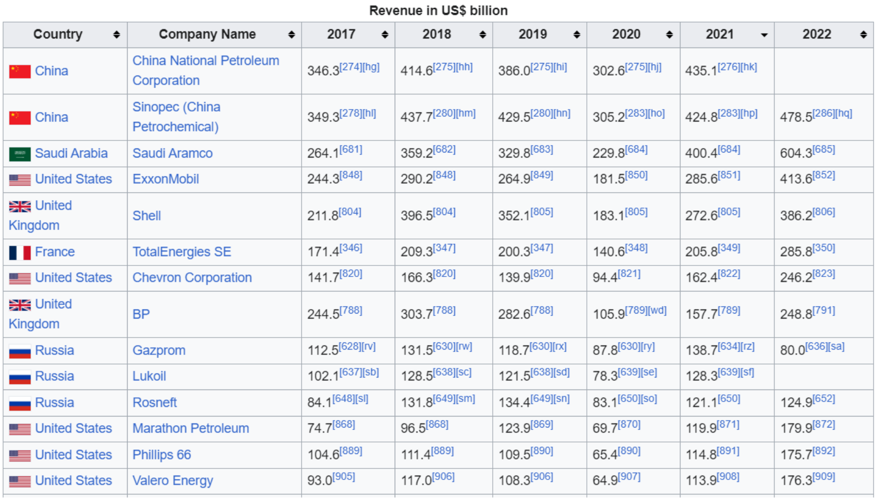
2.2. Fossil fuel industry Selection

The first important choice in research was the pool of fossil fuel industry companies that would be analysed in this research. The choice came down to research on latest news on the topic [11] [37] [12], as well as the amount of papers sponsored according to research database used. This combined effort has helped identifying the companies that sponsored largest amount of research papers, both as seen in Scopus, as well as companies most mentioned in media. Since ”Funding Affiliation” category in Scopus produces thousands of results, not only companies but also governmental and non-profit institutions, filtering out all fossil fuel industry companies became impossible, as database would simply not provide names of smallest institutions. Therefore, having identified what are the largest fossil fuel industry research sponsors, the available list in Scopus was checked, and largest companies were selected for the subsequent query of ” ”Shell” OR ”ENI” OR ... etc.” among funding sponsor. In total, as attempting to capture all of research by all of companies and their foundations was deemed too complex and fruitless, 8 companies were identified as by far the most frequently appearing funding sponsors.

In order to check this claim, various terms, namely ”oil”, ”climate”, ”fossil fuel”, ”coal”, ”natural gas”, ”lng” and ”carbon” were checked in the Scopus database, for each of these looking at the list of funding sponsors. Due to large amount of papers, only fundings sponsors that were mentioned in more than 50 papers were displayed, as Scopus was not displaying any institutions with a smaller number of papers funded, and the inclusion of those smaller institutions was deemed unnecessary to receive results. Therefore, 8 fossil fuel companies, primarily in oil industry, were identified, and included in this list.

As mentioned, it was noted that most research is funded by oil industry, as opposed to coal industry or natural gas. While potential explanations, such as research in English language that favours Western companies, which operate in many fields or shift away from coal, are possible, reader is invited to read about this in Chapter 4. For now, 8 largest companies were chosen: Shell, ENI, TotalEnergies, British Petroleum, Chevron, ExxonMobil, Saudi Aramco and Chinese National Petroleum Company (further shortened to CNPC). This list largely corresponds to largest fossil fuel (oil and gas) companies by revenue. as seen in the Figure 2.2 [48], though further research could well extend it to include many other companies, such as Equinor, SOCAR, Valero Energy, Gazprom, Sinopec or Petrobras (all companies with a 100 billion USD revenue per year). For the purpose of this research, the amount of papers sponsored by these companies was deemed too little (0-20 papers) to add significant value to the pool of papers funded by the companies included in the list. Furthermore, ENI is added to the list of fossil fuel lobby despite not being among the largest companies, due to large amount of papers available as well as reports on this company being involved in research funding. [27]

Figure 2.2: List of largest oil companies in the world according to revenue in 2021, Wikipedia



Revenue in US$ billion

Country	Company Name	2017	2018	2019	2020	2021	2022
China	China National Petroleum Corporation	346.3[274][hg]	414.6[275][hh]	386.0[275][hi]	302.6[275][hj]	435.1[276][hk]	
China	Sinopec (China Petrochemical)	349.3[278][hl]	437.7[280][hm]	429.5[280][hn]	305.2[283][ho]	424.8[283][hp]	478.5[286][hq]
Saudi Arabia	Saudi Aramco	264.1[681]	359.2[682]	329.8[683]	229.8[684]	400.4[684]	604.3[685]
United States	ExxonMobil	244.3[848]	290.2[848]	264.9[849]	181.5[850]	285.6[851]	413.6[852]
United Kingdom	Shell	211.8[804]	396.5[804]	352.1[805]	183.1[805]	272.6[805]	386.2[806]
France	TotalEnergies SE	171.4[346]	209.3[347]	200.3[347]	140.6[348]	205.8[349]	285.8[350]
United States	Chevron Corporation	141.7[820]	166.3[820]	139.9[820]	94.4[821]	162.4[822]	246.2[823]
United Kingdom	BP	244.5[788]	303.7[788]	282.6[788]	105.9[789][wd]	157.7[789]	248.8[791]
Russia	Gazprom	112.5[628][rv]	131.5[630][rw]	118.7[630][rx]	87.8[630][ry]	138.7[634][rz]	80.0[636][sa]
Russia	Lukoil	102.1[637][sb]	128.5[638][sc]	121.5[638][sd]	78.3[639][se]	128.3[639][sf]	
Russia	Rosneft	84.1[648][sl]	131.8[649][sm]	134.4[649][sn]	83.1[650][so]	121.1[650]	124.9[652]
United States	Marathon Petroleum	74.7[868]	96.5[868]	123.9[869]	69.7[870]	119.9[871]	179.9[872]
United States	Phillips 66	104.6[889]	111.4[889]	109.5[890]	65.4[890]	114.8[891]	175.7[892]
United States	Valero Energy	93.0[905]	117.0[906]	108.3[906]	64.9[907]	113.9[908]	176.3[909]

It needs to be noted that the scope of this research avoids any foundations that could potentially have indirect ties to fossil fuel industry, or political institutions (such as US Department of Energy), that may also have lobbying effects present. These, however, are incredibly difficult to establish and are thus not included.

2.3. Selection of lobbied universities

For the pool of data to investigate the lobbying effort of fossil fuel industry, whistle-blowing data and news are used to compile top universities accused of taking largest amount of donations from fossil fuel industry [11]. From these, looking at the number of research papers available in Scopus, 10 largest universities by amount of research papers available were singled out: University of California at Berkeley, Stanford University, Massachusetts Institute of Technology (MIT), George Mason University [41] [42] [29], University of Cambridge, Oxford University, The Norwegian University of Science and Technology (NTNU), ETH Zürich, Complutense University of Madrid and Imperial College London [33] [14]. It needs to be noted, that these may not necessarily be universities with largest total funding received, as such data is difficult to compile and was deemed out of scope of this research. While various research singles out other universities, the 10 chosen were mentioned multiple times, and have a sufficient amount of research papers available in order to answer the research sub-question 3. Moreover, in many instances these universities have produced significantly more research in any given time than fossil fuel industry sponsored pool, requiring random sampling or normalizing to achieve the results already. Research from these universities is thus singled out and compared to fossil fuel and general pools of research based on identical criteria, to establish whether any differences or patterns exist.

2.4. Search Strategy

With ”climate change” being a broad topic, results have a high risk of being unclear due to research pools being compared against something completely unrelated to them. For instance, simply comparing any papers funded by fossil fuel industry against any papers in general makes no sense, as taking absolutely random papers would include research areas absolutely unrelated to fields interesting for fossil fuel industry, let alone climate studies that are of interest for this research.

For this reason, various pools of data were extracted and tested before coming to the Flowchart 2.1 that we see at the beginning of this chapter. As an example, fossil fuel industry funded research was first analysed without any keywords, in order to record what kind of topics does the industry involve itself in. This way, research funded by fossil fuel industry itself consisted of 10,153 papers over the last decade. While some papers existed from the years before it, the numbers of available papers were much smaller compared to general research (total amount of papers funded by fossil fuel industry of all times was 13,192), therefore they were excluded as less representative of current research and its trends. After that, this research pool was further broken down to include terms such as ”climate”, ”carbon”, ”emissions” and many others. However, it was noted that such simple comparison produces a very small amount of papers, and is also difficult to compare to general research pool with vastly more papers involved. For this reason, break-down by subject areas was introduced.

Further research, likewise in general, lobbied university and fossil fuel pools, focused on top four distinct areas of research, together accounting for roughly half of all fossil fuel funded research: Environmental Science, Engineering, Energy and Earth & Planetary Sciences pools. In cases when further narrowing is needed, search terms are added that are related to climate research, such as: ”Climate”, ”Greenhouse gases”, ”carbon dioxide” etc. These serve as complimentary searches in order to identify and pinpoint patters, or include papers from other subject areas, with most analysis performed on broader searches in order to have a sufficient pool of fossil fuel funded data. All searches were performed in a span of a week, from April 29th to May the 3rd in 2024, and described further along with search numbers and results in chapter **??**, in chronological order.

Search queries would start with fossil fuel industry, putting a query ”Shell” OR ”ENI” OR ”TotalEnergies” OR ”British Petroleum” OR ”Chevron” OR ”ExxonMobil” OR ”Aramco” OR ”CNPC” in Funding Sponsor section, then filtering it by date, subject areas, and further by terms when needed. For general data, funding sponsor information would then be cleared and simple overview of subject areas, or terms would be added. Then, university pool would be checked by entering all universities mentioned above as query in ”Affiliation” section, followed by aforementioned filters. Due to a large amount of terms experimented with, those that are included in this analysis, along with their queries and amount of papers produced, are presented separately in Chapter **??**.

2.5. Data Analysis And Tools

The analysis of text, network and citation data was performed using VOSViewer [46] software tool, largely inspired by various scientometric papers [23] [22]. This was deemed sufficient and convenient as compared to other software tools or scripting, though the latter is done to validate the results of the research. The papers for VOSViewer were imported using .csv files exported from Scopus database. This database was chosen as opposed to Dimensions and Web of Science due to a larger amount of papers available and easily accessible funding information, which could be filtered by and exported to VOSViewer for further analysis. It needs to be noted that Scopus has a limit of exporting 20,000 papers at once. In cases where the pool of data was too broad, it was either narrowed down, or random sorting of data was performed to extract papers for analysis. For some of statistical figures, as well as validation, Python ntlk, gensim, pyLDavis and pandas libraries were used along with a few more standard ones, coded in Jupyter Notebook. Chat GPT LLM model was used to upload and analyse this paper for consistency and improvement suggestions.

Following the searches, each pool is analysed in VOSViewer to find out key topic clusters and most frequent terms in order to answer research questions about terminology and methodology, their average publication date in order to see if any patterns emerge over time, and citation relevance scores of their authors to check for potential patterns arising there, as well as additional Python scripts employed to compare the pools between each other, by for instance, finding common terms and their appearances normalised between the pools of data, or by comparing presence of selected terms. The last of these methods is used to answer all four research questions, since terminology and methodology answer the first three sub-questions, and pattern identification can also be considered with these results. The analysis is done starting with fossil fuel industry, then focusing on industry's specific subject areas, and comparing general and university research in those areas. The results of such analysis, its order and other important decisions are shown and discussed in Chapter 3.

Additionally, temporal analysis was performed by splitting data into two sections by date: 2014-2019 and 2020-2024 pools. Even within those, VOSViewer allows for analysis of temporal data in order to check for topic importance within a timespan, which served as additional analysis to gain insights into changes of patterns of research over time. While it was considered to perform a temporal analysis split by other dates, or by, for instance, significant events (such as Paris Climate Accords Agreement in the end of 2015), the amount of research in decades prior was deemed too small, and time frame considered deemed enough to be able to see any patterns occurring. Therefore, in order to have comparable sets of data, a split in the middle of the decade was chosen.

2.6. Validation strategy

As seen in road safety research of third-world countries by Haghani et. al. [23] [22] and many others, the method of scientometric mapping is used in many various sectors, and is already proven to be successful in establishing key term differences between various pools of data, acting as a validation of methodology. Validation of results, however, proves difficult due to lack of existing literature to compare it to.

For validation strategy, results can be compared with existing literature on biases in industry-funded research, such as those mentioned by Fabbri et. al [17]. While this is not a direct comparison, as no such literature exists on this specific topic, patterns from other industries may be examined for similarities and potential ways to improve the analysis, leading to a limited validation of the results.

Another validation strategy involves developing a Python script for topic modelling, which allows for scanning of pools of research and clustering them into several topic groups as well as its keywords, thus confirming the findings of VOSViewer software. This may additionally reveal whether patterns noted during software analysis are visible by other means of analysis, leading to validation of results.

3
Results

In this Chapter, analysis of research paper pools is presented. First, all of fossil fuel industry funded papers are analysed. Since it is impossible to compare this pool to a general pool due to its vast variability of research, pools are then divided into four key subject areas, which are most represented among fossil fuel industry funded research. These are compared to general and university research pools. Finally, additional term searches are performed to get a different angle of analysis, without division by subject areas, followed by explanation of citation and temporal analysis performed on each of the research pools. The order of research described here corresponds to chronological order of analysis, following a top-down approach, and then by additional explanation and analysis.

Each research pool is analysed by first producing the word map of all terms occurring in the research pool, with minimum of 10 occurrences per term, and minimum of 80 % relevance score to reduce computational cost and ensure relevance of the results. Clusters are then identified by their topics. The results are then analysed for their relevance score to see which clusters have higher relevance score or in other words, are most common in research. Finally, these results are analysed for the average publication time of terms, indicating trends of whether the terms are newly appearing, or otherwise, less likely to be mentioned with the progression of time. Unfortunately, the software analysis does not allow for identification of specific research papers among those analysed, but rather sums them up into a single piece of data. Before comparison to other research pools, top terms by their occurrence are extracted, and if necessary, pools are compared by the term occurrence or by largest difference between occurrence of terms. Each score is normalised to ensure that the difference in the amount of papers is not affecting the results. This way, the first two research sub questions are answered, before university pool analysis answers the third research subquestion. Final, fourth research subquestion, is discussed in Chapter 4 having analysed all the results. Finally, citation analysis is performed by taking a look into network of authors and their citations in VOSViewer, plotting them into a similar relevance map as displayed in Figure A.3.

This chapter shows the analysis and results, leading to conclusions that fossil fuel industry research has three main differences: lack of ESG terms, increased focus on some specific technologies, and less literature-analysis based methodology. University pools show no significant differences compared to general research. Citation analysis shows no significant patterns, while temporal analysis shows similar trends in both fossil fuel and general research pools.

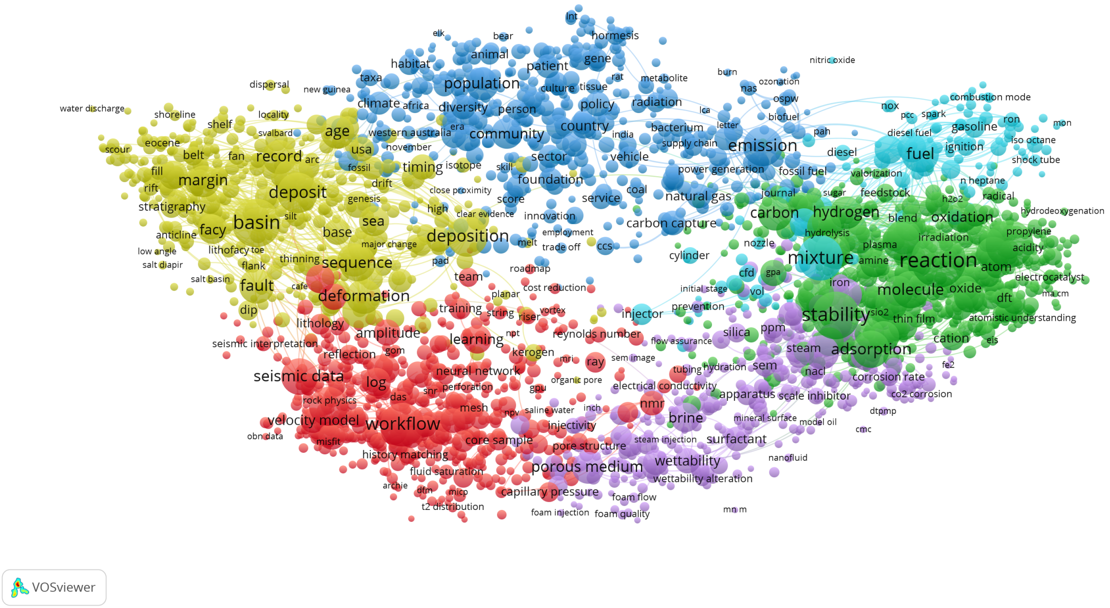
Important note is that some terms in VOSViewer analysis may appear as one term in one pool, and broken into multiple smaller ones in other pools (like terms ”climate change”, ”climate mitigation”, ”climate risk” etc.). In cases when these terms are relevant and analysed, these smaller terms are summed together before being compared to other research pools. In all cases, software analyses terms that appear at least 10 times across research. Despite fossil fuel industry papers being significantly less represented, such cut-off was deemed to be of no significance, since the software simply combines all mentions of a single word, rather than combinations of words in such case. Therefore, should term ”climate mitigation” be mentioned 7 times, term ”climate risk” 8 times, and ”risk mitigation” 5 times, all three terms ”climate”, ”risk” and ”mitigation” would still appear in software as separate terms.

Lastly, for the general and university pools, the amount of research was incredibly vast (usually in the range of 50-100 thousand papers). Scopus only allows for extraction of 20,000 papers, which were sorted by highest citation number and then downloaded. As other methods of sorting were either alphabetical or by date, this was deemed the least biased method of extracting such papers. Of course, there would still be more papers of older date, as they tend to accumulate more citations, but as temporal analysis showed, there was a sufficient presence of newer material. Another potential issue could be overrepresentation of established fields as compared to minor, lesser known journals and topics. However, for comparison against funded research, this was deemed acceptable, as it would allow for comparison with the popular trends in academic community of the time. Moreover, such highly cited papers ensure quality of research, and with 20,000 papers deemed as sufficiently vast dataset while still allowing for computationally efficient analysis.

3.1. Fossil Fuel Industry

To begin the analysis, all papers sponsored by companies listed in the methodology section (henceforth shortened to fossil fuel industry) within the last decade were collected. As mentioned in Chapter 2, the search query consists of 8 fossil fuel company names in ”Funding affiliation” section in Scopus, further filtered to consist of papers from 2014. This amounted to 10,153 papers, spread around various industries and fields. The large variation of topics can be seen in the word map divided into clusters by topic, as seen in the figure below:

Figure 3.1: Key terms map of Fossil Fuel Industry Research over the last decade. Bubble size indicates frequency of terms occurring, colours indicate belonging to a cluster of terms connected by relevance scores to join them into topics. Dark blue cluster indicates terms that are not directly related to technical research, but towards policy, climate, and stakeholders.

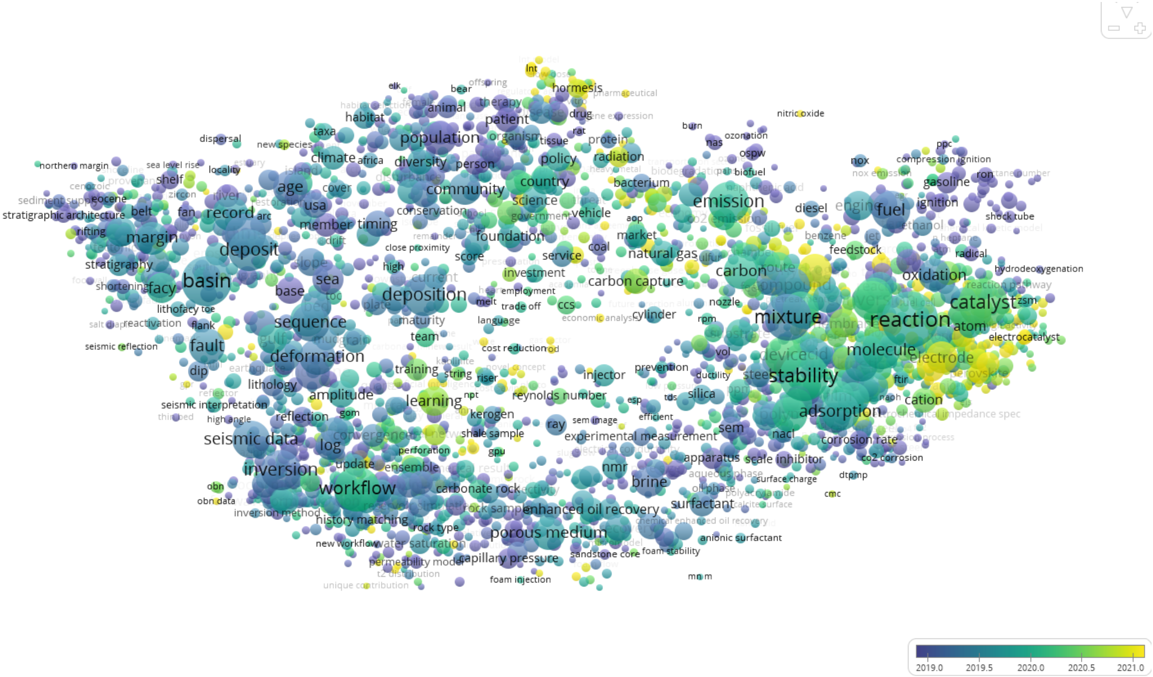


As discussed before, cluster analysis is conducted first. The terms are divided into 6 clusters, roughly corresponding to fields of geology (yellow), extraction (red), engineering (blue), chemistry (green), fuel and mechanics (light blue) and human (blue). These topics were identified by taking a look at their most frequent terms, and comparing to terms in other clusters, which is a straightforward method employed in other research with VOSViewer [28]. The human part, for instance, included pharmaceutical/medical terms such as ”drug”, ”vitro” and ”therapy”, as well as more policy related terms, such as ”economy”, ”emission”, ”investment”, ”employment” or ”tax”, which are not present in any other cluster.

Most climate change related terms fall within the last cluster, or in some cases overlapping with cluster of chemistry and mechanics (top right quarter of figure 3.1). This includes terms such as ”emission”, ”coal”, ”oil”, ”natural gas”, ”greenhouse gases”, ”carbon”, ”hydrogen”, as well as ”climate change”, ”emission reduction”, ”paris agreement”, ”alternative fuel” or waste.

A few interesting insights can be gained from this section. Firstly, three terms ”new technology”, ”price” and ”carbon capture” are clustered tightly together, indicating the frequent co-occurrence of these terms. In fact, ”carbon capture” occurs 134 times across these papers, roughly equivalent to term ”climate change” (135 occurrences), and more than ”co2 emission” (101 occurrences) or ”environmental impact” (77 occurrences). Such finding likely indicates the importance and hopes of carbon capture technology placed by fossil fuel industry as ways of mitigation of future climate-related risks, and the main hurdle associated with such technology.

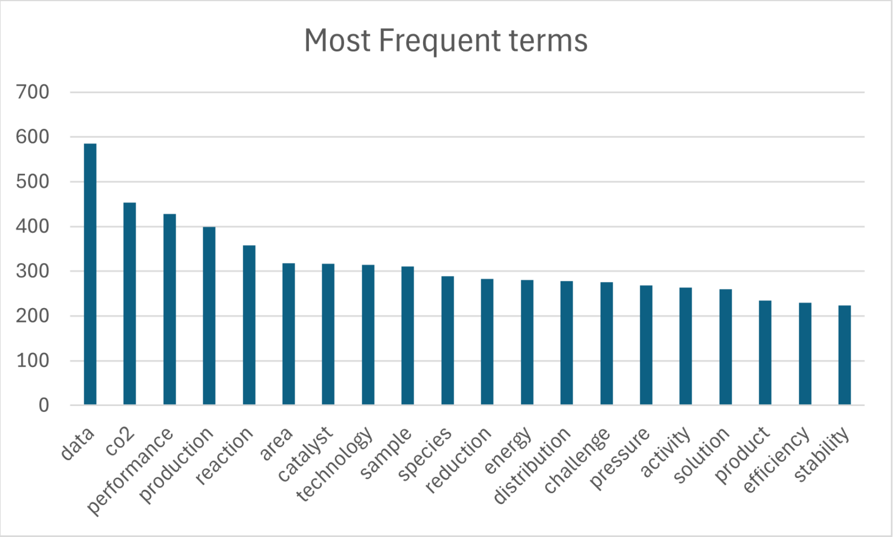
Figure 3.2: Overlay map of Fossil Fuel Industry Research over the last decade. Colour indicates average publication date, being the average date of a term occurring in research. Purple colour indicates earlier average date, or decreasing popularity of term over time, while yellow indicates newer date, or terms rising popularity. A combination of new or disappearing terms shows a trend in research.



Another finding can be seen in an overlay map, showing the weighted average time of publication of those terms (or the trend of their appearance in papers over time), as seen in figure 3.2. Here, brighter colour indicates newly appearing terms, with almost all terms related to hydrogen, renewable energy (”carbon emission”, ”hydrogen production”, ”electric vehicle”, ”energy transition” or ”decarbonization”) or batteries (such as ”electrocatalyst”, ”electrode”, ”water splitting”, ”battery”) being marked as yellow (their weighted average publication time being latest). This indicates a significant change of interest towards climate-change related topics, compared to industry-related research.

While such insights can seem interesting, and many more small notions can be found, this standalone analysis of fossil fuel industry sponsored research has no meaning in comparison to general research. However, unlike this pool of data, any next step needs to be scoped down to analysis by topic or area, as general pool of papers far exceeds any reasonable computational effort and would prove meaningless in comparison to this first step. It makes no sense comparing ”all fossil-fuel industry funded papers” to ”all papers together”, as topic variation and the amount of papers selected would be immense. Therefore, as discussed in Chapter 2, in the next steps, all pools of data will be narrowed down.

Figure 3.3: Most frequent terms of Fossil Fuel Industry Research. Terms are mostly technical in nature, and no conclusions can be drawn on this number standalone.



3.2. Energy field

The previous section established a significant, while logical, interest of fossil fuel industry in the fields related to energy and its sources. Therefore, for the next part, the analysis will focus on Energy subject area, as a filter in Scopus for all the papers retrieved. This field will be analysed from the perspective of fossil fuel funded research, then compared to general, and then the university pool is checked to see if there are any differences. All of these pools will be checked in two data timeframes: 2014-2019 and 2020-2024, in order to note and identify any potential changes in trends of research. In all cases, terms with minimum of 10 occurrences are included, and of those 80% of most closely correlated terms are plotted into a map in order to conserve computation costs (for each pool of 15-20 thousand papers, computation takes roughly an hour).

3.2.1. Fossil Fuel Industry

For the fossil fuel industry funded research, the pool is reduced to roughly 20% of the previous section. The pool of data for 2014-2019 consists of 1319 papers, while the pool from years 2020-2024 consists of 1515 papers. Unlike university or general pool, the period of 2020-2024, while unfinished, already has more papers published with fossil fuel industry affiliation, which may indicate growing ties and/or interest of the field in newest research.

For both of time frames, terms are divided into four main clusters, which can be explained as related to geology, chemistry, a connecting cluster (including terms like ”oil”, ”inhibition”, ”treatment” and ”recovery”), which combines terms related to both aforementioned clusters, and a fourth one of miscellaneous terms, relating, among others, to emissions, production and economy.

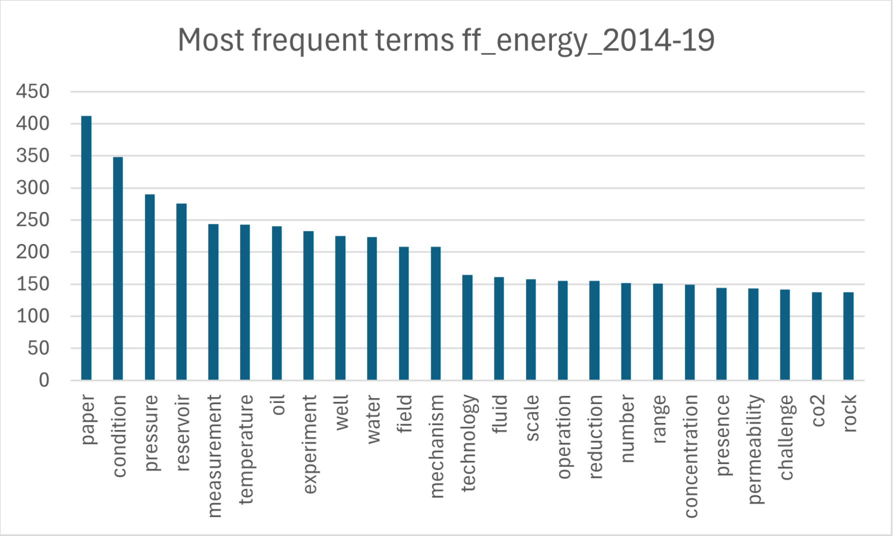
As can be seen in the figures 3.4 and 3.5 below, the terms of two pools are very similar, with main difference being lack of keyword ”data” in one pool and ”paper” in the other. This is explained by software choice to combine these keywords into pairs with other terms, such as ”field data” or ”production data” and ”research paper” and ”industry paper”. In both cases, term frequency is comparable when added.

Interesting patterns arise when looking at temporal data (graphs of which can be found in Appendix A as figures A.1 and A.2). As in section 3.1, most important increasing trend is related to the term ”hydrogen”. In both pools of data, weighted average publication time of the term is in the last years, and the increase in term usage is significant. While in the 2014-2019 period term ”hydrogen” was used 34 times, in the period of last five years, this has increased to 100 times. Similar thing can be said about ”carbon capture”, although increase in term use is less significant, from 23 to 51 times. Similarly, one can note the terms that have become a fresh ”trend” in the 2020-2024 pool, while being absent from the previous five year term, such as ”electrolysis”, which is closely linked to hydrogen and its related terms, and a small cluster of ”policy”, ”decarbonization” and ”energy transition”, having 30, 18 and 17 occurrences respectively in the latest pool.

The trend of seemingly more environmentally friendly terms is confirmed when looking in the opposite direction, and noting terms that are missing from the second pool of data, or have a significant decrease in its use. Most of such cases lie in the ”chemistry” cluster, with terms such as ”gasoline” or ”ignition delay” being oldest in the 2014-2019 pool and missing from the second one. While there are other cases of such terms falling out, or of fast-passing trends (term ”wettability” was among the latest in 2014-2019 pool with 33 occurrences, but in 2020-2024 pool it was one of the oldest, with 24 occurrences, signifying a decrease in interest), most of these relate to chemical terms and in fact correspond to a decrease of interest in combustion technology, with a significant increase in newer technologies, commonly perceived as climate-friendly.

As a final note, bibliographic analysis of co-authorship was performed for both pools. In both cases, out of hundreds of authors only a fraction was considered of significant connection (having 5 papers each and clustered). Interestingly, though probably not surprisingly, the authors are largely the same, with a few minor changes. For further interest, reader is referred to Appendix A, figures A.3 and A.4.

Figure 3.4: Most frequent terms of 2014-2019 pool of fossil fuel industry in Energy field. Similar to combined research, there is a prevalence of technical terms.



3.2.2. General Paper Pool

Unlike fossil fuel industry funding, this pool of papers is incredibly vast, with more than a 100,000 papers in each time frame. Therefore, papers were sorted by citation number, and a top 20000 papers were chosen to analyse for the sake of this section.

Important difference at the start is that the 2014-2019 data pool is divided into six clusters, while the latest data is divided into four. One of the two extra clusters is related to computational fluid dynamics (CFD) and is a significant outlier compared to the rest of terms, therefore for the sake of this analysis it can be discarded. The other five clusters can be described as chemistry, thermodynamics, ecosystem, human and electrical. Compared to that, 2020-2024 data pool removes the electrical pool of data, merging most of its terms to chemistry or thermodynamics.

Comparison of these two data pools between each other shows that they are very similar, since all key terms repeat throughout both pools such as ”temperature”, ”co2”, ”fuel”, ”warming”, ”economic growth” and so forth (figures A.7 and A.8 in Appendix A). From the top 30 terms listed by occurrences in both datasets, it is observed that the terms are identical, indicating a strong thematic consistency. This suggests that there is no significant change of key topics within the realms of energy and climate, emphasizing areas such as renewable energy, carbon emissions, climate change, and related technologies and policies.

Figure 3.5: Most frequent terms of 2020-2024 pool of fossil fuel industry in Energy field. No significant differences compared to 2014-2019 pool exist, indicating consistency in research.

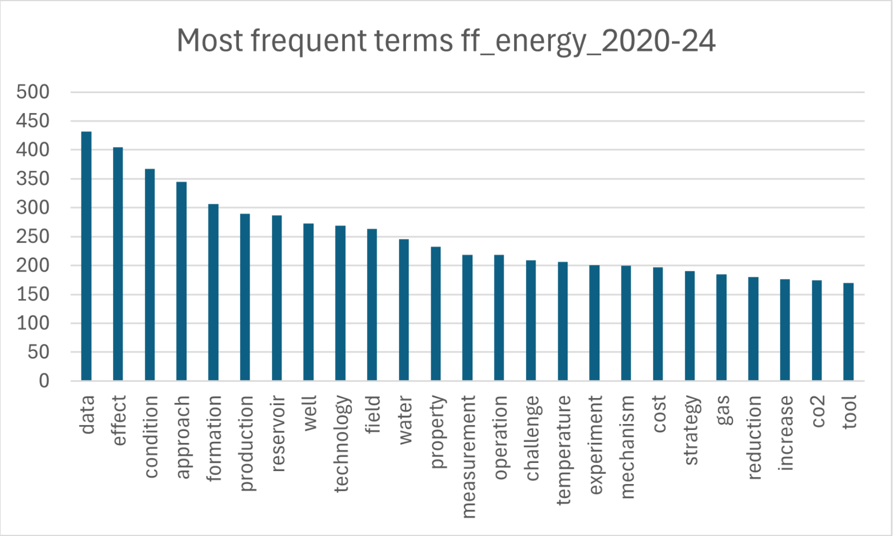
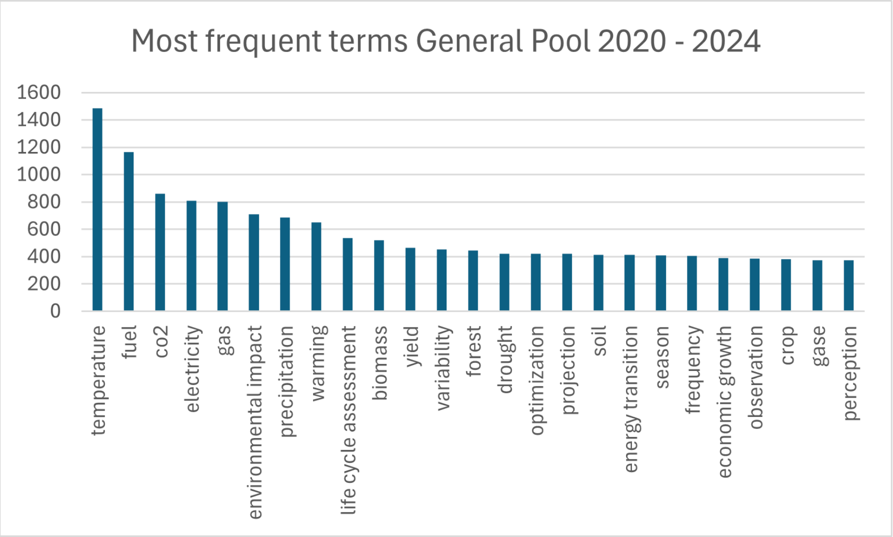


Figure 3.6: Most frequent terms of 2020-2024 pool for General Energy Data Pool. Presence of terms ”environmental impact”, ”life cycle assessment”, ”energy transition” and ”economic growth” are at odds with terms present in Fossil Fuel research, and confirm smaller focus of general research pool on technical research.



Interestingly, while temporal analysis shows that term ”Hydrogen” is relatively new in both data sets, there is no significant difference in the number of term occurrences, indicating that within general field of Energy research, hydrogen has been a subject of study for some time and is not gaining significant traction increase compared to previous years. The main difference between two datasets is presence of terms like ”low emissions analysis platform” and ”lower carbon footprint” in the 2019-2024 dataset, indicating new emerging terms and interests.

Comparing the two pools to fossil fuel ones, some key differences can be distinguished. The most logical one, is that terms like ”well,” ”reservoir,” ”oil,” ”gas,” and ”pressure” are prevalent in the fossil fuel industry research but absent from the top terms in the general pool, since this is the main focus of the fossil fuel industry. Conversely, terms like ”CO2,” ”emission,” ”climate,” and ”warming” are common in the general pool but not in the fossil fuel one.

The presence of terms like ”policy,”, ”stakeholder”, ”warming”, ”economic growth”, ”carbon tax”, ”livelihood” and many others in the general pool are more policy-oriented, indicating a focus on climate policy, environmental impacts, and mitigation strategies. While the term ”policy” occurs in fossil fuel search 30 times within 2020-2024 pool, general one in the same time perspective has 32 different terms with the world ”policy”, such as ”eu policy”, ”adaptation policy” etc., ranging from 16 to 108 occurrences for each term. Notably, terms ”stakeholder”, ”mitigation” and ”warming” are absolutely missing from fossil fuel industry pool, while each occurring more than 300 times in each of general industry pool, which may indicate less interest in decision-making and climate policy by fossil fuel funded research.

Figure A.9 in Appendix A shows all terms that can be found in both general pools and 2020-2024 pool of fossil fuel industry, normalising the score by dividing the occurrence weight of general papers by ratio of general paper pool to fossil fuel energy pool. It shows that only one term within the list of common terms is related to climate policy, namely the term ”energy transition”. Coincidentally, it is one of the few terms in that figure that is way more frequent in general pool rather than fossil fuel one, further indicating the lack of climate or policy consideration in research funded by fossil fuel industry.

Finally, citation analysis of this pool has not revealed any significant patterns that could be compared to fossil fuel industry research. As can be seen in figures A.5 and A.6, the clusters are messy and interconnected, indicating a significant overlap and cross-citation of various works. There are also no noticeable author correlations with fossil fuel industry pool of research.

3.2.3. Universities Pool

Finally, the selected university research is analysed, with 2014-2019 pool consisting of 18924 papers, and 2020-2024 pool consisting of 16895 papers. This means that no sorting or reduction of pool was necessary.

Just like in the case of general research pool, university pool of 2014-2019 was divided into six clusters by relevance score, with four clusters for the 2020-2024 pool, with one significant difference being the presence of more technical terms in the pool of ”ecosystem”, such as ”biodiesel”, ”biofuel”, ”pyrolisis”, ”cellulose” etc. In this case, both ”extra” clusters are merged into larger ones in 2020-2024 pool and can be described as: chemistry, electricity grid, batteries and miscellaneous/human.

This cluster naming already indicates first major difference with general pool of data - much less emphasis on soil/ecosystem terms, and a much larger than general emphasis on battery/electrical research. Such discrepancy may largely be attributed to university specialization and does not show a pattern in standalone. It is necessary to note, that terms such as ”agriculture”, ”ecosystem” or ”forest” are still prevalent, but are more closely related to ”human/miscellaneous” cluster of terms.

In fact, these two clusters of data show by far most consistency, as shown by almost identical top list of frequent terms between two datasets (once again reader is referred to Appendix A, namely figures A.10 and A.11). Similarly to general data, this pool includes a lot of climate, policy or mitigation/risk management related terms, in sufficient quantities (a range of 100 occurrences or more). Figure A.12 is a great example of diversity of university research. For a large amount of co-occurring terms, there is a large variety of discrepancy in which pools these terms occur more frequently. As such, it can be concluded, that university research in this field has no significant patterns differing from general pool of research, while having a significant difference with fossil fuel industry related research.

Once again, citation analysis reveals no patterns or correlation of significance, as no overlapping authors can be found in both fossil-fuel funded and other pools, and the variability of researchers does not tell much. Reader is referred to section 4.8 of this chapter for further information on citation analysis. Clusters, while plenty, are tightly connected, and show no significant overlap with either fossil fuel or general pool of research. Figure A.13 is included in the appendix as proof of such lack of correlation.

3.3. Environmental Science

Second subject area to be checked is environmental science. Being directly related to climate research, this subject area is highly relevant for the subject at hand, and while it is not the main focus area for fossil fuel industry, there is a large enough pool of fossil fuel data to allow for an adequate comparison.

3.3.1. Fossil Fuel Industry

As mentioned above, fossil fuel industry has relatively little interest in this area compared to energy sector. Nevertheless, 741 documents were identified as having industry affiliation in the 2020-2024 timeframe, and 613 additional documents in 2014-2019.

In both cases, around 300 identified terms with 80% relevance score were converted into four clusters. These clusters, in both cases, can be described as two technical and two non-technical, with one of each leaning towards environmental side of terms, and the other towards fuel/chemistry side of terms.

Temporal analysis shows a growth of interest in natural gas and hydrogen in 2014-19 pool, while the later pool shows the establishment of hydrogen and natural gas as a research topic (it is not emerging anymore), with more attention being paid to ”cancer risk assessment” in the last couple of years. As a reverse trend, terms related to ”ozonation” or ”ozone” have an average publication year of 2015 and are not present in 2020-2024 studies. One thing in common with energy industry is prevalence and emergence of term ”carbon capture”, with 20 occurrences and average publication date of 2018 in 2014-2019 pool, and 44 occurrences in 2020-2024 pool, with an average publication date of 2021.

In general, the 2020-2024 pool of data shows much more interest in terms such as ”greenhouse gases” or ”sustainability”, emerging in the second pool only, similarly to the term ”policy”. Somewhat surprisingly, the same pool of data lacks the term ”oil”. However, further comparison of these two pools of data shows no significant difference in frequencies or changes of top terms, and only comparison to general pool of data may be able to inform us about patters or trends at hand.

As a final note, citation analysis has shown no patterns. Interestingly, unlike in Energy subject area, even pools of 2014-2019 and 2020-2024 had entirely different authors involved.

3.3.2. General Pool

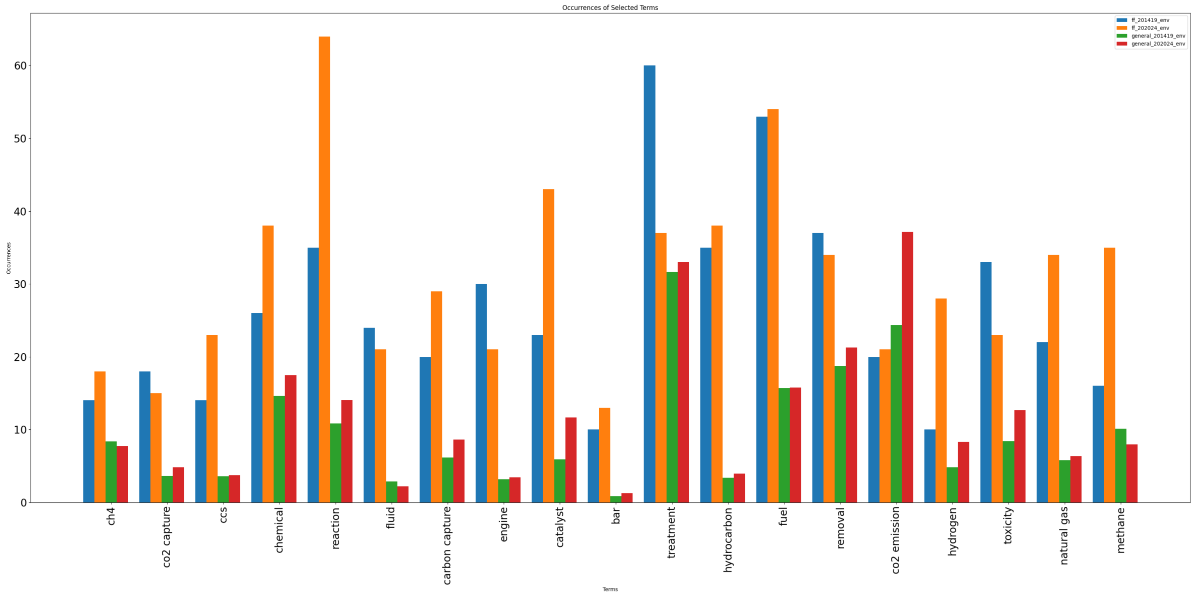
Pool of Environmental Science is incredibly vast, and even after limiting it to containing the term ”climate” anywhere in abstract or title, it comes out to more than 280,000 papers in 2014-2019, and more than 422,000 papers in 2020-2024. As such, once again 20,000 papers were chosen to analyze based on highest citation numbers.

In fact, the pools are so vast, that terms that are mentioned more than 10 times and have at least 80% relevance score are in the range of 5000-6000, leading to 6 different clusters of data in 2014-2019 pool, and seven such clusters in data pool of 2020-2024. While most clusters are similar to those encountered previously, the seventh cluster of 2020-2024 pool is dedicated entirely to methodology (with terms such as ”machine learning”, ”algorithm”, ”forecast”, ”deep learning”, ”neural network” etc.). These methods are also highly visible in word maps of 2014-2019 pool, though they are a part of larger weather/temperature cluster. More details on methodology and its comparison will be available in section 3.6 of this chapter.

When comparing the two data pools one significant difference becomes the emergence of Covid-19 pandemic. In the 2020-2024 pool term ”covid” is mentioned 847 times, with word ”pandemic” occurring 507 times. This is the main difference between pools, as most key terms, relating to environmental impact, emissions or various chemical/fuel compounds are almost identical. Figures A.16 and A.17 show this, with main difference being a term ”concentration” being absent from 2020-2024 pool, as it was divided into sub-terms such as ”pm10 concentration”, ”mass concentration”, ”ozone concentration” and so on. Temporal analysis shows no significant patterns in the research, with most newly emerging terms relating to artificial intelligence or other methodology, a trend similar for both pools of data.

More differences arise when comparing the general pools of data with fossil fuel industry. For instance, in the previous section it was noted that ”ozone” and related terms are disappearing from the radar of fossil fuel industry. This could not be further from the truth for the general environmental research, with 8 various ”ozone” related terms occurring around 200 times total in each pool, with varying publication year averages (indicating a spread-out and disproving the claim that ozone related studies have gone down in numbers over time). Figure 3.7 below compares the terms that are present in all four datasets, normalised to take into account the number of papers present in the pools. There are only 19 terms in total that occur frequently in all of them. Similarly to the Energy field, term ”carbon capture” occurs much more frequently in fossil fuel industry related search, while ”co2 emission” and ”carbon emission” are much more prevalent in general research, as do many other climate change related terms. Once again, however, there is a vast difference in policy or mitigation/risk assessment related terms. In fact, 7 different terms with the word ”mitigation” are present in general pools of both time frames, with occurrences in the range of 100, with no mention of this word in fossil fuel industry sponsored research. Term ”policy”, while occurring 24 times in fossil fuel industry pool of 2020-2024, is significantly underrepresented, given that the general pool of data contains 40 different terms with word ”policy” for 2014-2019 pool, and 39 such terms in the 2020-2024 pool.

Figure 3.7: Normalised comparison between common terms in fossil fuel industry and general data pools. Blue and orange bars indicate term occurrence in fossil fuel industry pools, green and red the general pools. Scores are normalised to even out the number of papers in the pool. X-axis shows the term, Y-axis shows the score. Nearly all common terms across research are more prevalent in fossil fuel research, with the exception of ”co2 emission”. Note the large discrepancies in ”ccs”, ”carbon capture” and ”co2 capture” frequency.

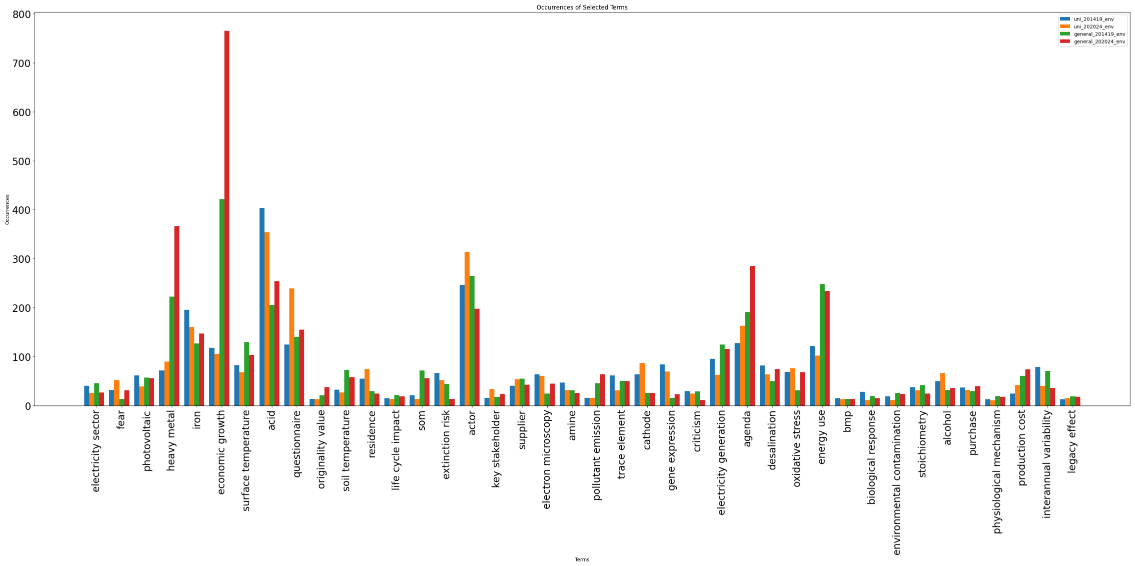


3.3.3. University Pool

Finally, university pool is being checked for the environmental sciences subject area, to see any patterns or trends emerge. These universities alone produced 25,973 papers between 2014 and 2019, with another 26,771 papers in five years since. Once again, top 20000 by citation numbers are selected.

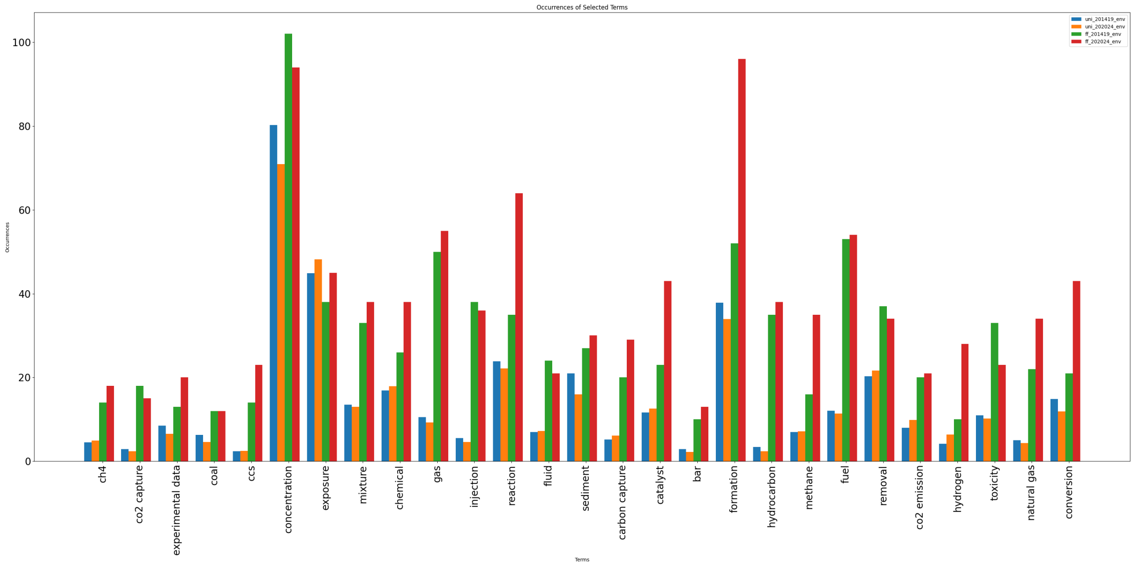
Figures 3.8 and 3.9 below show normalised comparison of common terms between university pool of research and fossil fuel industry, as well as general pool, respectively. The aim of such comparison is to directly note the similarities and differences between these pools of data, and establish, which ones do university pools resemble more. As is visible, the university pools has many more terms co-occurring with general pool of data than with fossil fuel industry, and once again, the same pattern emerges with ”carbon capture” and any emission-related terms. Naturally, any fossil fuel terms are much more prevalent in fossil fuel funded industry, but as can be seen in A.18 and A.19, similar to previous patterns of policy and mitigation terms being much more prevalent in research with no established industry funding affiliations.

Figure 3.8: Normalised comparison between common terms in university and general data pools of Environmental research. Due to a large number of common terms, only top 20 terms are selected by their average occurrence. Note higher variability in pools with largest occurrence, and a larger prevalence of topics related to environment or policy/economics.



As with general data, temporal analysis shows an emerging pattern for use of artificial intelligence and deep-learning methods in environmental research. No significant patters of terms missing or disappearing over time could be established.

Figure 3.9: Normalised comparison between common terms in fossil fuel industry and university data pools of Environmental research. Similar to fossil fuel and general comparison, terms are almost exclusively more represented in fossil fuel industry research, are more technical, and a large focus is paid towards ”carbon capture” and synonymous terms.



As a final note, citation analysis shows some smaller clusters separated from the others, unlike in the case of general data. This is likely an influence of researchers related inside university. Most of citations are tightly interconnected in both university and general pools of data, allowing for no significant pattern establishment.

3.4. Earth and Planetary Science

A third important subject area to analyse is Earth and Planetary Sciences. This subject area is the largest among papers that have been funded by fossil fuel industry, which comes with no surprise, since it involves geologic scanning for discoveries of new fossil fuel locations, among many other important research areas. Insights into what interests fossil fuel industry in this area, as well as particular university research and the general pool can shed more light into whether any research biases exist.

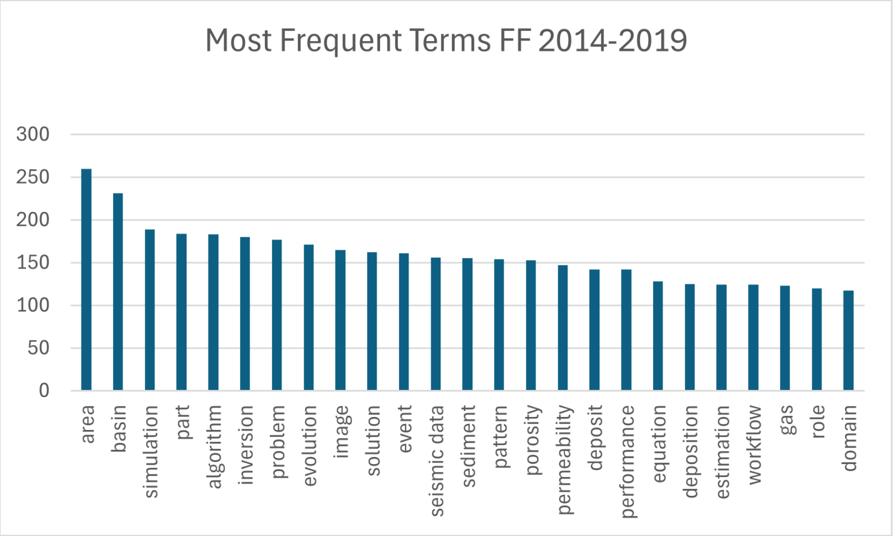
3.4.1. Fossil Fuel Industry

As mentioned previously, this subject area has the largest amount of papers that were funded by fossil fuel industry, though not by a large margin. For 2014-2019 period, this amounts to 1842 papers, and another 1626 papers in the next five years. Their analysis is presented below.

The first interesting insight is the number of clusters. While the 2014-2019 pool still has a usual number of 4 clusters (roughly relating to soil, extraction, data management and structural analysis), 2020-2024 pool skips the last cluster and only has 3 of those in total. There are no noticeable or unsurprising patterns in variation of terms. Figures 3.10 and 3.11 show the most common occurring ones in the research. Perhaps the only difference is once again terms that are broken into a few different terms including the same keyword in the other dataset, such as ”oil”. Once again, there is no presence of any terms relating to policy, risk assessment or any other stakeholders involved, only terms strictly relating to the industry. Once again, term ”carbon capture” is present in the 2020-2024 pool with the latest average publication date of all terms present (2022.4), indicating a freshly emerging interest.

Interestingly, temporal analysis in this case tells a lot about focus on specific countries. For instance, term ”Canada” is present among the oldest in 2014-2019 pool, while not present at all in the next pool. Vice versa, newly emerging terms in both cases are ”Netherlands” and ”Brazil”, with the first one being mentioned 11 times in both cases, while Brazil term jumping from 22 occurrences in the first pool to 113 searches in 2020-2024 pool.

Figure 3.10: 25 most frequent terms in fossil fuel industry Earth & Planetary Sciences pool for years 2014-2019. Similar to other fossil fuel industry research by topics, technical terms prevail.

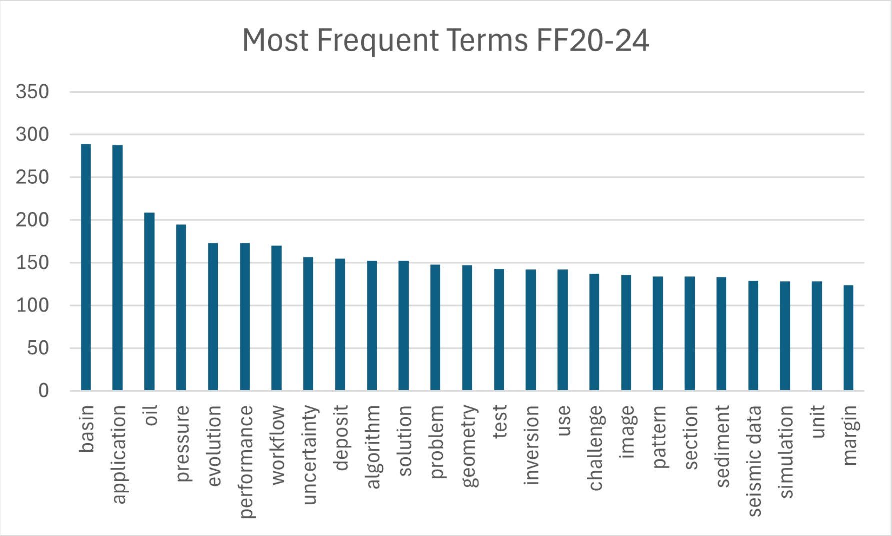


3.4.2. General Research Pool

With general research pool being over a million papers, search term ”climate” reduced it to 205,000 papers for 2014-2019 pool, and 234,483 papers for 2020-2024 pool. Sorting by citation scores, top 20,000 papers were chosen for the analysis.

Figure A.20 shows comparison of common terms between fossil fuel related papers and general ones. Unlike comparison in other industry, almost all papers here are much more frequently represented in Fossil Fuel research. Moreover, most of these terms are not directly related to climate, policy, stake-

Figure 3.11: 25 most frequent terms in fossil fuel industry Earth & Planetary Sciences pool for years 2020-2024. No significant differences compared to 2014-2019 pool, and technical terms only once again.



holders or risk assessment, being technical in nature. This explains why all such terms are more frequently represented in fossil fuel research - general research pool is more diverse in nature, and has less focus on specific areas of research that fossil fuel industry prefers.

Part of this can be seen in Figure 3.12. A whole cluster (green) is dedicated to terms that are generally missing from fossil fuel industry sponsored research, related to consequences, such as terms ”policy”, ”adaptation”, ”decision-making”. More notably, here we see mentioning of many smaller stakeholders, such as ”smallholder farmer”, ”rural community”, ”citizen”, ”local level” etc. These terms are equally present in both 2014-2019 and 2020-2024 pools.

Some terms that are present in general research pools are missing from fossil fuel industry research for a logical reason of having little to do with relevant research. For instance, quite a lot of it is related to ice, with terms such as ”glacial”, ”ice age” or ”last glacial cycle”. While directly related to climate research, it is not surprising that any industry that is not directly related to this phenomenon would be absent from research in this area.

Lastly, temporal analysis in this area reveals little about trends in research other than methodology. Logically, use of ”machine learning” or ”deep learning” methods increases substantially in the 2020-2024 pool, with their occurrence count increasing roughly twice, and averaging 2023 year publication date. While there are many smaller terms that appear or disappear from research in various years, all the main terms remain identical, and little can be said about trends further in research.

3.4.3. University pool

Finally, university pool, consisting of 22,779 papers for years 2014-2019 and 18,424 papers for 2020-2024, is analysed below.

Most notable difference, as presented in Figures 3.13 and 3.14 is in presence of ”Planetary” part of research, with terms like ”galaxy”. While in general pool these were filtered out due to presence of term ”climate”, it was not filtered out of university research and forms a large part of research. As can be seen in Figure A.21 in the Appendix, this pool of research has a lot of terms that occur as often as in general research, with a huge variation in what terms are prevalent in which pool of data.

Once again, we see similarities with general pool in terms of consideration of smaller stakeholders, policy and risk mitigation. Moreover, temporal analysis once again relates mostly to methodology, which will be expanded upon later in this chapter. Concluding, this research pool comparison is very similar to the ones done on other subject areas - university pools are quite similar to the general pool both in terms of topic diversity and trends of research.

Figure 3.12: Map of frequent terms used in general research pool of years 2020-2024 in Earth & Planetary Sciences. The green cluster with policy and stakeholder information on bottom left is of key importance here as it contains largest differences compared to fossil fuel industry.

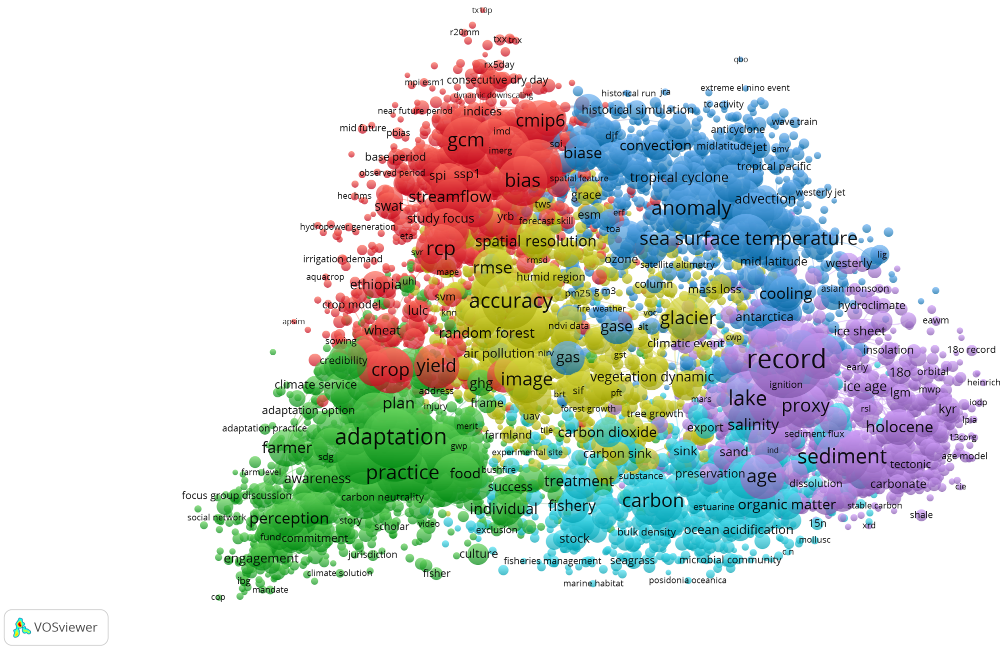


Figure 3.13: Map of terms used in university research pool of years 2014-2019 in Earth & Planetary Sciences. Note an uncommon shape and large disconnection between two sides of clusters, mainly due to focus of research in certain universities, which are much more related towards "Planetary" part of the subject area. This part is logically unconnected to anything fossil fuel related. Green cluster on the left side contains climate related terms that are similar to general research pool.

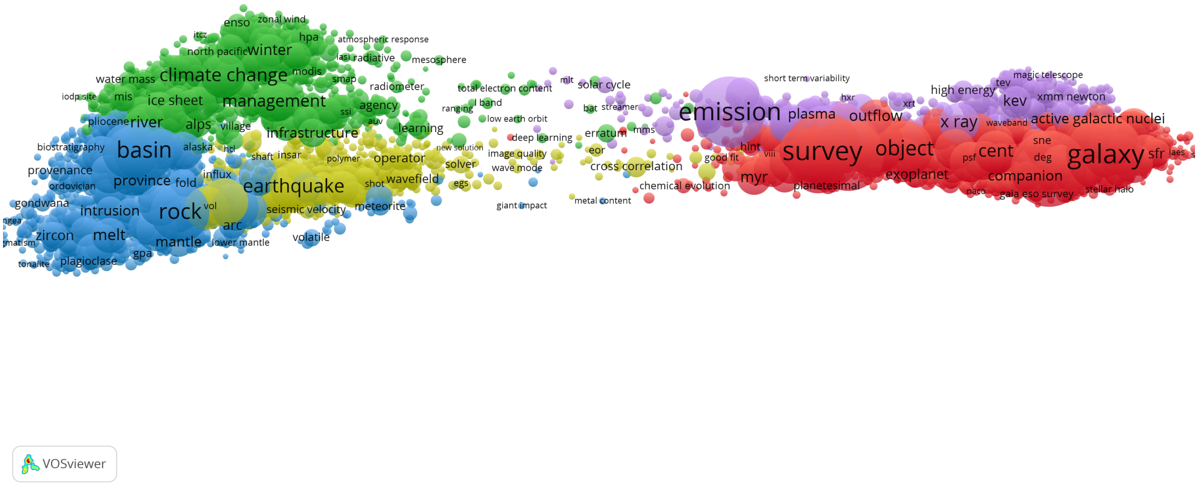
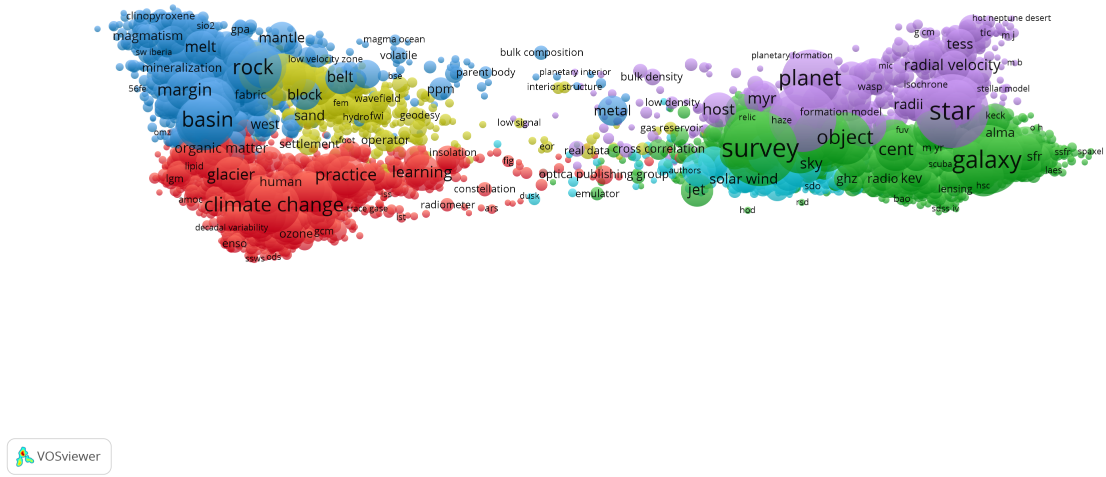


Figure 3.14: Map of terms used in university research pool of years 2020-2024 in Earth & Planetary Sciences. Similar to 2014-2019 pool, it has differences compared to general pool in terms of ”Planetary” part of research, yet has more climate related terms than fossil fuel funded research (here represented by red cluster on the left side).



3.5. Engineering

Finally, engineering section closes the top 4 subject areas that are of interest to fossil fuel industry. This analysis will consist of 851 and 1043 papers for fossil fuel industry in 2014-2019 and 2020-2024 pools respectively, as well as a selection from over 126,000 and 196,000 papers from general access, as well as 45,026 papers from university pool in 2020-2024, and 51,298 papers for five years prior to that.

3.5.1. Fossil Fuel Industry

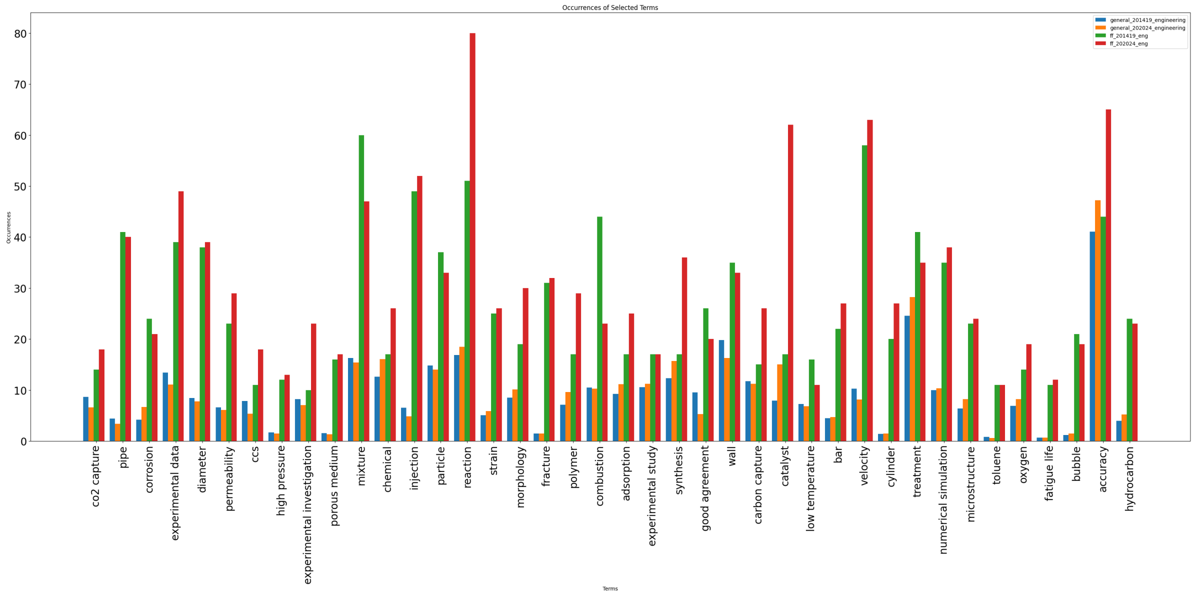
First noteable mention in this section is that despite a relatively low number of terms, both time frames have terms divided into five clusters, including chemistry, fluid dynamics, combustion technology, geological terms and general terms. Similarly to previously analysed subject areas, there are some mentions of methodology (terms like ”machine learning”, ”field data”, ”review” etc.), some general terms like ”technology” or ”cost”, and many fossil fuel related terms, from ”co2” to ”natural gas”. Once again, ”carbon capture” technology is prevalent, spread into two terms ”carbon capture” and ”co2 capture”, as well as neutral term ”capture” that may or may not refer to the same thing. Even excluding the latter, two terms combined occur 29 times in 2014-2019 pool of data, jumping to 44 in 2020-2024 data. Another similar increase is prevalent in term ”emission”, coupled with ”co2 emission” - these terms occur 63 times in the first five years, jumping to 105 occurrences in 2020-2024. Of fossil fuel terms, there is a surprising lack of term ”oil” in both pools of data. ”Natural gas” seems to stay constant in interest, scoring 25 and 27 occurrences respectively, while ”coal” makes an entrance in 2020-2024 pool with a minimal number of 10 occurrences.

Another similarity to previous subject areas, is lack of general terminology on policy, risk management, stakeholders or such. Figure 3.15 compares common terms between fossil fuel and general pool of data, showing once again that common terms relate only to directly engineering related terms, with ”carbon capture” and ”co2 capture” technology being vastly more popular among industry financed research.

Interestingly, in this case temporal analysis data shows a small scale of 1 year in both pools, indicating that all terms are somewhat evenly spread out across the time scale. Given the relatively small amount of terms, it seems logical, as the essential terms captured are not groundbreaking in nature, and research including these terms is long-term.

Key terms direct comparison shows a decreasing use of term ”algorithm” (which has 5 occurrences less in 2020-2024 pool of data despite having more papers), and logically, a significant increase in machine learning technologies. Term ”prediction” is absent from 2014-2019 pool of data, jumping to 80 occurrences (12th most common term) in 2020-2024 pool of data.

Figure 3.15: Comparison of normalised common terms between fossil fuel and general pools of data for engineering papers. Most terms are once again more prevalent among fossil fuel research. Note the term ”hydrogen”, which has a significantly larger use in both fossil fuel and general research in 2020-2024 pools compared to their earlier equivalents.



3.5.2. General Pool

Naturally, this pool of research is much more diverse than fossil fuel industry related one, most notably focusing more on construction (for instance, term ”concrete” is among top-30 terms in both 2014-2019 and 2020-2024 data pools respectively. Being so much more diverse, general pools can tell us a bit about patterns in research that should be visible even in funded papers.

Large part of this is ”business/management” side of research. Terms such as ”economic growth” or ”supply chain” are directly relevant to the field of engineering and to the fossil fuel industry, yet are not present in funded research, similarly to terms ”mitigation” or ”governance”, as well as various stakeholders (”employee”, ”worker”, ”project manager”). The largest gap present here is related to climate. General pools of data have a large presence of a whole variety of terms like ”environmental degradation”, ”sustainable development goal”, ”ecological footprint” or even ”environmental kuznets curve”. These are not small terms, but each besides the last one occurring more than a 100 times in the 2020-2024 pool.

The fact that the last pool was mentioned is no surprise. Temporal analysis of both pools of data shows a very clear trend towards an enlargement and diversification of various terms related to ecology (even among technical terms, latest average publication time is in terms like ”battery”, ”hydrogen” and ”photovoltaic”, in both pools. Reader is referred to Appendix A figures A.22 and A.23 for more insights.

3.5.3. University Pool

Finally, the university pool is analysed in order to find out if any patterns persist. Both pools look very similar with one outlier cluster, related to ”dark matter” and space-related terms, since these universities may specialise in such terms. This outlier is unrelated to both general or fossil fuel funded research, hence is ignored from now on.

In general, university research contains more specific and technical terms like ”molecular imaging,” ”monolayer mos2,” and ”trabecular bone,” indicating a deeper dive into niche research areas. It includes terms like ”academic literature” and ”academic research,” suggesting a context rooted in academic research and scholarly analysis.

This contrasts with general research, which, while also technical, includes broader terms such as ”absolute error,” ”academic performance,” and ”total cost,” suggesting a more generalized overview of engineering topics. Terms such as ”wearable electronic,” ”weber number,” and ”wellbeing,” indicate a broader range of application contexts, potentially including industry and consumer applications.

While in comparison, many more terms are common between the general and university pools of data, with more variability in terms of which terms is more popular in which research, some important differences arise. University research pools are generally much more methodologically diverse, including terms such as ”convolutional neural network”, ”optimization”, and dozens of different ”schemes”, it has less focus on climate-specific terms, which, while still in abundance, are significantly less prioritized compared to general research. Thus ”sustainable development goal” is mentioned 50 times compared to 367 in 2020-2024 pools (of identical amount of papers), while ”environmental degradation”, let alone ”environmental kuznets curve”, like many other medium-size terms of general pool, are lacking in university research. This may be explained by more technically focused research in these specific institutions, but it also indicates, unlike in other subject areas, a sufficient discrepancy between general and university pools of data.

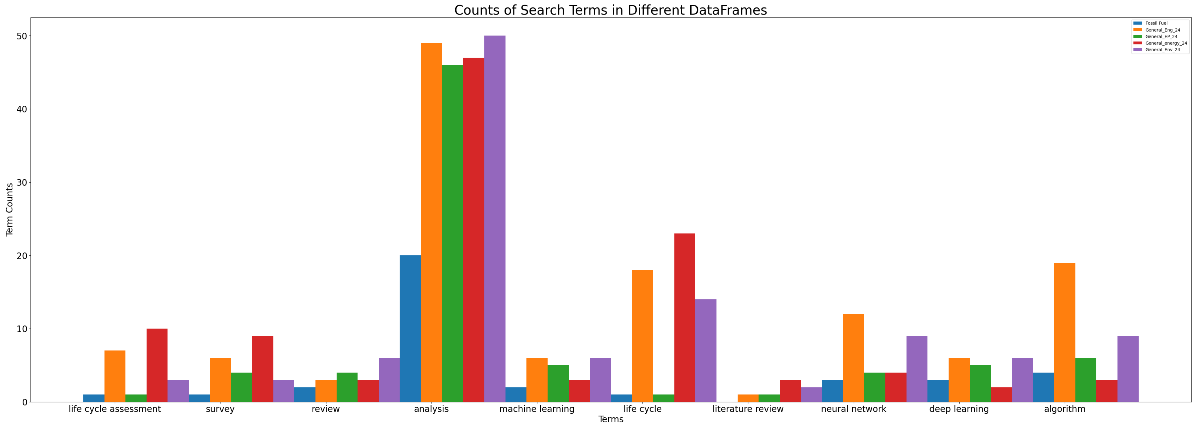
3.6. Methodology comparison

In this section, Research Sub-Questions 1 and 4 are addressed from the perspective of methodology. This means, that all research pools mentioned above are analysed specifically for terms relating to research strategy, methods employed and their frequency. Fossil fuel pool, for ease of comparison, was analysed in its initial stage, with all papers from 2014-2024 combined altogether, adding to 10,153 papers. This is still twice smaller than any dataset that general research produces, which is taken into account when scoring.

The documents share a common ground in their emphasis on empirical research methods and design methodologies. Terms related to machine learning, specific subject area terms (such as ”electron microscopy”, ”microstructure analysis”) etc. are prevalent in all areas, and once again reader may be welcome to review comparison figures in the Appendix to note such similarities. Perhaps one main exception is ”environmental kuznets curve”, which does not appear in fossil fuel reviews, however, when normalising the scores it becomes evident, that the term may well be included in some fossil fuel industry funded papers, yet not appear in the review.

A noteable difference occurs in a small representation of terms related to ”survey”, ”literature review” or ”hypothesis”. In fossil fuel data, term ”seismic survey” appeared 48 times, with the word ”survey” not being present anywhere else. In general pool of energy for 2020-2024 alone, this term appeared 321 times, while engineering pool has this term occurring 343 times in various forms, from ”online survey” to ”field survey”. With term ”review”, it appears 33 times in fossil fuel industry, compared to 213 occurrences of ”systematic literature review” alone in general pool of engineering in 2020-2024, jumping to 273 occurrences if all terms with ”review” are considered”. Term ”hypothesis” does not appear in fossil fuel industry research at all, while in general research it appears 100-200 times in all pools except for energy field (where ”ekc hypothesis” term occurs 24 times in 2020-2024 and 3 various ”hypothesis” terms occur 37 times in 2014-2019 pool). Finally, the term ”life cycle/lifecycle/life-cycle” appears in fossil fuel industry pool as ”life cycle assessment” a total of 17 times, while in every general pool this term appears roughly 500 times, usually in 10-18 different term combinations. This term shows the most striking difference between approaches of fossil fuel industry and general pool of research. Normalised term comparison of such selected terms can be seen in Figure 3.16:

Figure 3.16: Comparison of normalised terms between fossil fuel and general pools of data for specific methodology terms. Here, blue bar indicates fossil fuel funded research, while others are general pools of various subject areas. Note the complete absence of ”literature review” among fossil fuel industry research, as well as decreased presence of ”life cycle” terms. Other terms, while still less prevalent compared to general research, do not have such a significant discrepancy.

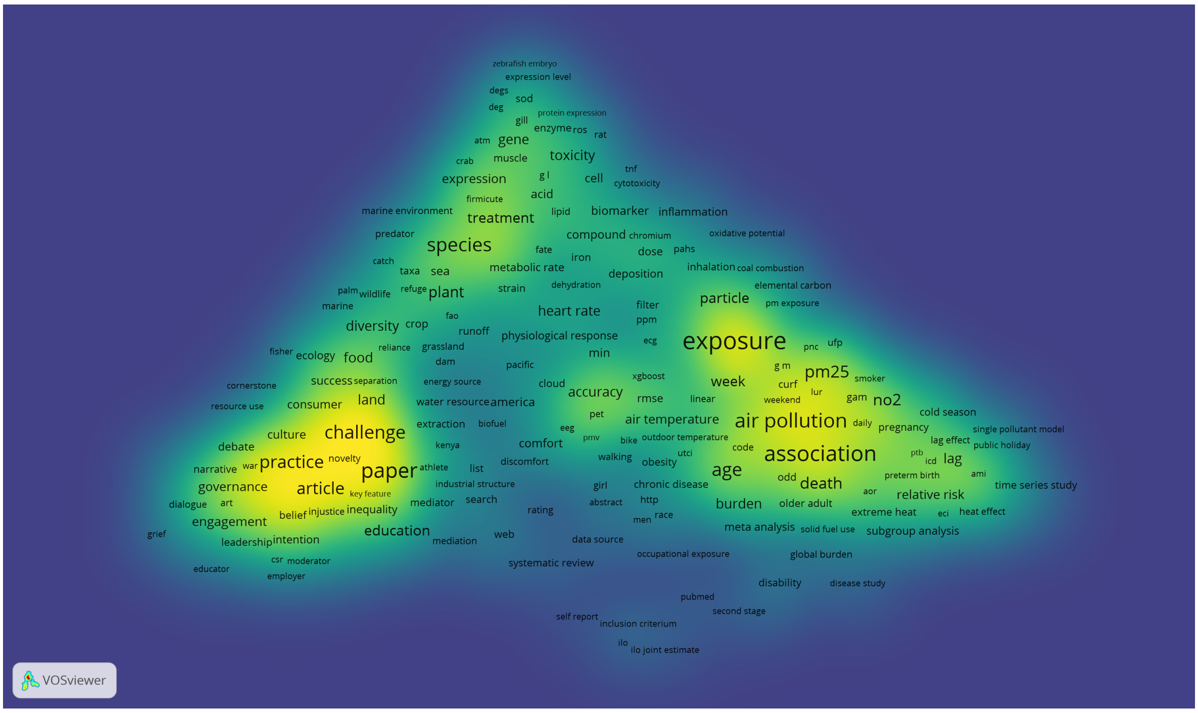


Many potential explanations exist for such phenomenon, such as publication bias (there might be a publication bias where industry-funded studies are more likely to be published in industry-specific journals that focus on technical advancements rather than broader academic journals that emphasize comprehensive reviews and analyses), perceived value of such research and its focus or strategic interests of funding source. Industry-funded research might be more collaborative with a focus on specific applied research projects rather than fundamental research, which often includes extensive reviews and analyses aimed at advancing scientific understanding in a broader sense. Analysis of these possible causes is, however, outside of the scope of this project.

3.7. Term Search

In this section, mainly due to a large amount of general research, the search results were further limited to include the word ”climate” and later other terms, since this topic area is most important for this research project and thus helpful in establishing relevant paper pools. Since previous results seem to indicate an interest of fossil fuel industry funded research in topics closely related to their operations, the goal of this section is to move away from those topics and see if focusing more on ”climate” would bring a difference in results. For this purpose, the pool of all fossil fuel funded papers is limited to those including term ”climate”, decreasing the pool from over 10000 papers to 6,907. The results can be seen in the figure 3.17 below:

Figure 3.17: Map of most frequent terms used in Fossil Fuel Industry Funded papers with the term ”climate”. Yellow bright spots indicate highest relevance score, containing both frequently appearing and commonly appearing together terms. Large frequency of ”air pollution” and ”exposure” cluster on the right side indicates an interest in consequences of fossil fuel industry activities, though policy or stakeholder analysis is still largely irrelevant.

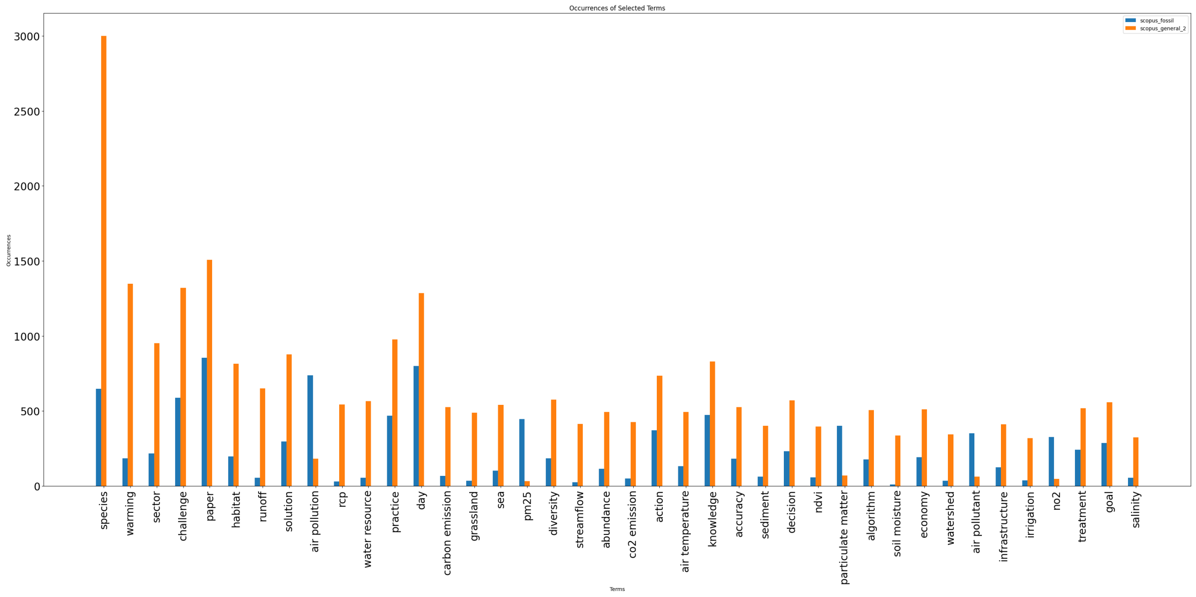


What may seem surprising at the first glance is the lack of term ”climate” in this map. However, such insight is misleading, as the word is present in 16 different terms, ranging from ”climate action” to ”climate warming”, all together occurring 330 times across papers. This, however, would not make it the most frequent occurring term by far, as for instance term ”air pollution” occurs 738 times, while term ”exposure” is present 1245 times.

Similarly, word ”policy” occurs 123 times across 7 different terms, and word ”stakeholder” a further 177 times. Moreover, even methodology terms ”survey” and ”literature review” appear, a total of 50 times. There are still no mentions of the word ”mitigation”, however. Figure A.24 in the Appendix A shows the comparison between common terms of fossil fuel and general papers with term ”climate” without division by subject area. While the similarities are much more prevalent than in the previous sections, it is interesting to note significant differences in some terms. For instance, the first thing that catches attention is the huge difference between general and fossil fuel papers in regards to terms ”species” or ”habitat”. Similarly, large differences exist with regards to terms ”sea”, ”diversity”, ”rcp” (standing for Representative Concentration Pathway [13]), ”co2 emission”, ”carbon emission”, ”ecological restoration” and a few others. The terms that are significantly more represented in fossil fuel papers are ”theoretical basis”, ”cumulative effect”, ”pm25”, ”air pollutant” and ”death”. Figure 3.18 shows terms with largest normalised difference between their occurrences between general and fossil fuel pool of research, sorted from largest difference. Terms ”species”, ”warming” and ”carbon emission” are much more present in general research, while fossil fuel research prefers terms ”air pollution” or ”pm25”. Such differences once again are likely to arise from focus of fossil fuel industry on what directly affects their industry and actions, though any further discussions are left to chapter 4.

In order to check the patterns, seeing if anything else may emerge, a number of other terms were included as a search term, such as ”greenhouse gases”, ”carbon”, ”emissions” or ”oil”. However, it soon became clear, that having analysed fossil fuel industry funded papers in their entirety, exclusion of papers to having a specific term would either result in an insufficient amount of papers in a research pool (for instance, term ”greenhouse gases” produced only 163 papers funded by fossil fuel industry in the last 5 years, and 106 papers in 2014-2019), or would not differ significantly from inclusion of word ”climate” (the terms ”carbon” and ”emissions” produced around 1000 papers for each pool, with results not differing from previous analysis). Therefore, these terms are not presented here, as further exclusion and reduction of number of papers is counterproductive. Should any further research attempt to replicate the analysis, it is advised to include countless other fossil fuel industry companies, even if each of them sponsors a very small amount of papers, for the sake of more inclusivity, as discussed further in recommendations in Chapter 5.

Figure 3.18: Largest normalised differences in term occurrences between fossil fuel and general pool of research for ”climate” term across 2014-2024 dates. X-axis displays terms, Y-axis displays their occurrence numbers for a single pool. Orange bar shows general pool, while blue one shows fossil fuel one.



3.8. Journals

An interesting note can be made on journals that are most frequently represented in each of research pools. As could be expected, these differ greatly. For instance, among Fossil Fuel Industry funded research, technical journals such as ”Journal of Petroleum Science and Engineering”, ”Marine and Petroleum Geology”, ”International Journal of Hydrogen Energy” and ”SPE Annual Technical Conference and Exhibition” were most prevalent, among others. These are mostly publications directly related to Petroleum science or Geology, or publications made at relating conferences. For both university and general pools, names are entirely different, ranging from ”Journal of Climate” to ”International Journal of Climatology”, ”Geophysical Research Letters”, or ”Science of The Total Environment”. while no more significant information, or conclusions, can be drawn from this, such differences show a general discrepancy between studies funded by fossil-fuel industry, and general research in various fields, where popularity of climate topics vastly outnumbers that of petroleum engineering.

3.9. Citation Analysis

Figures A.2-A.6 in the Appendix provide an example of citation network maps in both fossil fuel and general research pools. Maps are similar to what was noted in all other research pools. It can be noted that for fossil fuel funded research, citation networks were much smaller and less diverse than general pool, which itself was much more complicated and had less structure. In fossil fuel research, one can clearly see division between earlier and later research (as noted by different colours of author names), which can be explained by a small number of researchers involved, thus making citation maps clear to read. There seems to be no overlap between researchers of different data pools. Finally, Figure A.13 shows such clusters among university pools, which is very logical due to inclusion of 10 vastly different universities, and clusters corresponding to researchers within these institutions. No significant information could thus be extracted from such citation network map.

The disadvantage of such analysis is that one cannot note which authors outside of the data pool were cited by authors inside the data pool, only cross-referencing between them. This explains why there is no overlap of author names across data pools, be it general versus fossil fuel, or fossil fuel versus fossil fuel in different subject areas. In fact, due to small amount of fossil fuel funded research, no clusters of cross-citation can be noted, unlike in general research pools, where some authors cite each other more than rest of research (thus being grouped together far from other papers, like in Figure A6). As such, this analysis concludes there are no patterns of interest to be noted here, although future analysis recommendations are included in Chapter 5.

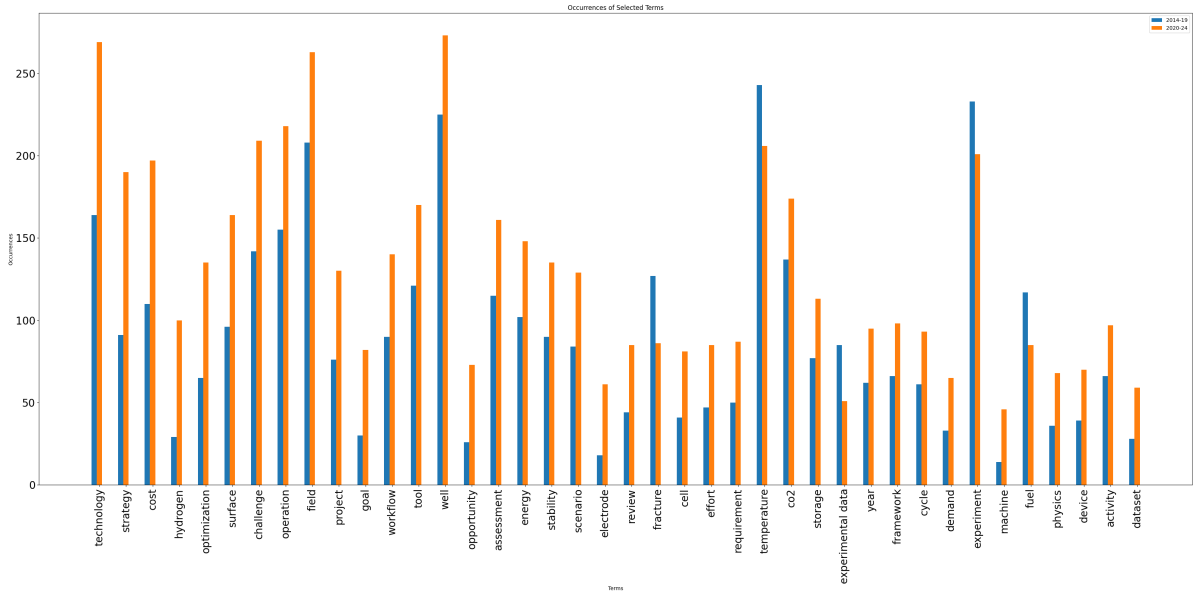
3.10. Temporal Analysis

In this section, the results of temporal analysis from all previous research pools are analysed and explained with more clarity. This analysis was performed between 2014-2019 and 2020-2024 data pools, which corresponds to a halfway cut-off date between all research analysed (within the last decade, as per research question of this thesis). This cut-off date was deemed appropriate due to similar number of papers in both data pools, and due to software ability to note trends and average publication dates down to a month in any case, making any other cut-off date unnecessary. While it was considered to include cut-off dates related to significant events (such as Paris Climate Agreement, adopted in December 2015) or European Green New Deal (adopted in December 2019), these events were either considered too early for a cut-off date of 2015, or too irrelevant (other than ”paris agreement”, no significant geopolitical or other event besides ”covid” appeared in term searches across any of data pools). ”paris agreement” term has a relatively similar showing in fossil fuel and general papers, with its average term used being 2017.5 in 2014-2019 pools (or late average), and 2021.2 in 2020-2024 pool (early average). This signifies the fading importance of an agreement, likely as a result of new regulations taking more spotlight.

Analysis in pools shows three major trends: an increase in ESG terms, increase in ”carbon capture” and ”hydrogen” terms, as well as an increase in machine learning or artificial intelligence related terms. Figure 3.2 shows such new terms in yellow, meaning that their average publication date is much later than for other terms. On the right side of the figure, various electric technology terms are prevalent, from ”battery” to ”electrocatalyst”. Much less visible, but newly appearing are terms such as ”policy” or ”decarbonization”. Similarly, newly appearing terms include ”learning”, ”deep learning”, ”training” or ”simulation model”, all with average publication date past 2021 for a pool of 2014-2024. Such trends are seen across both fossil fuel and general pools, in all date ranges - slight yet constant increase in term presence, as well as their late average publication date. Figures A.22 and A.23 in Appendix show such temporal analysis maps of General research pools, all with identical trends noted.

As an example, Figure 3.19 shows terms with largest absolute difference in their occurrence across fossil fuel energy subject area between pools of two different dates. One can note an increase in ”co2” term occurrence over the years, as well as large increase in ”hydrogen” or ”cell” terms, signifying a trend of more climate related terms, also in funded research. This trend is not as visible compared to others due to such terms being already prevalent in older research pools, but an increase in focus in evident.

Figure 3.19: Largest normalised differences in term occurrences between fossil fuel Energy subject area research pools. X-axis displays terms, Y-axis displays their occurrence numbers for a single pool. Orange bar shows 2014-2019 pool, while blue one shows papers from 2020-2024.



On the other hand, there is no such persistence in ”disappearing terms”. This means, that for terms with early average publication date, there is no consistent picture across data pools. Most often these are technical terms such as ”drilling performance” or ”molecular dynamics simulation”, other times these are general terms such as ”weight”. Such early publication date does not indicate that these terms disappear in later research, nor that it is a trend across fossil fuel, university and general pools. Rather, it usually includes dates close to average of pool publication, or terms that are not relevant for citations over the last couple of years. As it stands, no significant patterns are noted among disappearing terms.

3.11. Validation

In this Section, validation of the results is discussed, both in terms of validating software and in terms of result comparison to other similar research.

3.11.1. Topic Modelling

For the first part of validation, topic modelling is employed. It is a technique including unsupervised machine learning that provides clusters of similar words, giving an idea of whether the topics identified by VOSViewer clustering are similar to those found by the Python-Sklearn library script. The code for the script can be found in Appendix B.

One initial disadvantage of such topic modelling is that the number of clusters needs to be pre-specified by the user, hence analysis was done separately for the range of 4-8 clusters, similar to amounts seen in VOSViewer and small enough to be computationally feasible (hypertuning within this range took 16 hours to complete with relatively large 0.1 steps for alpha and beta values). While many pools of data were divided into 4-6 clusters (with some reaching even higher numbers), the number of 8 clusters was deemed to have the highest coherence score following hyperparameter tuning of Latent Dirichlet Allocation (LDA) model with varying alpha and beta parameters in a range from 0.01 to 1. LDA was chosen over other similar techniques due to its probabilistic nature and scalability for large datasets. LDA model was found optimal after its hyperparameter tuning and due to explainability with PLDavis library. In the end, alpha parameter (relating to distribution of topics within documents, influencing how many topics are represented in each document) was kept minimal at 0.01, while beta parameter (controls the distribution of words within topics, influencing how many words are important within each topic) was kept at 0.11, which allowed to maximise coherence values across clusters at around 0.35.

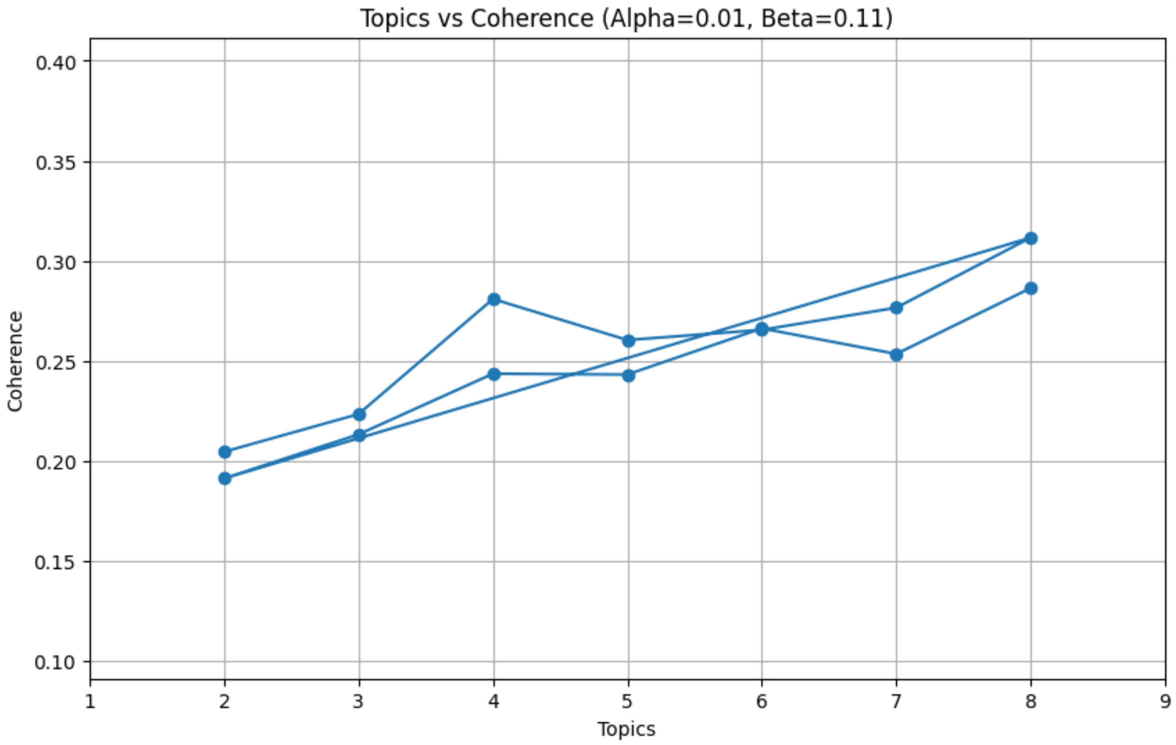
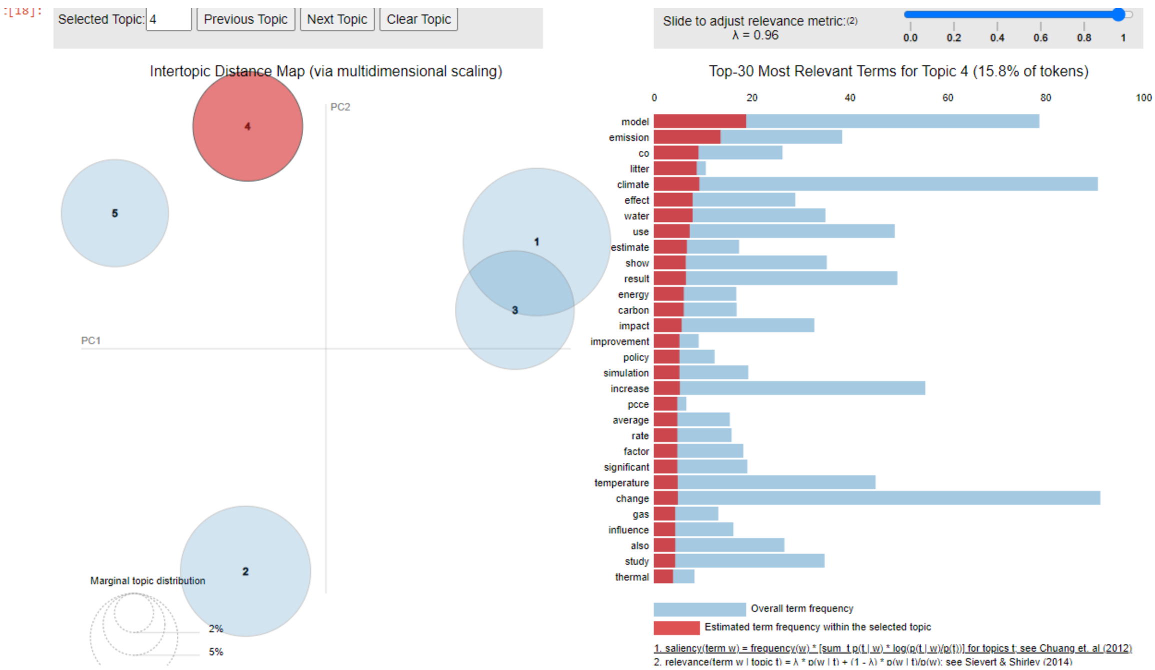


Figure 3.20: Coherence scores for number of clusters as a result of hyperparameter tuning. Optimal results are seen with the higher number of clusters, 8 in this case.

Figure 3.20 presents the coherence score results for this case of alpha and beta values. In cases of other values, overall coherence score was slightly smaller, while the results versus number of clusters stayed consistent. Nevertheless, analysis was performed using 5 and 8 values, due to them having smallest and highest coherence scores within this range, allowing to check for similarities or disparities compared to VOSViewer.

Such model for validation had one problem, namely some terms (like most frequent ones, ”climate” or ”co2”) were present in multiple clusters, as model calculated probability of those terms appearing in a cluster as highest (which comes as no surprise due to these terms having large variability and being tied to many other terms). Nevertheless, the terms that are not common in all clusters are identifiable as similar to topics in VOSViewer. Consider a case with 5 clusters, which is more similar to number of clusters in VOSViewer in the case of analysed pool. For this instance, Figure 3.21 shows top 30 terms in one of the clusters of General Energy Pool for 2014-2019 data, which roughly corresponds to the red ”policy” cluster in Figure A.7. This pool is a perfect example to showcase for validation purposes, as it also includes five clusters, with very similar overlap compared to ones computed by LDA model. One can note purple cluster (corresponding to number 3 in figure below) being a ”subset” of a larger cluster 1 (blue cluster in Figure A.7), with green cluster of ”chemistry” corresponding to cluster 2, as evident by presence of terms ”ice”, ”catalyst” or ”biodiesel”. While there are noteable differences with the VOSViewer model (often due to same terms having similar probability of being in two or more clusters), topics of each cluster are still found to be roughly corresponding to each other in almost all cases by finding distinct terms that do not belong to any other cluster. Figure 3.22 shows similar results with 8 clusters, where compared to 5 cluster analysis, clusters 1,2,3 from Figure 3.21 were broken down into smaller clusters, while clusters 4,5 of ”policy” terms were merged together due to larger similarity in terms.

In the end, such validation was performed on a random sample of 5 pools of data (General Energy 2014-2019, fossil fuel environmental 2020-2024, university energy 2020-2024, fossil fuel with term ”climate” 2020-2024 and general engineering 2014-2019). In all cases, clustering was somewhat similar to VOSViewer results, allowing for validation of software results.

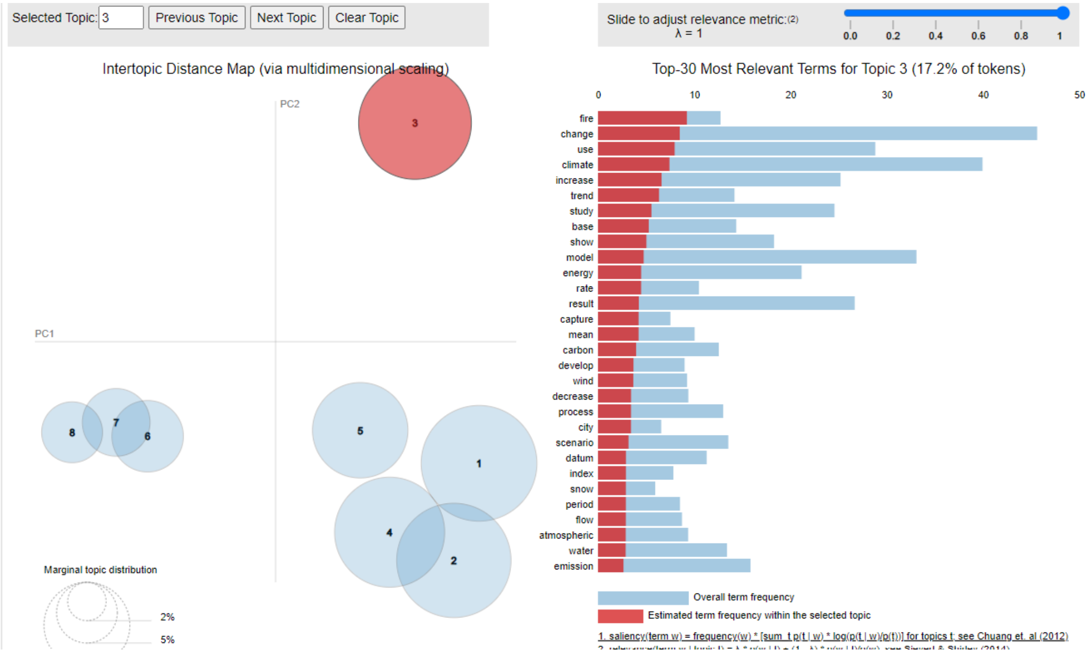


Selected Topic: 4 | Previous Topic | Next Topic | Clear Topic

Slide to adjust relevance metric:(2)
λ = 0.96

1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang et. al (2012)
2. relevance(term w | topic t) = λ * p(w | t) + (1 - λ) * p(w | t)/p(w); see Sievert & Shirley (2014)

Figure 3.21: Visualisation of a "policy" cluster within General Energy Pool of 2014-2019 within 5 clusters. Shown is cluster number 4, corresponding to General cluster of policy and stakeholders, together with its most frequent terms, including "emission", "climate" and "policy". Note that in this model, terms can occur in multiple clusters.



Selected Topic: 3 | Previous Topic | Next Topic | Clear Topic

Slide to adjust relevance metric:(2)
λ = 1

1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang et. al (2012)

Figure 3.22: Visualisation of a "policy" cluster within General Energy Pool of 2014-2019 within 8 clusters, showing cluster 3 terms of "General" cluster. terms like "policy", "trend", "climate" or "city" are of high relevance here, indicating the cluster topic.

3.11.2. Literature Validation

There are a couple of research papers available that have performed similar research in other industries, and whose findings are compared to findings of this research in this section. It is important to note, however, that almost all of such research is within the scope of medicine and biochemistry, hence their findings, while demonstrating plausible industry-funding behaviour, may be different in industry specifics compared to fossil fuel industry research.

For start, a 2012 review in PLOS Medicine assessed the quality of reporting in industry versus non-industry-funded studies. It found that industry-funded studies often had better reporting of methodological details but were more likely to exhibit biases in study design and reporting of results [2]. Such findings are consistent with what was identified within this research, where methodological details are much more technical and practical than general research, but the choice of topics exhibits certain biases towards topics that are of particular interest for the industry.

Rather similar methodological finding was found in research published in Health Affairs (2015), that found that non-industry-funded studies are more likely to conduct comparative effectiveness research, which compares different interventions to determine which works best. Industry-funded research, however, tends to avoid direct comparisons with competitors' products, focusing instead on demonstrating the efficacy of their own products [35]. While this does not directly relate to fossil fuel industry, such finding is consistent with diminished amounts of ”review” methodology found in topics of industry funded research.

Industry-funded research tends to focus on marketable products and treatments that can lead to profitable outcomes. A study published in 2011 in BioSocieties found that pharmaceutical companies prioritize research on drugs and treatments with high commercial potential, often neglecting less profitable areas such as rare diseases or non-pharmaceutical interventions [32]. This may correlate with findings on ”carbon capture” and ”hydrogen” term prevalence, though not necessarily due to profitability, but due to industry priority in terms of climate mitigation.

Similarly, a review in The Lancet (2009) showed that non-industry-funded research is more likely to address broader public health issues, including prevention, health systems, and diseases prevalent in low-income settings [9]. In contrast, industry-funded studies often align with the commercial interests of the sponsors, potentially leading to a misalignment between research efforts and public health needs. This again corresponds with findings that industry does not necessarily align with broader public interest, rather focusing on their own commercial interest. According to this study, non-industry-funded research is more likely to include health services and policy research, which examines the organization, delivery, and financing of healthcare. Industry funding rarely supports this type of research, as it does not directly contribute to product development. Once again, the finding is consistent with conclusion of this research, that industry funded climate research does not focus on policy or governance topics, rather focusing on directly fossil fuel industry related products.

In conclusion, industry research in the fields of medicine and pharmacology show consistent industry behaviour with what was noted in this research relating to fossil fuel industry, in all three findings - lack of interest in policy, more direct and less comparative methodology, as well as more interest in commercially profitable/beneficial to the industry products.

4 Discussion

In this Chapter, the results of analysis from Chapter 3 are discussed, along with potential explanations, assumptions and further analysis required to develop further information based on insights gained. After more than 100,000 research papers have been analysed, concerning research in four different subject areas, and research strictly related to climate, both funded by fossil-fuel industry and not, certain conclusions can already be made about the results and alignment of fossil-fuel industry funding with goals of broader research community. The analysis, despite its limitations that are discussed further in this chapter, is able to shed a light onto the bigger picture of fossil-fuel industry influence on research, where its interests lie, leading to ability to draw conclusions of whether this state of research funding is desireable, or whether some changes are necessary to address certain shortcomings. The results are therefore significant in establishing context for further discussions on whether and how to change the relationship between the industry and climate-change researching academia community, given the established conflict of interest.

4.1. Research Questions

Firstly, results are used to answer the research questions, specifically subquestions 1-3, as the other subquestion and final main research question are answered in the section 4.2 below. First research sub-question was concerning the methodology employed by fossil-fuel industry funded research, as opposed to general research patterns. As seen in Figure 3.16 in the previous chapter, fossil-fuel industry funded research has a considerably smaller amount of ”non-technical” research methods employed, such as ”survey” or ”review”. Similarly, ”life-cycle” related assessment terms are considerably lacking in the aforementioned research. However, no significant differences were seen in ”technical” aspects, such as machine learning research techniques, or field-related research such as ”sample” or models of various techniques. Such differences lead to a conclusion that research funded by fossil fuel industry are mostly concerned with direct technical aspects relating to their operations, and are much less concerned with perception of their work, consequences or end-of-life aspects of their products. However, a potential counter-explanation may lie in the fact that theoretical review or survey-based research requires much less funding, therefore requiring the help of the industry much less. While the answer to this research subquestion sheds some insight, it alone cannot lead to any certain conclusions.

Such conclusions, however, can be drawn more clearly after answering the second subquestion, which delves more into terms that are differing between fossil-fuel funded and general research pools. While most of these results are discussed along with patterns in section 4.2, it can shortly be mentioned here that fossil-fuel industry funded research has a significantly smaller amount of ”ESG” terms mentioned, such as ”policy”, ”governance”, ”stakeholder”, or many smaller ones, ranging from ”farmer” to ”community”. Such lack of terms supports the notion that industry funding is concerned mostly with technical aspects of research and operations, and much less with climate-change consequences, impact on livelihood around areas of industry operations or more globally, or even by what potential mitigation strategies can be employed, be it by industry itself or governing bodies. Another interesting notion, discussed in the section below, is a more frequent mention of ”carbon capture” and ”hydrogen” terms by fossil fuel funded research, indicating industry's interest in these two technologies. Such notion was to be expected, as many people have reported on the surge of interest in these technologies by fossil fuel industry. [12] Potential explanations for that lie in the fact that both technologies allow for continuation of operations for the industry, either by mitigating their effects in the case of carbon capture, or by reusing some infrastructure and taking an early lead in emerging market, as in the case of hydrogen. Many other explanations, from tax cuts to government subsidies can be mentioned, yet what is absolutely clear from this study, is that unlike solar, wind or any other sustainable energy source, hydrogen and carbon capture technology are the two mitigation techniques that are more frequently used by fossil fuel industry than by general research.

Finally, a third subquestion was concerning indirect funding by fossil fuel industry in the form of university research. For this reason, research from 10 universities accused of serious financial ties to fossil fuel industry was analysed, as explained in Chapter 2. However, if any correlation exists, it has not been noticed in this research. University research tended to be versatile in terminology and methodology, and its results were much more similar to general research pools than to the one funded by fossil fuel industry. While differences did exist, they would vary between time and various subject areas, most likely due to focus of specific researchers in specific universities on one area rather than the other, yet have not produced anything comparable to fossil fuel funded research. Therefore, this research subquestion can be answered that indirect funding research differs significantly from direct funding, and there is no conclusive way that can show that such funding has ever significantly affected research results, at least with the method employed in this research.

4.2. Patterns Identified

As discussed in Chapter 3, while university pool has seen no significant differences in terminology and methodology compared to general pool of research, the situation with papers funded by fossil fuel industry is different. Differences exist in both terminology and methodology, and those differences persist across various subject areas, term searches and time frames. Therefore, it is possible to identify those as patterns, which are further discussed here.

First pattern identified, or rather a lack thereof, are terms related to policy, decision-making and risk mitigation, as well as less diversification of stakeholders. While a term ”policy” became more frequent when the term ”climate” was included, it is still much less diverse than in general research pool, which includes dozens of different terms related to policy, which occur much more throughout research. Similarly, terms related to smaller ”stakeholders” appear much less - other than ”stakeholder”, this includes terms like ”farmer”, ”local community”, ”business owner” etc. Terms including ”decision-making”, ”local government”, ”climate risk”, ”governance”, or ”responsibility” either do not occur entirely, or are represented in one of analysed fossil fuel pools by a bare minimum amount of occurrences (10 - 15 times across the pool). Special consideration here goes towards term ”risk”, which comes across general pools of data in many forms, ranging from ”climate risk”, to ”risk mitigation”, ”risk management', ”drought risk” etc. All together, these terms appear 500-1000 times across each general pool of data analysed, while in fossil fuel pool of data it becomes roughly around 50-100. Even after normalising the scores for the amount of papers involved, this is still a much smaller occurrence. Other terms related to climate change, such as ”carbon emission” or ”co2 emission”, as well as ”life cycle assessment”, ”adaptation” or ”reduction” occur less frequently among fossil fuel industry pools as well.

Second pattern identified relates to methodology, though it is much easier to explain. In this regard, fossil fuel industry papers include much less terms such as ”survey”, ”review” or ”hypothesis”. When it comes to methodologies directly related to the subject area, be it ”machine learning”, ”field experiment”, ”theoretical study”, ”spectroscopy” etc., frequency of occurrence is similar, or higher in fossil fuel industry research. As discussed in chapter 3, this indicates a pattern of funding papers directly related to the field of operations of companies, rather than studies on the effect of those, though this will be elaborated upon in the next section.

Third pattern relates to terms that are found more frequently across pool of fossil fuel funded research than across general pools of research. The main term in this regard is ”carbon capture”, occurring roughly twice more frequently in fossil fuel funded energy research as well as in research undivided between subject areas, and 50 % more frequent among environmental science. Similarly, term ”hydrogen” occurs in fossil fuel industry funded research roughly 30-50 % more frequently, along with its related terms, such as ”co2 hydrogenation”, ”hydrogen production” or ”hydrogen sulfide”. While there are more terms that occur more frequently in fossil fuel research in one pool or another, there are no significant patterns there to be found that would continue among various pools of data consistently.

4.3. Literature

As already discussed in 3 validation section, findings appear to be consistent with some of the literature that can be found on this topic. While most of these sources, like Lancet [9] or BioSocieties report [32] have studied mostly other industries, like pharmacy and biotechnology, they have found that studies funded by industry tend to focus on what is profitable for the industry, and less on anything else, which can be considered a bias. Similarly here, heightened interest in hydrogen or carbon capture technology, together with interest in mostly petroleum related studies, signifies that fossil fuel industry does not consider funding of climate consequences or ESG risks important. Unlike in tobacco industry, where results of studies are clearly affected by funding [5], this research in its current form does not support any claims that results of studies were impacted by the funding from an impartial source.

Since such literature assessment has been used as a partial validation tool for study results, some of its limitations need to be addressed. One of them is of course the fact that it is impossible to read and check every source in every language, meaning that there may have been sources that contradict or support the findings of this thesis further, which have not been addressed. Moreover, while none have been found, similar scientometric studies could have been performed on fossil fuel industry, in which case comparison with this thesis would be of significant importance. Finally, none of literature findings in terms of other industries, be it tobacco or pharmacology, can act as evidence or validation of claims that fossil-fuel industry funded research has been maliciously affecting the studies or their topics.

4.4. Reflection

The observed differences in fossil fuel industry funded research compared to general research can be attributed to several varying factors, which include the strategic interests of the fossil fuel industry, the nature of industry-driven research, and the specific priorities of these companies. These are examined here for each of three identified patterns.

Firstly, there is a possibility of fossil fuel companies avoiding emphasizing terms related to policy and governance because these areas often highlight the need for stricter regulations and oversight, which can be perceived as threats to their business models. By downplaying topics related to governance and risk mitigation, the industry might aim to shape public and policy-maker perceptions, suggesting that their activities are less risky or problematic than they might be portrayed by independent researchers. [4] While this is a logical and intuitive insight, it is difficult to prove malicious intent, therefore it is listed here as one potential reason, with more research required into whether or not this is the actual case. Another, less malicious explanation could be that the industry may prioritize research that leads to immediate technological and economic benefits rather than research that could lead to increased regulation or governance changes, or research that simply focuses on long-term climate-related risks with no immediate economic benefit of such research to the companies involved.

Having said that, the conspicuous absence of ”life cycle assessment” (LCA) and pollution-related terminology in research funded by fossil fuel companies raises significant concerns about the objectivity and comprehensiveness of such studies. LCA is a crucial tool for evaluating the environmental impacts of a product or process from cradle to grave, encompassing all stages of production, use, and disposal. By neglecting these critical assessments, fossil fuel-funded research may overlook or deliberately downplay the full extent of environmental damage caused by fossil fuel extraction, processing, and consumption. This omission not only skews public perception but also hampers the development of more sustainable energy solutions by failing to provide a transparent and holistic understanding of the environmental costs associated with fossil fuels. While reasons for such omission of terms are not clear, their results are, and more focus should be paid by fossil fuel industry to mitigation and rehabilitation of consequences of their activities.

In terms of methodology, the explanation of ”direct-involvement-only” research is prevalent. Industry funded research often aims at practical solutions that can be implemented quickly to improve processes, reduce costs, or enhance efficiency. This contrasts with academic research, which can afford to be more exploratory and theoretical. Moreover, practical methodologies can lead to quicker implementation and commercialization of new technologies or processes, aligning with business goals of short-term returns on investment.

Finally, the prevalence of ”carbon capture” and ”hydrogen” related technologies needs to be discussed. As temporal analysis has shown, it is a recent and growing trend, that requires a lot of practical research, rather than theoretical, which could therefore lead to prevalence among fossil fuel funded industry research for reasons similar to practical methodology. However, one can also attribute such findings to increasing pressure to reduce industry carbon footprint. Emphasizing research on carbon capture and hydrogen production might help the industry to demonstrate commitment to addressing climate change while continuing its core business operations, or it could be a hope of some companies, that such research could help them reduce costs of changing their business model in the medium-to-long term. Even if such technologies are not entirely the focus of the industry, investing in hydrogen production and carbon capture could allow these companies to diversify their energy portfolios and prepare for a future where renewable energy sources become more dominant.

Ultimately, it is difficult to assess the specific intent behind the funding of such topics and patterns by the fossil fuel industry. The policy of such funding may differ from company to company, all of which may pursue different strategic interests. Without information on decision-making in such companies, patterns identified cannot be assessed qualitatively on the reasons of their occurrence, though several reasons listed above can be perceived as most likely explanations.

4.5. Influence

Finally, while this research does not investigate anything relating to influence of fossil fuel industry funded papers on policy and governance, the discussion focuses on the potential influence of patterns identified. Such potential influence needs to be addressed, as there is a clear risk that decisions are made on research that has been funded by a party with a conflict of interest. After all, such research has been found to be used for lobbying of decision-making [15], shaping the policies or delaying their implementation [3]. Similarly, research that aligns with fossil fuel industry narratives can reinforce the idea that fossil fuels can continue to play a major role in the energy mix if coupled with technological solutions. This can undermine efforts to promote renewable energy sources and more sustainable consumption patterns. For instance, carbon capture technology has long been accused of being a ”PR campaign of fossil fuel industry” aimed at creating the idea that fossil fuel phaseout can be prevented by large-scale CO2 capture from the atmosphere [18]. Similar accusations follow the hydrogen technology, with fossil fuel industry companies being able to use their existing infrastructure to capture the market early on and solidify their influence [21].

While the lack of some terms in fossil fuel industry funded research points towards lack of focus, which is therefore less relevant in terms of influencing public or policy-making opinion, it is the terms that are frequent in such research that raise concerns. Further research might thus be necessary into research specifically about ”carbon capture” and ”hydrogen” in order to discover differences in results among such research, and where such research is being cited and used. This report offers no insights into policy-making or whether fossil fuel industry funded research influences such policy-making, yet its findings may well be useful for any further research focusing on it.

4.6. Risks and Limitations

Given that the methodology of this research is present in a vast amount of research, well-documented and with sufficient experience on the part of research supervisors, not many risks lie within the methodology, though there are significant limitations. For instance, relying solely on quantitative metrics (e.g., h-index, impact factor) can overlook qualitative aspects of research impact and innovation. This may explain results in methodology differences of data pools, by including many theoretical or survey-based research in the general pool. Additionally, such quantitative focus puts emphasis on citation metrics can encourage practices like ”salami slicing” (publishing many small papers from one study) or prioritizing quantity over quality. Methodology itself cannot look directly into papers due to their vast amount, therefore potentially missing some context of research, and direct comparison between industry funded and general pool of research inherently includes research that is not directly related to either climate-change studies or industry operations, potentially adding many unnecessary terms in general pool of research.

Similarly, limitations exist within citation analysis, which proved to be of low importance in this particular research. Accurately attributing contributions in multi-author papers can be challenging, affecting individual assessments. Adversely, collaborative papers often receive more citations, which might not reflect individual contributions accurately. Citations can be influenced by factors unrelated to the quality of the research, such as the author's reputation or journal impact factor, and their practices vary significantly between disciplines, making cross-disciplinary comparisons difficult. All such considerations have resulted in a risk of having no significant results from citation analysis, which is recommended for further research.

Having said that, considerable risks exist within data search. For instance, research funded by the fossil fuel industry may have concealed acknowledgments, or be of insufficient quantity to fully assess the patterns emerging. Extension of inclusion criteria, both in terms of time range and range of research, as well as previously mentioned utilization of whistle-blowing organization data, seem like a good source of mitigation, yet the data available is still considerably lower in quantity.

The analysis may be constrained by the availability and accuracy of funding information in published articles. To this point, there is not much to achieve, as anything that is not sourced as funded by fossil fuel industry is almost impossible to detect on a mass scale, requiring a full investigation if such hiding takes place.

Additional risks exist in the results, which failed to deliver any identifiable patterns, specifically in the university pool of data. As scientometric analysis produces similar results for both data pools, this could serve as a conclusion that no biases are found. However, there is a chance that differences may be insignificant or unable to present a clear picture on whether any pattern exists. In such a case, this is still considered a significant contribution to research on this question, and further research would be recommended to build-up upon the findings. Any funding ties to institutions, rather than researchers, is vague and has not lead to any causation links found. Ties to specific researchers as opposed to institutions can be investigated, though finding such researchers is another problem altogether.

Rather importantly, the number of papers found with proven ties to the fossil fuel industry is considerably lower than the pool of general data, which may lead to unsatisfactory confidence in the results. For instance, any research in specific subject areas was already showing much less keyword available than general data, and attempting to specify terms further would lead to even less terms shown.

The study is also limited to the scope of publicly available research data and may not capture unpublished or proprietary research. Additionally, a limitation of having little research to compare it to was already discussed in chapter 2. Final thing to add is limitations on the validation method. Topic modelling, as a technique used to validate software, is essentially a simplified model of VOSViewer software, therefore results of validation were expected, and only validate the software rather than the method itself. This was attempted to overcome via comparison to other literature, as discussed in Chapter 3 as well.

5 Conclusion

In this Chapter, findings are summarised, limitations are acknowledged, and recommendations for further research are presented.

5.1. Summary of findings

As discussed in Chapters 3 and 4, the fossil fuel industry research was found to pay less attention to policy, governance, risk mitigation (ESG terms) and lower stakeholders, which is likely tied to a finding of their methodology being mostly focused on technical processes, rather than academic and theoretical research. Additionally, a large discrepancy was found in research into carbon capture and hydrogen technologies, something that was already noted by many observers before, though with no academic proof. Such findings show focus of the industry funding, and while unclear in terms of their intent, may already be used as to inform of lack of interest, or bias towards, certain aforementioned aspects of research. Temporal analysis has shown the interest in carbon capture and hydrogen technologies increase over time, both in fossil fuel and general research pools, therefore more attention is necessary in this regard. It has also shown a trend common through all data pools of interest growing in all things electricity, as well as machine learning and it's varying techniques. Such trends have been very consistent and are likely to stay so for the foreseeable future. The findings are also very consistent with existing literature, or rather media reports about fossil-fuel industry funding, as discussed in Chapter 4.

On the contrary, universities with large fossil fuel funding have not had any clear patterns identified, having a lot of overlap with general research pools of data. Such findings are not surprising, since generally large universities have a lot of researchers, and most of research may have been performed without any industry involvement. However, more investigation is needed into specific researcher ties with fossil fuel industry, which was outside of scope for this research.

5.2. Recommendations

As already mentioned, further research recommendations include looking into ties of specific researchers with fossil fuel industry, as well as monitoring any whistle-blowing information on potential research that is not openly sponsored by fossil fuel industry. Additionally, a selected pool of papers could be analysed in their entirety, in order to check whether any patterns arise within their text.

Citation analysis could be expanded to see cross-pool ties (for instance, which fossil fuel funded researchers cite papers from general pool), as well as isolated to further check contents of papers from some specific researchers with most citations. This was deemed out of scope for current research.

Enhanced data collection could include not only research articles, but also patents, technical reports and other types of documents. An expansion to many more fossil fuel industry companies may also prove insightful, just as the analysis in languages other than English.

Another interesting point of future research could be comparing fossil fuel industry funding to other funding, rather than general pools. This could be a comparison with governmental funding (general, or of a specific state), funding from large climate-centric organisations, or other industry-specific actors. Similarly, one could break down fossil fuel industry funding, in order to check whether the findings differ between oil- and gas- primary companies, or between companies of different states. In this report, research was mostly analysed by oil companies, as those were most influential and reported on, and for reasons discussed in Chapter 2. Further breakdown of those companies could pinpoint some further issues.

Finally, additional research could focus on isolated parts of patterns found or other terms, by for instance analysing how notable events such as Paris Climate Accords have affected or shifted the focus of research, both within and outside of industry funded research scope.

References

[1] World Meteorological Organization (WMO). “The Global Climate 2011-2020: A decade of accelerating climate change”. In: *WMO* (2023). DOI: https://library.wmo.int/idurl/4/68585.

[2] A. Lundh et. al. “Industry sponsorship and research outcome.” In: *The Cochrane database of systematic reviews, 12* (2012). DOI: https://doi.org/10.1002/14651858.MR000033.pub2.

[3] Raquelle. Bañuelos. “Oil Spill: How Fossil Fuel Funding Corrupts British Cultural Institutions”. University of Illinois, 2021. URL: https://www.proquest.com/openview/4011a0e6ea642d61e0161c550a793eab/1?pq-origsite=gscholar&cbl=18750&diss=y.

[4] Noam Bergman. “Impacts of the Fossil Fuel Divestment Movement: Effects on Finance, Policy and Public Discourse”. In: *Sustainability* 10.7 (2018). ISSN: 2071-1050. URL: https://www.mdpi.com/2071-1050/10/7/2529.

[5] Deborah E. Barnes; Lisa A. Bero. “Industry-Funded Research and Conflict of Interest: An Analysis of Research Sponsored by the Tobacco Industry Through the Center for Indoor Air Research”. In: *J Health Polit Policy Law* (1996). DOI: https://doi.org/10.1215/03616878-21-3-515.

[6] Eric Bonds. In: *Human Ecology Review* 22.2 (2016), pp. 3–23. URL: https://search.informit.org/doi/10.3316/informit.324306883375423.

[7] R.J. Brulle. “The climate lobby: a sectoral analysis of lobbying spending on climate change in the USA, 2000 to 2016.” In: *Climatic Change* 149 (2018), pp. 289–303. DOI: 10.1007/s10584-018-2241-z.

[8] Talha Khan Burki. “Conflicts of interest in tobacco industry-funded research”. In: *Lancet Oncology* 22 (2021 June). URL: https://doi.org/10.1016/S1470-2045(21)00281-3.

[9] P. Chalmers I. Glasziou. “Avoidable waste in the production and reporting of research evidence.” In: *Lancet, 374(9683)* (2009). DOI: https://doi.org/10.1016/S0140-6736(09)60329-9.

[10] United Nations Framework Convention on Climate Change (UNFCCC). *The Science Behind Climate Change*. 2023. URL: https://unfccc.int/ (visited on 07/22/2024).

[11] Inci Sayki Jimmy Cloutier. *Oil and gas industry spent* $124.4millionon federallobbyingamidrecordprofitsin2022$. 2023. URL: https://www.opensecrets.org/news/2023/02/oil-and-gas-industry-spent-124-4-million-on-federal-lobbying-amid-record-profits-in-2022/ (visited on 02/22/2023).

[12] Jimmy Cloutier. *Burgeoning hydrogen industry draws* $41millioninfederallobbyingfromfossilfuelcompanies$. 2023. URL: https://energynews.us/2023/12/12/burgeoning-hydrogen-industry-draws-41-million-in-federal-lobbying-from-fossil-fuel-companies/ (visited on 12/12/2023).

[13] CoastAdapt. *What are RCPs*. URL: https://coastadapt.com.au/infographics/what-are-rcps (visited on 05/24/2024).

[14] Ilana Cohen. *Why Are Fossil Fuel Companies Funding Climate Change Research?* 2022. URL: https://www.thenation.com/article/politics/oil-gas-university-greenwashing/ (visited on 08/15/2022).

[15] Jonathan S. Coley and David J. Hess. “Green energy laws and Republican legislators in the United States”. In: *Energy Policy* 48 (2012). Special Section: Frontiers of Sustainability, pp. 576–583. ISSN: 0301-4215. DOI: https://doi.org/10.1016/j.enpol.2012.05.062. URL: https://www.sciencedirect.com/science/article/pii/S0301421512004752.

[16] “Corporate funding and ideological polarization about climate change”. In: *PNAS* (2015). DOI: https://doi.org/10.1073/pnas.1509433112. URL: https://www.pnas.org/doi/full/10.1073/pnas.1509433112.

[17] Alice Fabbri et.al. “The Influence of Industry Sponsorship on the Research Agenda: A Scoping Review”. In: *American Journal of Public Health* 108 (2018). URL: https://doi.org/10.2105/AJPH.2018.304677.

[18] Jonathan Foley. “Opinion: Don’t Fall for Big Oil’s Carbon Capture Deceptions”. In: *Scientific American* (2023). URL: https://www.scientificamerican.com/article/dont-fall-for-big-oils-carbon-capture-deceptions/.

[19] France24. *Sultan al-Jaber under fire: COP28 president accused of conflict of interest*. 2023. URL: https://www.france24.com/en/tv-shows/middle-east-matters/20231206-sultan-al-jaber-under-fire-cop28-president-accused-of-conflict-of-interest (visited on 12/06/2023).

[20] et.al. Frumhoff P.C. “The climate responsibilities of industrial carbon producers.” In: *Climatic Change* 132 (2015), pp. 157–171. DOI: 10.1007/s10584-015-1472-5.

[21] Gabriel Gavin. *Fears grow that fossil fuel firms will capture booming hydrogen industry*. 2023. URL: https://www.politico.eu/article/fear-fossil-fuel-firm-capture-hydrogen-industry/ (visited on 12/08/2023).

[22] Milad Haghani. “What makes an informative and publication-worthy scientometric analysis of literature: A guide for authors, reviewers and editors”. In: *Transportation Research Interdisciplinary Perspectives* 22 (2023), p. 100956. ISSN: 2590-1982. DOI: https://doi.org/10.1016/j.trip.2023.100956. URL: https://www.sciencedirect.com/science/article/pii/S2590198223002038.

[23] Milad Haghani et al. “Road safety research in the context of low- and middle-income countries: Macro-scale literature analyses, trends, knowledge gaps and challenges”. In: *Safety Science* 146 (2022), p. 105513. ISSN: 0925-7535. DOI: https://doi.org/10.1016/j.ssci.2021.105513. URL: https://www.sciencedirect.com/science/article/pii/S0925753521003568.

[24] Julian Hartman-Sigall. *Do fossil fuel funders impact research? Researchers say it’s the other way around.* 2023. URL: https://www.dailyprincetonian.com/article/2023/03/princeton-university-dissociation-exxonmobil-researches-defend-funding-from-bp-divest-pushes-for-full-dissociation (visited on 03/29/2023).

[25] Fiona Harvey. “Universities must reject fossil fuel cash for climate research, say academics”. In: *The Guardian* (2022). URL: https://www.theguardian.com/science/2022/mar/21/universities-must-reject-fossil-fuel-cash-for-climate-research-say-academics#:~:text=Rowan%20Williams%2C%20the%20former%20archbishop,on%20them%20to%20reject%20all.

[26] Martina Igini. *Azerbaijan Appoints Ecology Minister and Ex-Oil Executive to Lead COP29*. 2024. URL: https://earth.org/azerbaijan-appoints-ecology-minister-and-ex-oil-executive-to-lead-cop29/#:~:text=Mukhtar%20Babayev%27s%20appointment%20to%20preside,held%20in%20Baku%20in%20November (visited on 01/08/2024).

[27] Martina Igini. *Italian Oil Firm ENI Sued For ‘Lobbying and Greenwashing’ For More Fossil Fuels Despite Knowing the Risks*. 2023. URL: https://earth.org/eni-italy-lawsuit/ (visited on 05/11/2023).

[28] Y. Jeong D. Koo. “Analysis of Trend and Convergence for Science and Technology using the VOSviewer.” In: *International Journal of Contents* (2016). DOI: 10.5392/ijoc.2016.12.3.054.

[29] Liam Knox. *Fossil Fuel Industry Gave Hundreds of Millions to Higher Ed*. 2023. URL: https://www.insidehighered.com/quicktakes/2023/03/06/fossil-fuel-industry-gave-hundreds-millions-higher-ed (visited on 03/05/2023).

[30] Viktoriya Lantushenko and Carolin Schellhorn. “The rising risks of fossil fuel lobbying”. In: *Global Finance Journal* 56 (2023), p. 100829. ISSN: 1044-0283. DOI: https://doi.org/10.1016/j.gfj.2023.100829. URL: https://www.sciencedirect.com/science/article/pii/S1044028323000248.

[31] Jie Li, Floris Goerlandt, and Genserik Reniers. “An overview of scientometric mapping for the safety science community: Methods, tools, and framework”. In: *Safety Science* 134 (2021), p. 105093. ISSN: 0925-7535. DOI: https://doi.org/10.1016/j.ssci.2020.105093. URL: https://www.sciencedirect.com/science/article/pii/S0925753520304902.

[32] R. Light D. Warburton. “Demythologizing the high costs of pharmaceutical research.” In: *BioSocieties, 6* (2011). DOI: https://doi.org/10.1057/biosoc.2010.40.

[33] Juliet Ferguson Chris Matthews. *European universities accept €260 million in fossil fuel money*. 2023. URL: https://www.investigate-europe.eu/posts/european-universities-accept-260-million-euros-fossil-fuel-money (visited on 11/28/2023).

[34] Juliet Ferguson Chris Matthews. *European universities accept €260 million in fossil fuel money*. 2023. URL: https://www.investigate-europe.eu/posts/european-universities-accept-260-million-euros-fossil-fuel-money (visited on 11/28/2023).

[35] Huseyin Naci and John Ioannidis. “How Good Is ”Evidence” from Clinical Studies of Drug Effects and Why Might Such Evidence Fail in the Prediction of the Clinical Utility of Drugs?” In: *Annual review of pharmacology and toxicology* 55 (Aug. 2014). DOI: 10.1146/annurev-pharmtox-010814-124614.

[36] Grace Nosek. “Climate discourse polluted: a cumulative effects analysis of the fossil fuel industry’s tactics to influence public discourse”. PhD thesis. University of British Columbia, 2023. DOI: http://dx.doi.org/10.14288/1.0431101. URL: https://open.library.ubc.ca/collections/ubctheses/24/items/1.0431101.

[37] OpenSecrets. *Oil Gas Lobbying*. 2024. URL: https://www.opensecrets.org/industries/lobbying?cycle=2024&ind=E01 (visited on 04/20/2024).

[38] Naomi Oreskes and Erik M. Conway. *Merchants of Doubt: How a Handful of Scientists Obscured the Truth on Issues From Tobacco Smoke to Global Warming*. Bloomsbury Press, 2010.

[39] IPCC Sixth Assessment Report. *Climate Change 2022: Impacts, Adaptation and Vulnerability*. 2022. URL: https://www.ipcc.ch/report/ar6/wg2/ (visited on 07/22/2024).

[40] Maria Sharmina. “Academia–industry ties under scrutiny”. In: *Nature Climate Change* 12 (2022), pp. 1086–1087. URL: https://doi.org/10.1038/s41558-022-01522-2.

[41] Molly Taft. *Research or Lobbying? New Documents Reveal What Fossil Fuel Companies Are Really Paying for at Top Universities*. 2024. URL: https://drilled.media/news/hearingdocs-universities (visited on 04/30/2024).

[42] Paul D Thacker. “Stealing from the tobacco playbook, fossil fuel companies pour money into elite American universities”. In: *BMJ* 378 (2022). DOI: 10.1136/bmj.o2095. eprint: https://www.bmj.com/content/378/bmj.o2095.full.pdf. URL: https://www.bmj.com/content/378/bmj.o2095.

[43] “The growth of climate change misinformation in US philanthropy: evidence from natural language processing”. In: *Environmental Research Letters* 14 (2019). DOI: 10.1088/1748-9326/aaf939. URL: https://iopscience.iop.org/article/10.1088/1748-9326/aaf939.

[44] Leiden University. *Leiden University publishes list of research partnerships with the fossil fuel industry*. 2023. URL: https://www.universiteitleiden.nl/en/news/2023/08/leiden-university-publishes-list-of-research-partnerships-with-the-fossil-fuel-industry (visited on 05/28/2024).

[45] Utrecht University. *Utrecht University Transparency Research Collaboration Fossil Industry*. 2023. URL: https://www.uu.nl/en/research/research-at-utrecht-university/transparency-research-collaboration-fossil-industry (visited on 05/28/2024).

[46] Nees Jan Van Eck Ludo Waltman. *VOSViewer Manual*. 2023. URL: https://www.vosviewer.com/documentation/Manual_VOSviewer_1.6.20.pdf (visited on 10/31/2023).

[47] Jenny B. White and Lisa A. Bero. “Corporate Manipulation of Research: Strategies Are Similar Across Five Industries”. In: *Stanford law and policy review* 21 (2010), p. 105. URL: https://api.semanticscholar.org/CorpusID:166860303.

[48] Wikipedia. *List of largest oil and gas companies by revenue*. 2024. URL: https://en.wikipedia.org/wiki/List_of_largest_oil_and_gas_companies_by_revenue (visited on 05/20/2024).

A

Figures

Figure A.1: Most frequent terms of 2014-2019 pool for Fossil Fuel Industry Energy-related papers. Size of bubbles indicate the term occurrences, while colour indicates average publication date, with brighter colours showing more recent terms, with electrical energy and machine learning topics being among brightest.

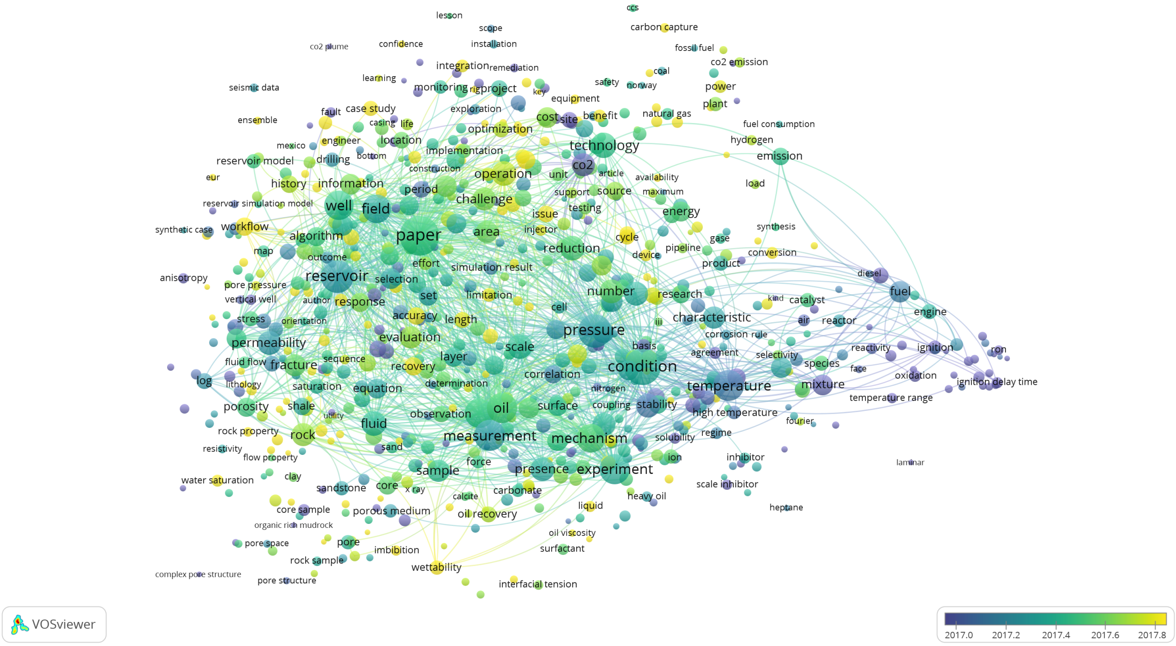


Figure A.2: Most frequent terms of 2020-2024 pool for Fossil Fuel Industry Energy-related papers. Size of bubbles indicate the term occurrences, while colour indicates average publication date, with brighter colours showing more recent terms, with electrical energy and machine learning topics being among brightest.

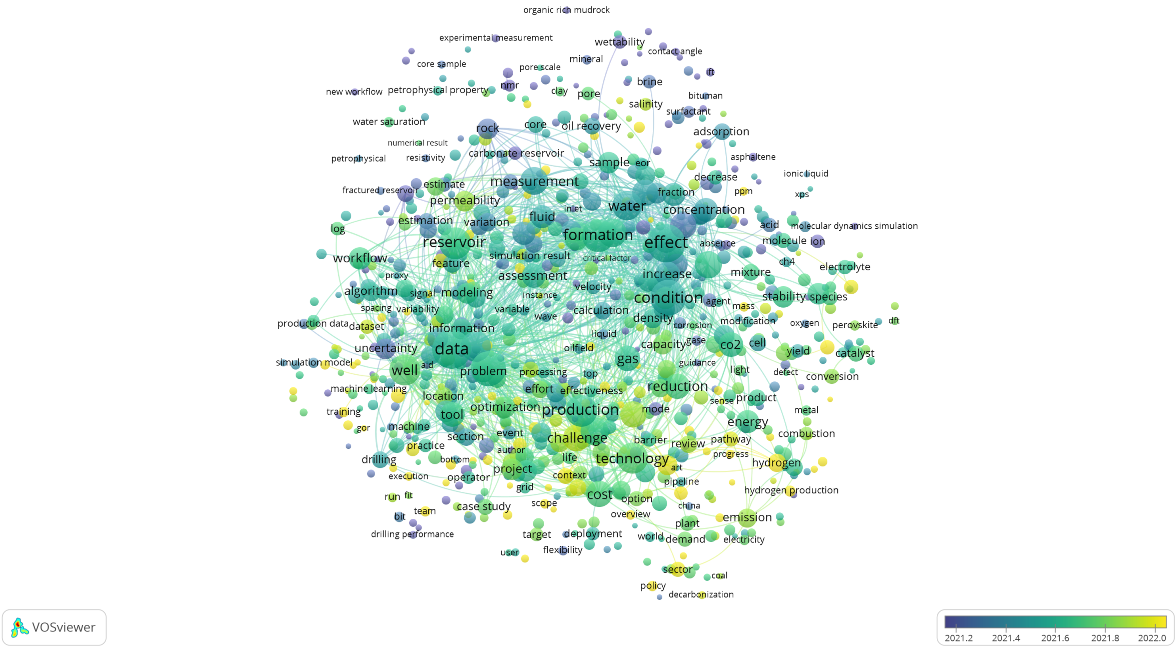


Figure A.3: Most frequent citations of 2014-2019 pool for Fossil Fuel Industry Energy-related papers. Colours indicate average publication date, and the graph is neatly divided into earlier and later research, with later research having lower citation scores.

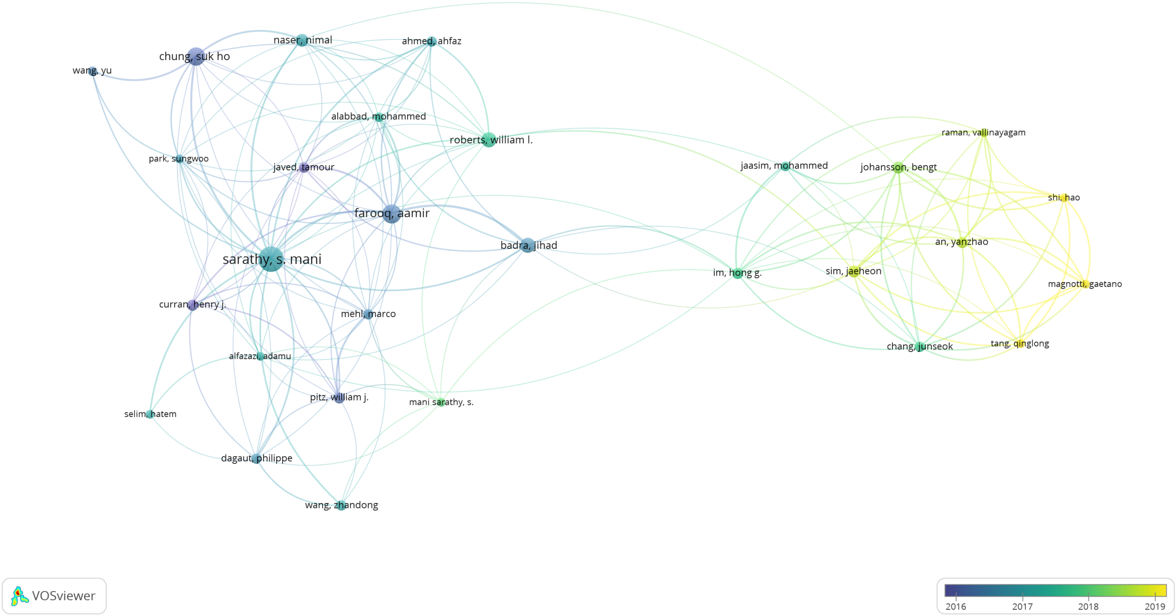


Figure A.4: Most frequent citations of 2020-2024 pool for Fossil Fuel Industry Energy-related papers. Colours indicate average publication date, and the graph is neatly divided into earlier and later research. Small number of authors here do not indicate a significant pattern.

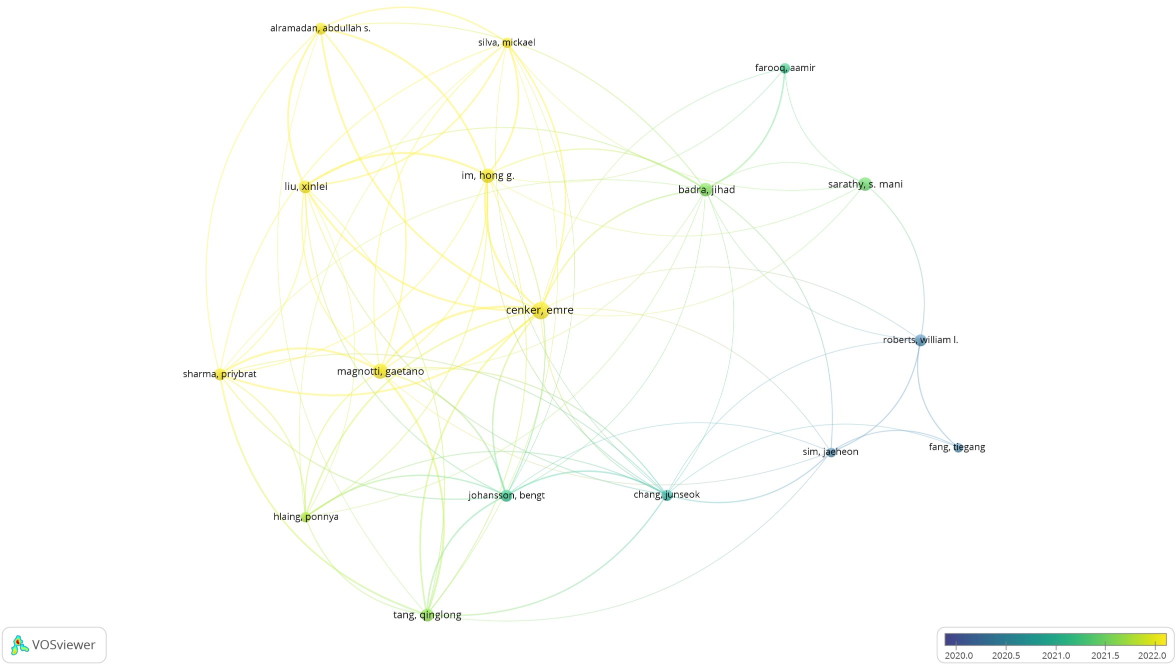


Figure A.5: Most frequent citations of 2014-2019 pool for general Energy-related papers. A large number of research authors result in no significant clusters seen and a lot of research cross-cited, resulting in no significant patterns. No authors from fossil-fuel industry research appear here.

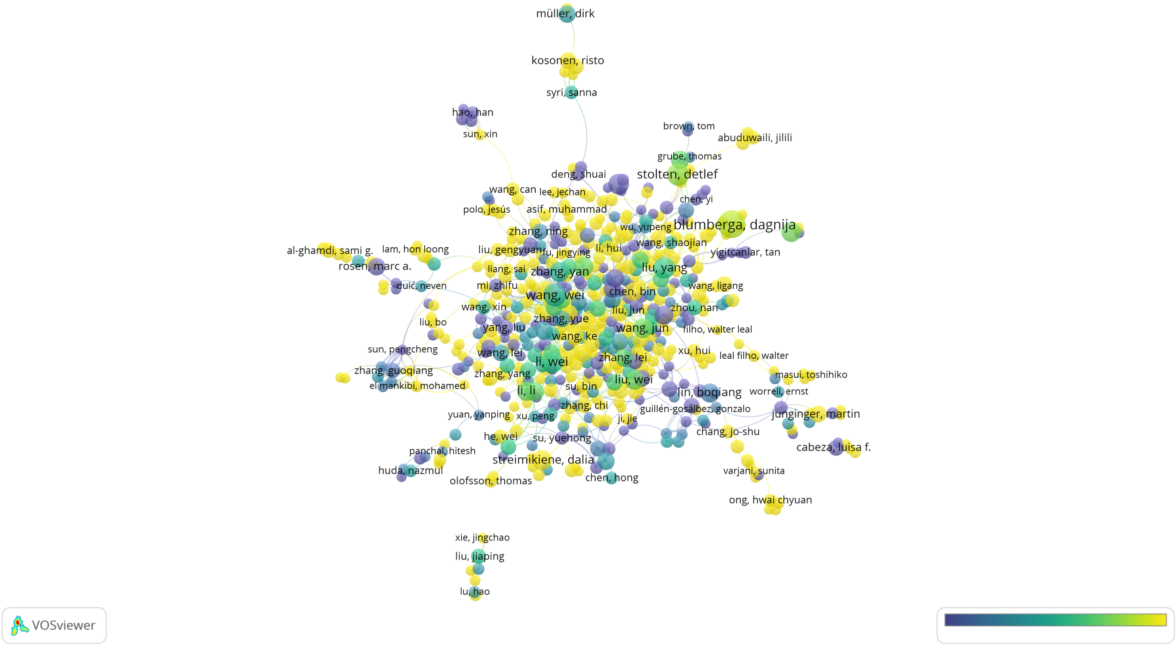


Figure A.6: Most frequent citations of 2020-2024 pool for general Energy-related papers. Large topic variability results in some clustering, yet no names related to fossil fuel industry research appear here, resulting in no patterns identified.

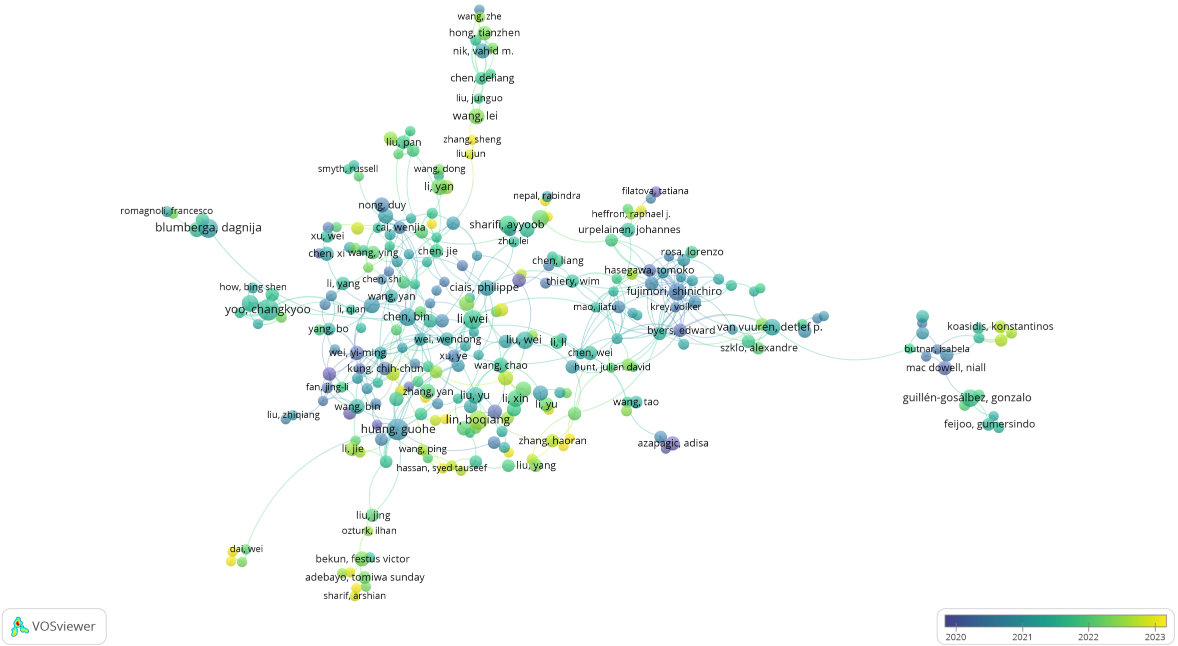


Figure A.7: Most frequent terms of 2014-2019 pool for general Energy-related papers, divided into clusters by topic. Here, red cluster of ”policy and stakeholders”, as well as green cluster of ”fuel” is of most relevance for comparison to other research.

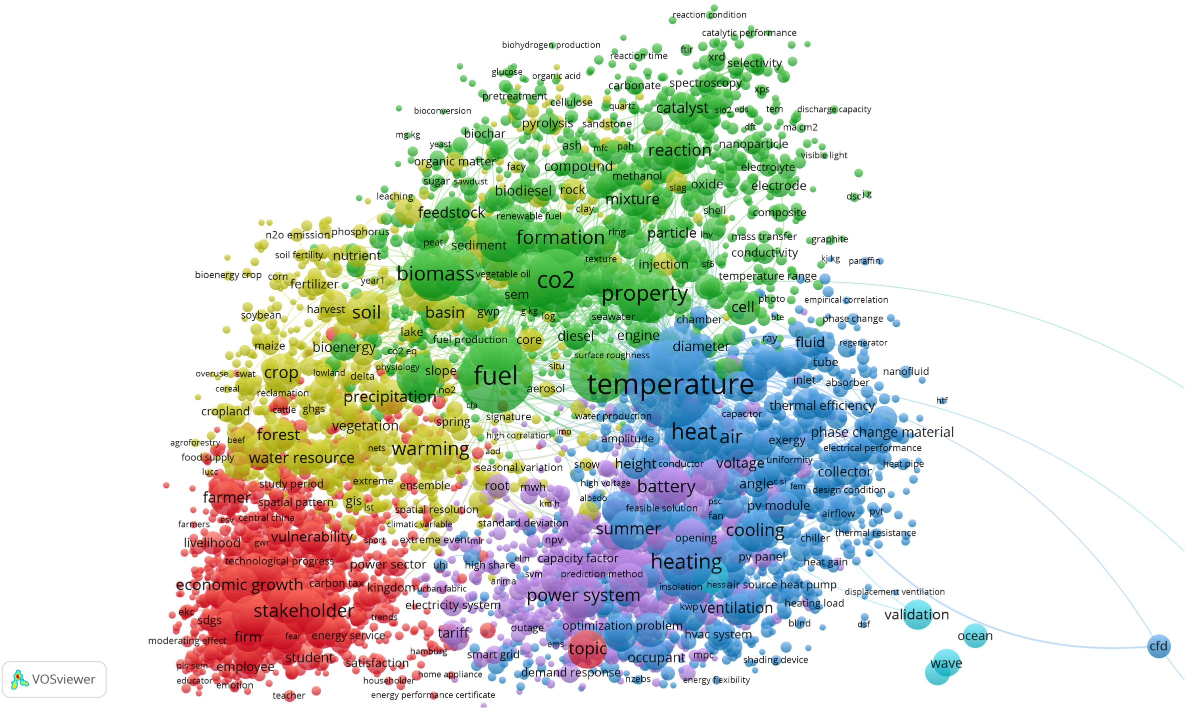


Figure A.8: Most frequent terms of 2020-2024 pool for general Energy-related papers, divided into clusters by topic. Here, green cluster of ”policy and stakeholders”, as well as blue cluster of ”fuel and consequences” is of most relevance for comparison to other research.

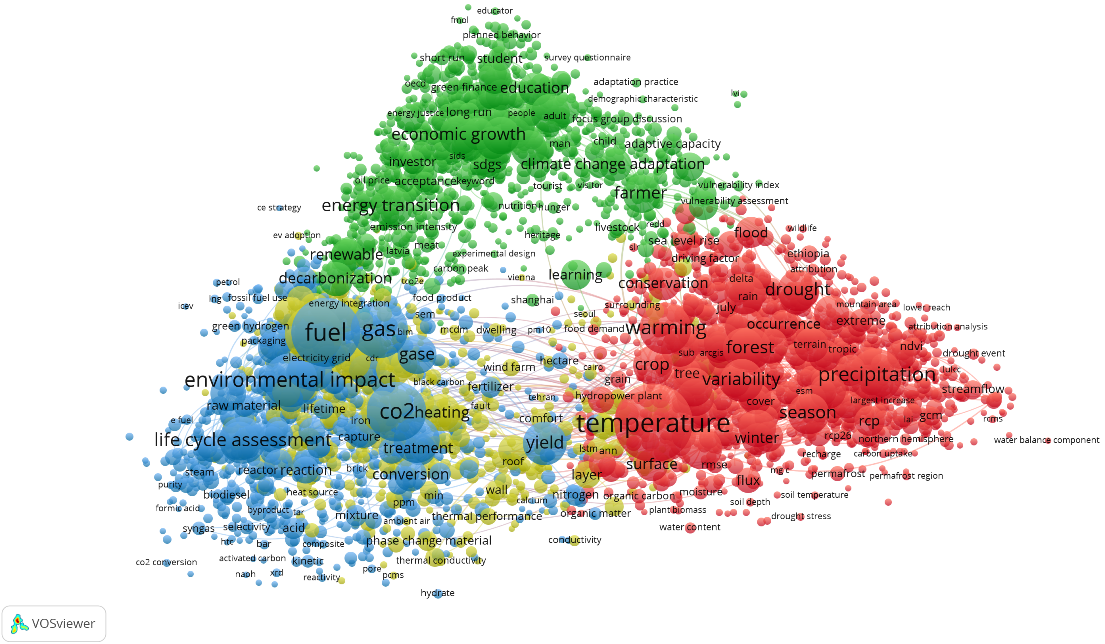


Figure A.9: Normalised most frequent common terms of Energy-related papers between general and fossil fuel industry pools of data. ”Carbon capture” and ”co2 capture” terms are once again more frequent.

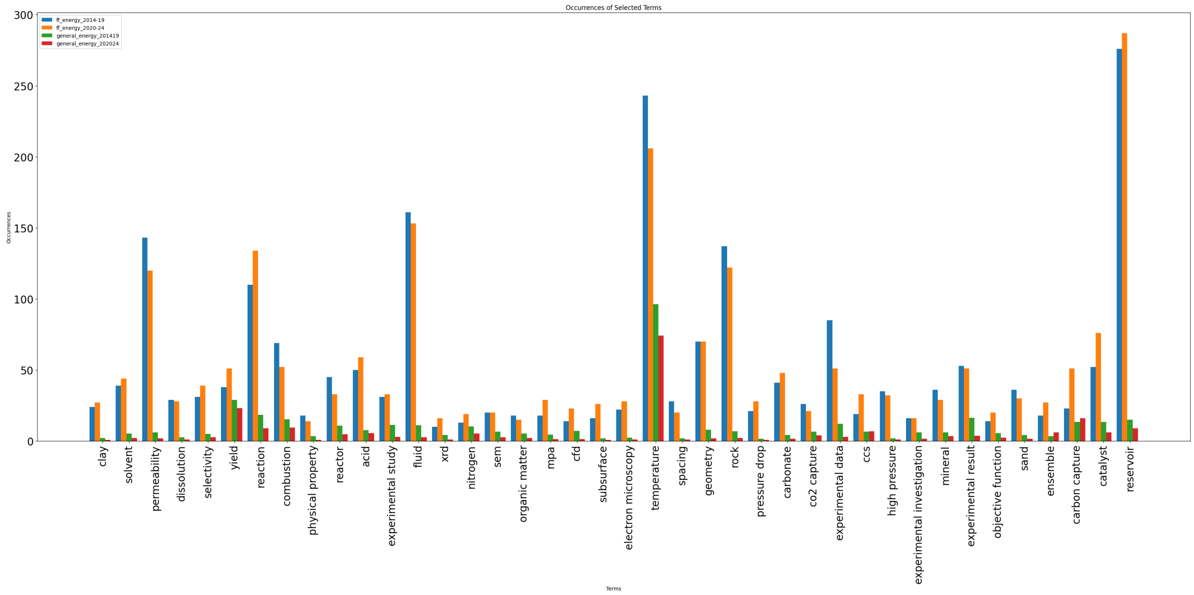


Figure A.10: Most frequent terms of 2014-2019 pool for university Energy-related papers. Presence of terms such as ”case study”, ”policy” or ”market” aligns them more to a general data pool rather than fossil fuel industry research.

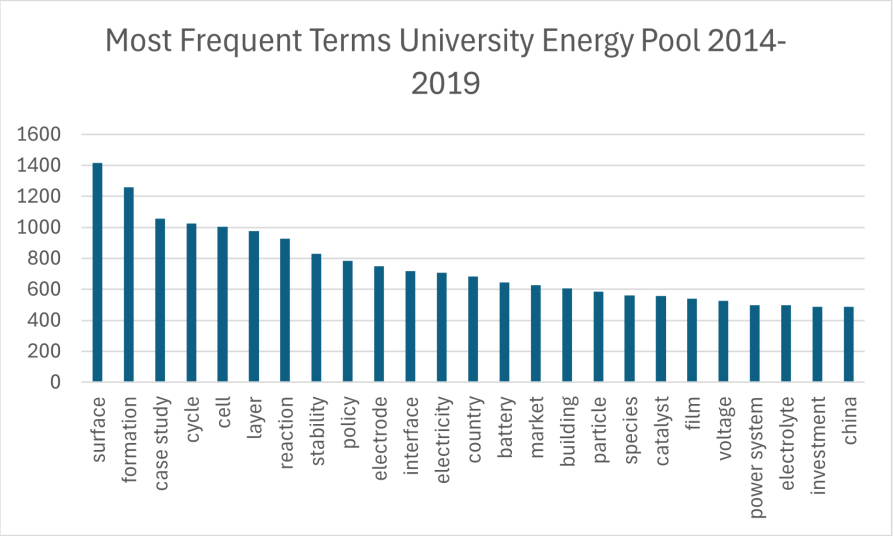


Figure A.11: Most frequent terms of 2020-2024 pool for university Energy-related papers. The results are quite similar to 2014-2019 pool.

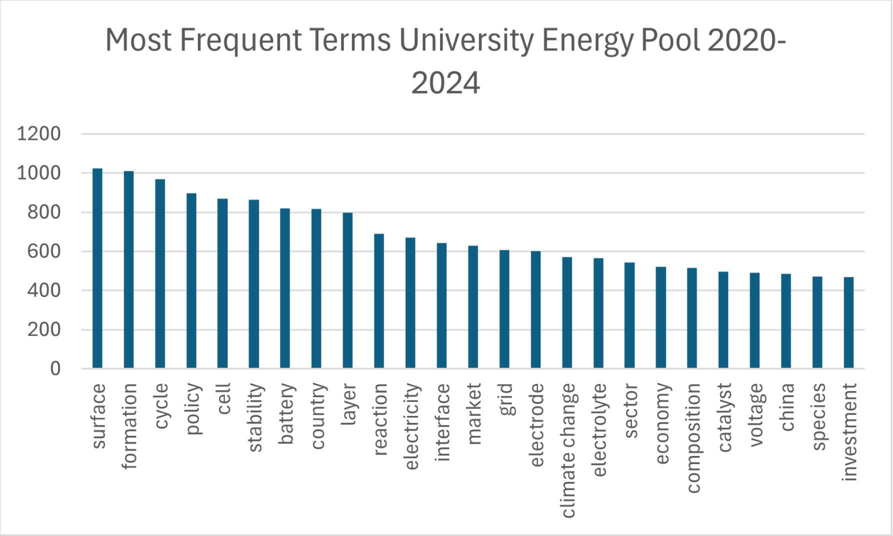


Figure A.12: Normalised 20 most frequent common terms of Energy-related papers between University and General Pools of Data. There is a large variability in where terms occur most frequently, and in subject of terms.

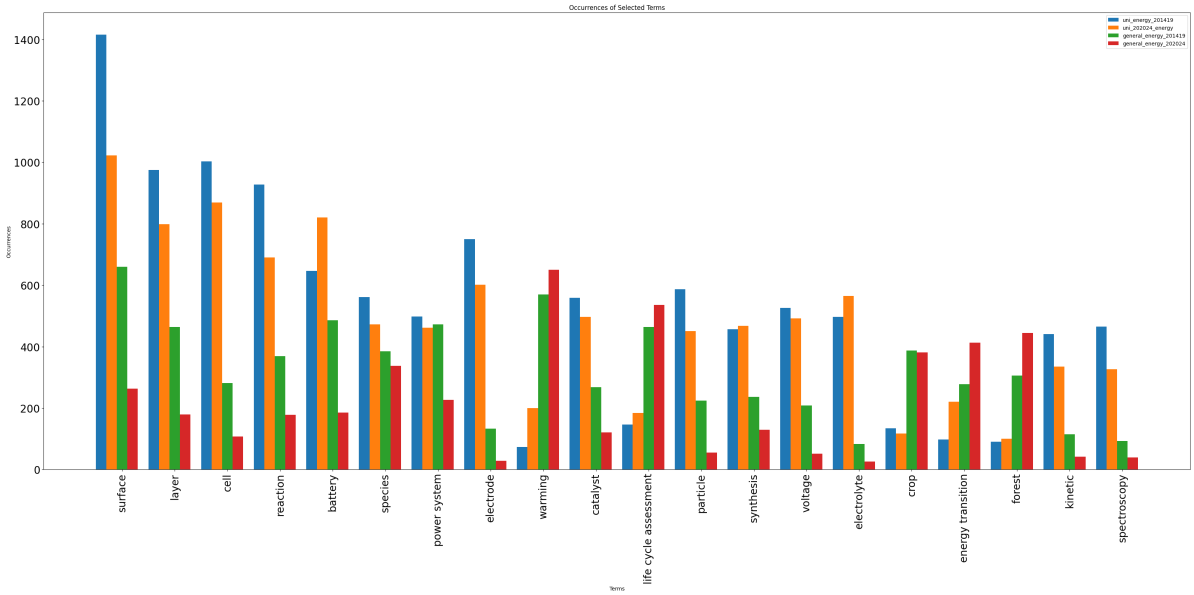


Figure A.13: Most frequent citations of 2020-2024 pool for University Energy-related papers. Large variability of clusters is explained by vastly different universities and the topics that their researchers work on.

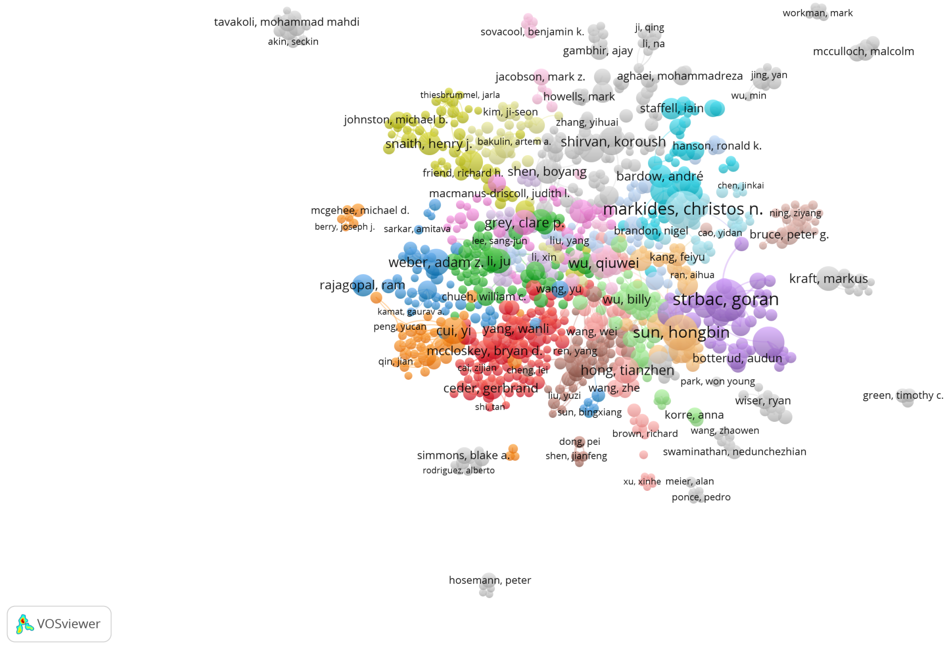


Figure A.14: Density of terms of 2014-2019 pool for general Environment-related papers. Bottom right side shows a very large prevalence of emissions and climate related terms, uncommon in fossil fuel industry funded research.

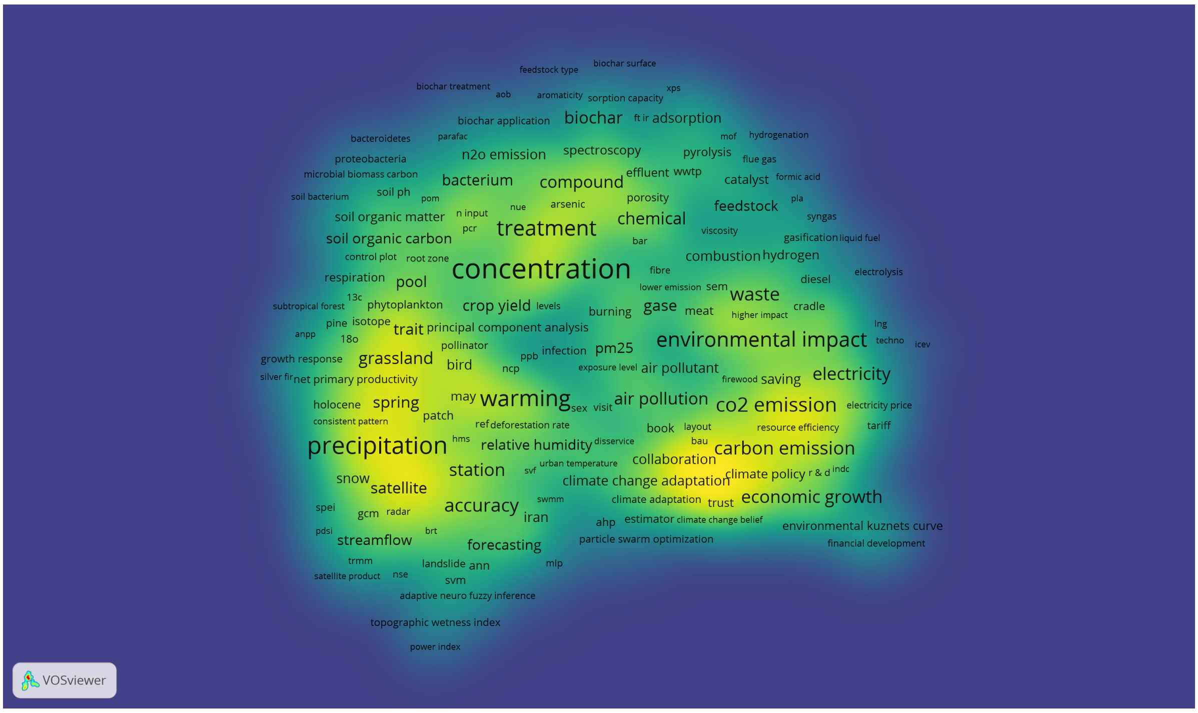


Figure A.15: Density of terms of 2020-2024 pool for general Environment-related papers. Note the right side shift from emissions to more generalised environmental quality and renewable energy interest.

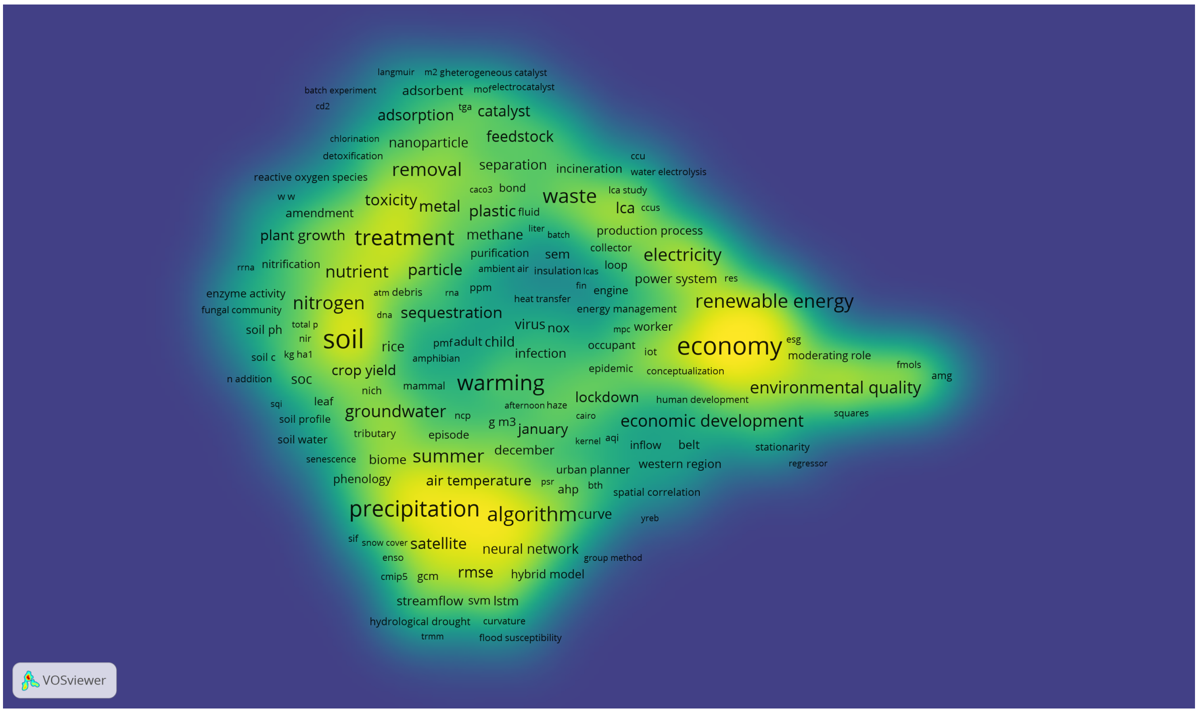


Figure A.16: Most frequent terms general of Environment-related papers in 2014-2019 pool. Most frequent terms are carbon emissions and greenhouse gas related, quite different from fossil fuel industry.

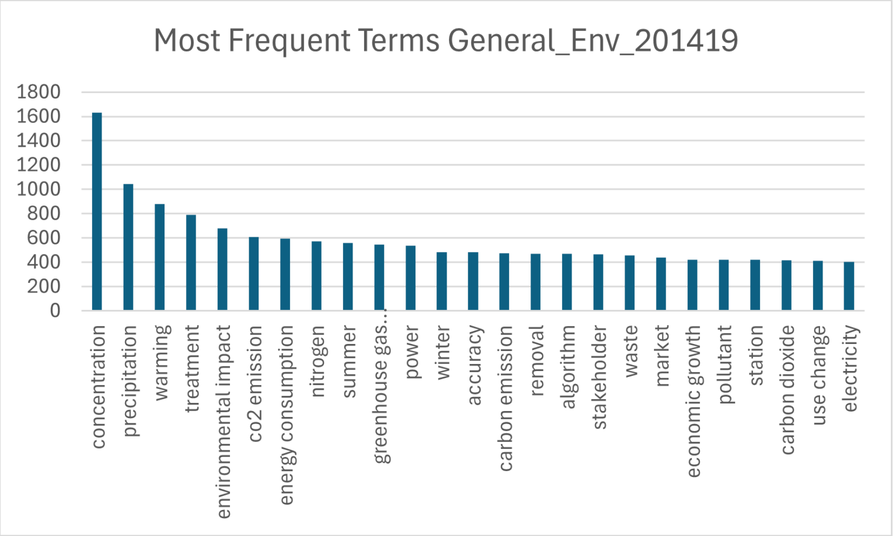


Figure A.17: Most frequent terms general of Environment-related papers in 2020-2024 pool. Results are quite similar to 2014-2019 pools of data.

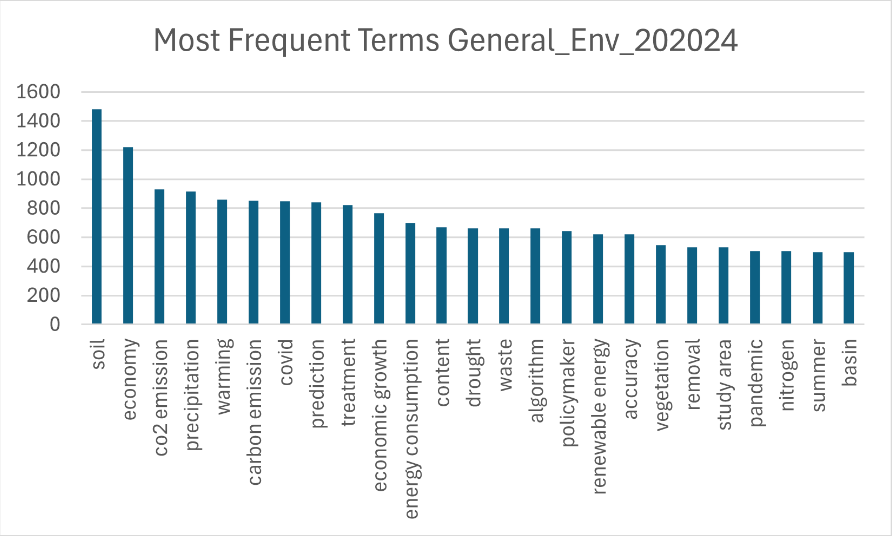


Figure A.18: Density of terms of 2014-2019 pool for university Environment-related papers. Top left side of the figure shows a very large relevance score for "policy" and related terms, in contrast to fossil fuel research.

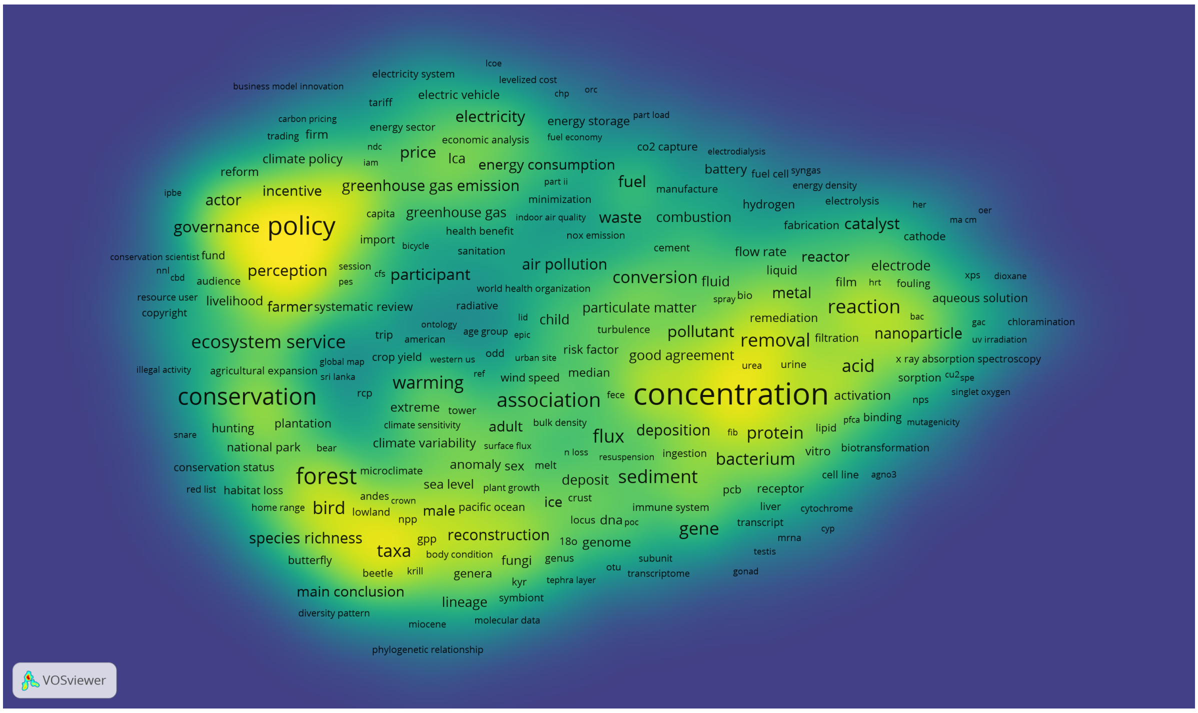


Figure A.19: Density of terms of 2020-2024 pool for university Environment-related papers. Once again, "policy" and "governance" terms have highest relevance terms, in bright yellow cluster on the left side of the figure.

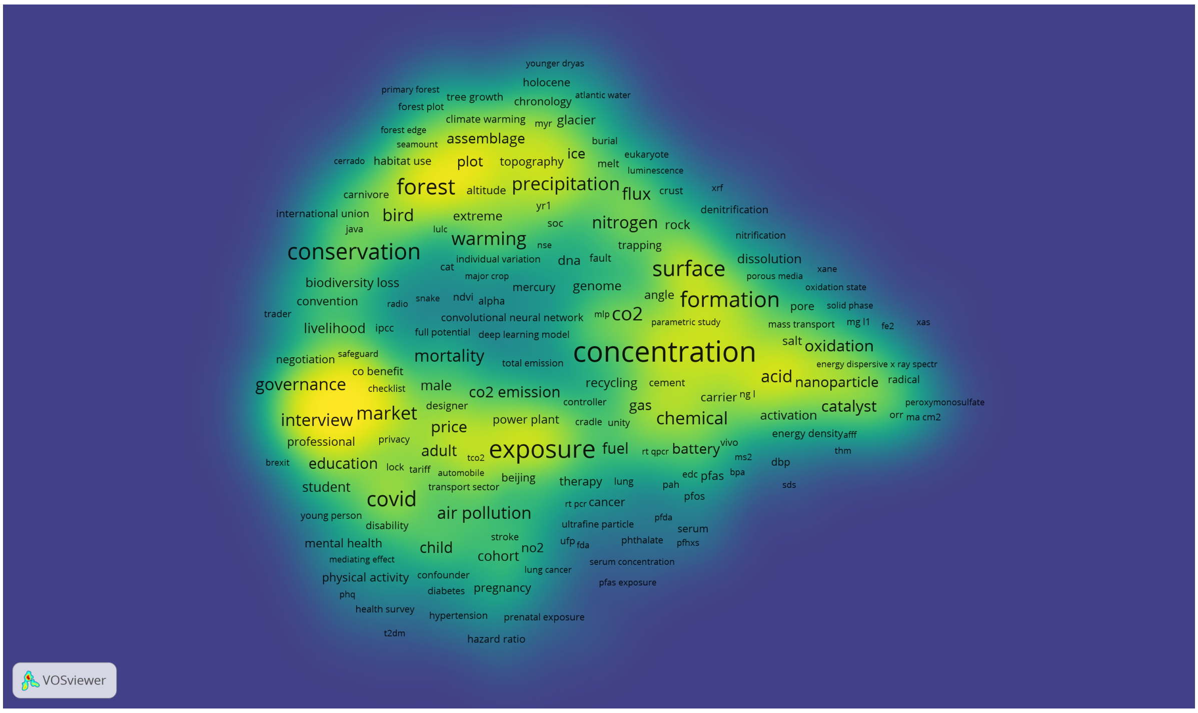


Figure A.20: Normalised most frequent common terms of Earth & Planetary papers between Fossil Fuel and General Pools of Data. Larger frequency of common terms in fossil fuel industry research indicates a smaller variability in industry focus than general research.

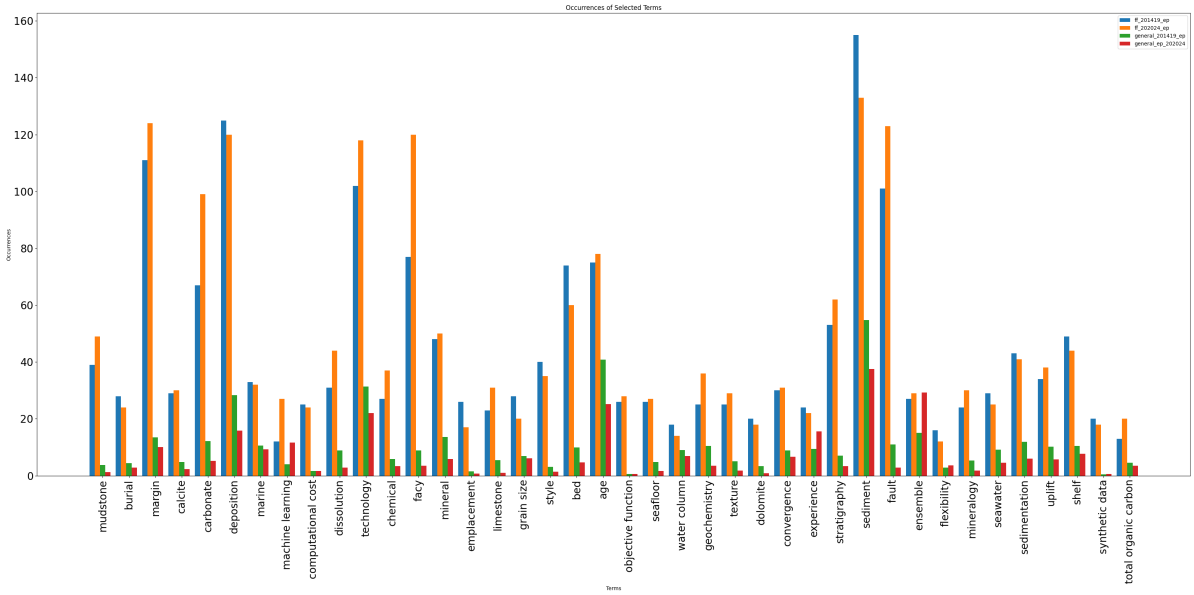


Figure A.21: Normalised most frequent common terms of Earth & Planetary papers between University and General Pools of Data. Once again, large variability of results and topics stands in contrast to fossil fuel and general research pools comparison.

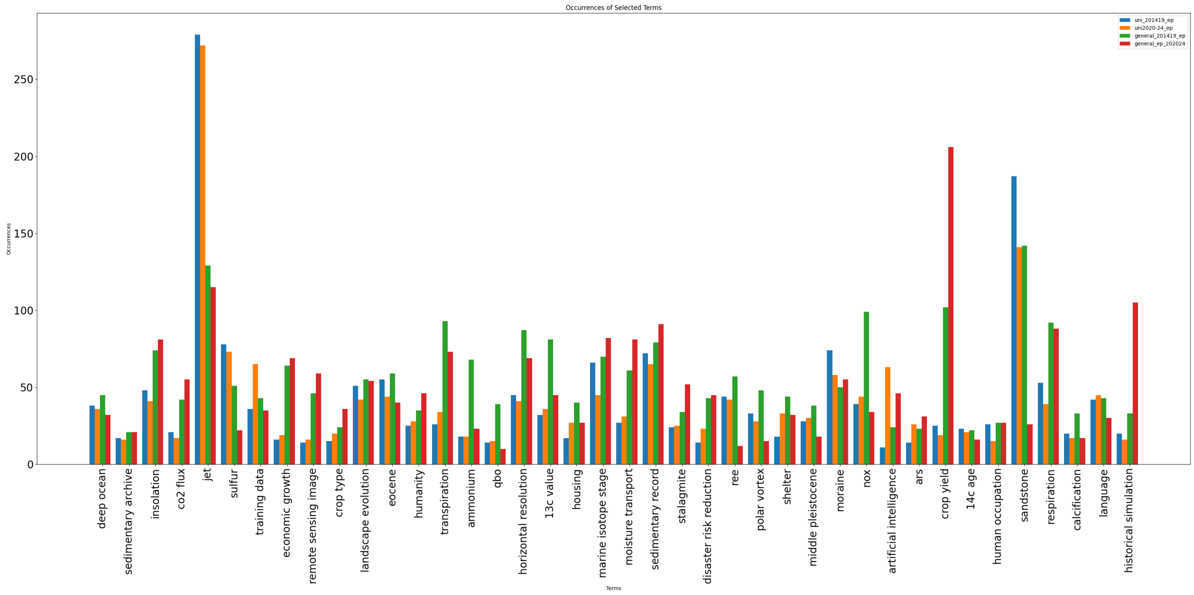


Figure A.22: Temporal Analysis of 2014-2019 general Engineering pool. Yellow bubbles indicate newly emerging terms, while dark blue terms indicate terms that appear earlier on average. Yellow bubbles indicate trends in electrical innovation, environmental impact and machine learning techniques.

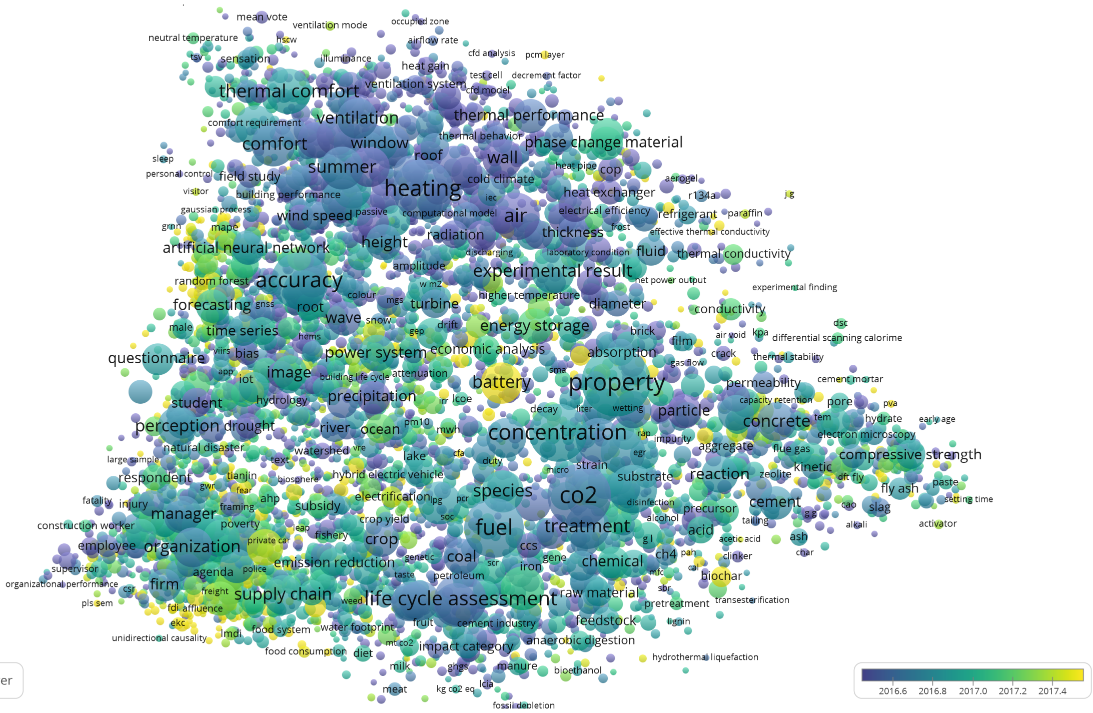


Figure A.23: Temporal Analysis of 2020-2024 general Engineering pool. Yellow bubbles indicate even more newly emerging terms, while dark blue terms indicate terms that appear earlier on average. Yellow bubbles indicate trends in electrical innovation, environmental impact and machine learning techniques. Compared to 2014-2019 pool of data, we see much larger shift towards newer terms in research on a similar longevity scale. This indicates a shift in topics discussed.

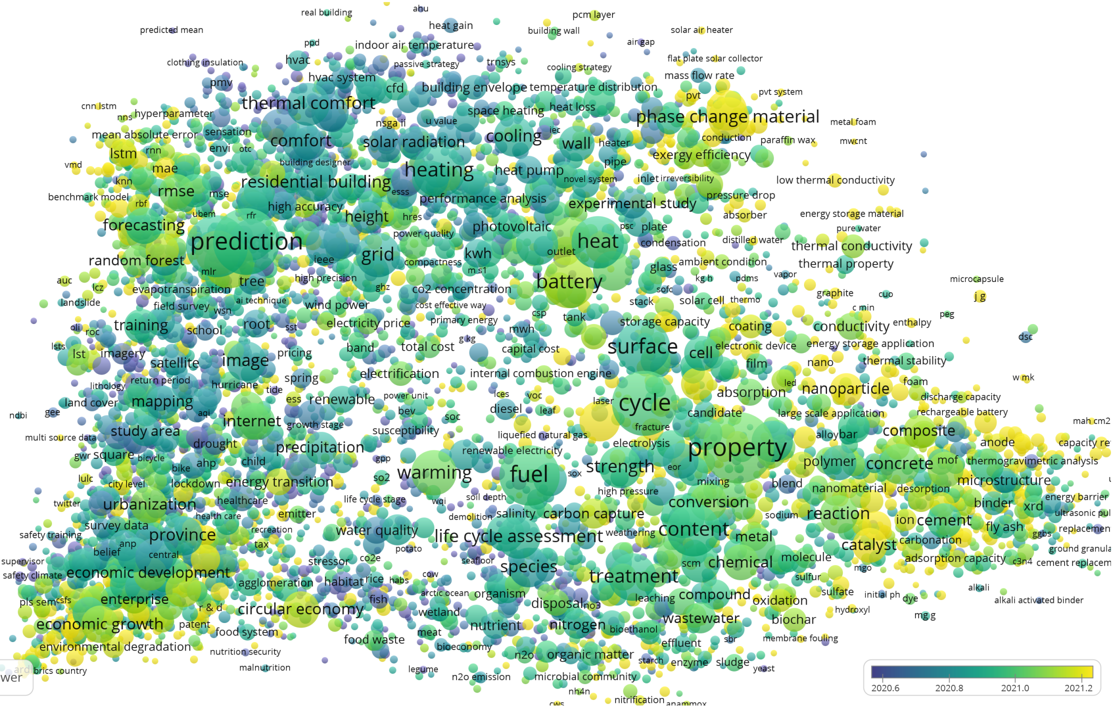
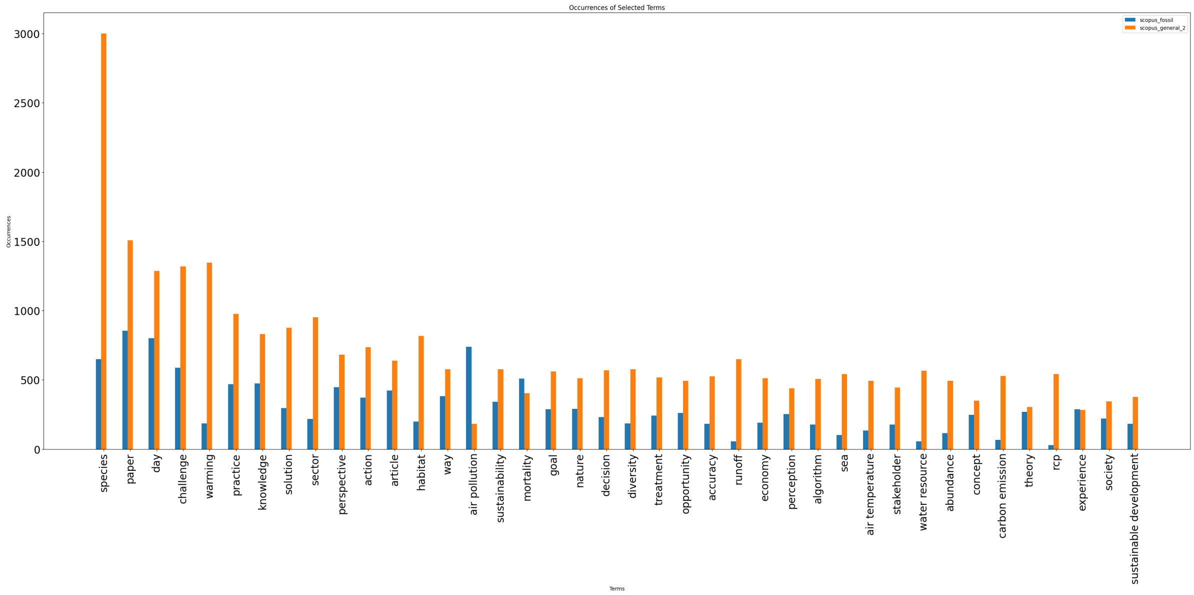


Figure A.24: Normalised 40 most frequent common terms of ”climate” term fossil fuel industry papers and General Pools of Data. Term ”climate” was broken into dozens of phrases with smaller occurrences, hence not making it to top 40. Note much larger presence of ”economy” or ”disaster risk reduction” in general pool of data. The only term in this figure that has more presence in fossil fuel research is ”pollutant”, once again indicating more technical focused research



B

Code

```
"""
Topic modelling code with hypertuning and visualisation. Executed in Jupyter Notebook and separated by cells
"""
import pandas as pd
import os

# Read the CSV file into a pandas DataFrame
papers = pd.read_csv('general_201419.csv')

# Print head
papers.head()
```

```
# Remove the columns
papers = papers['Abstract']

# sample only 100 papers
papers = papers.sample(100)

# Print out the first rows of papers
papers.head()
```

```
# Load the regular expression library
import re

# Remove punctuation
papers['paper_text_processed'] = papers.map(lambda x: re.sub('[,\.!?]', '', x))

# Convert the titles to lowercase
papers['paper_text_processed'] = papers['paper_text_processed'].map(lambda x: x.lower())

# Print out the first rows of papers
papers['paper_text_processed'].head()
```

```
import gensim
from gensim.utils import simple_preprocess

def sent_to_words(sentences):
    for sentence in sentences:
        yield(gensim.utils.simple_preprocess(str(sentence), deacc=True))  # deacc=True removes punctuations

data = papers.paper_text_processed.values.tolist()
data_words = list(sent_to_words(data))

print(data_words[:1][0][:30])
```

```
# Build the bigram and trigram models
```

```
bigram = gensim.models.Phrases(data_words, min_count=5, threshold=100) # higher threshold
    fewer phrases.
trigram = gensim.models.Phrases(bigram[data_words], threshold=100)

# Faster way to get a sentence clubbed as a trigram/bigram
bigram_mod = gensim.models.phrases.Phraser(bigram)
trigram_mod = gensim.models.phrases.Phraser(trigram)
```

```
# NLTK Stop words
import nltk
nltk.download('stopwords')
from nltk.corpus import stopwords

stop_words = stopwords.words('english')
stop_words.extend(['from', 'subject', 're', 'edu', 'use'])

# Define functions for stopwords, bigrams, trigrams and lemmatization
def remove_stopwords(texts):
    return [[word for word in simple_preprocess(str(doc)) if word not in stop_words] for doc
        in texts]

def make_bigrams(texts):
    return [bigram_mod[doc] for doc in texts]

def make_trigrams(texts):
    return [trigram_mod[bigram_mod[doc]] for doc in texts]

def lemmatization(texts, allowed_postags=['NOUN', 'ADJ', 'VERB', 'ADV']):
    """https://spacy.io/api/annotation"""
    texts_out = []
    for sent in texts:
        doc = nlp(" ".join(sent))
        texts_out.append([token.lemma_ for token in doc if token.pos_ in allowed_postags])
    return texts_out
```

```
!python -m spacy download en_core_web_sm
import spacy

# Remove Stop Words
data_words_nostops = remove_stopwords(data_words)

# Form Bigrams
data_words_bigrams = make_bigrams(data_words_nostops)

# Initialize spacy 'en' model, keeping only tagger component (for efficiency)
nlp = spacy.load("en_core_web_sm", disable=['parser', 'ner'])

# Do lemmatization keeping only noun, adj, vb, adv
data_lemmatized = lemmatization(data_words_bigrams, allowed_postags=['NOUN', 'ADJ', 'VERB', '
    ADV'])

print(data_lemmatized[:1][0][:30])
```

```
import gensim.corpora as corpora

# Create Dictionary
id2word = corpora.Dictionary(data_lemmatized)

# Create Corpus
texts = data_lemmatized

# Term Document Frequency
corpus = [id2word.doc2bow(text) for text in texts]

# View
print(corpus[:1][0][:30])
```

```
# Build LDA model
lda_model = gensim.models.LdaMulticore(corpus=corpus,
                                       id2word=id2word,
                                       num_topics=10,
                                       random_state=100,
                                       chunksize=100,
                                       passes=10,
                                       per_word_topics=True)
```

```
from pprint import pprint

# Print the Keyword in the 10 topics
pprint(lda_model.print_topics())
doc_lda = lda_model[corpus]
```

```
from gensim.models import CoherenceModel

# Compute Coherence Score
coherence_model_lda = CoherenceModel(model=lda_model, texts=data_lemmatized, dictionary=id2word, coherence='c_v')
coherence_lda = coherence_model_lda.get_coherence()
print('Coherence Score: ', coherence_lda)
```

```
# supporting function
def compute_coherence_values(corpus, dictionary, k, a, b):

    lda_model = gensim.models.LdaMulticore(corpus=corpus,
                                           id2word=dictionary,
                                           num_topics=k,
                                           random_state=100,
                                           chunksize=100,
                                           passes=10,
                                           alpha=a,
                                           eta=b)

    coherence_model_lda = CoherenceModel(model=lda_model, texts=data_lemmatized, dictionary=id2word, coherence='c_v')

    return coherence_model_lda.get_coherence()
```

```
import numpy as np
import tqdm

grid = {}
grid['Validation_Set'] = {}

# Topics range
min_topics = 2
max_topics = 9
step_size = 1
topics_range = range(min_topics, max_topics, step_size)

# Alpha parameter
alpha = list(np.arange(0.01, 0.5, 0.3))
alpha.append('symmetric')
alpha.append('asymmetric')

# Beta parameter
beta = list(np.arange(0.01, 0.5, 0.3))
beta.append('symmetric')

# Validation sets
num_of_docs = len(corpus)
corpus_sets = [gensim.utils.ClippedCorpus(corpus, int(num_of_docs*0.75)),
               corpus]

corpus_title = ['75% Corpus', '100% Corpus']

model_results = {'Validation_Set': [],
                 'Topics': [],
                 'Alpha': [],
                 'Beta': [],
                 'Coherence': []
```

```
                }

# Can take a long time to run
if 1 == 1:
    pbar = tqdm.tqdm(total=(len(beta)*len(alpha)*len(topics_range)*len(corpus_title)))

    # iterate through validation corpuses
    for i in range(len(corpus_sets)):
        # iterate through number of topics
        for k in topics_range:
            # iterate through alpha values
            for a in alpha:
                # iterare through beta values
                for b in beta:
                    # get the coherence score for the given parameters
                    cv = compute_coherence_values(corpus=corpus_sets[i], dictionary=id2word,
                                                  k=k, a=a, b=b)
                    # Save the model results
                    model_results['Validation_Set'].append(corpus_title[i])
                    model_results['Topics'].append(k)
                    model_results['Alpha'].append(a)
                    model_results['Beta'].append(b)
                    model_results['Coherence'].append(cv)

                    pbar.update(1)
    pd.DataFrame(model_results).to_csv('./lda_tuning_results.csv', index=False)
    pbar.close()
```

```
    import matplotlib.pyplot as plt
# Load the CSV file into a DataFrame
df = pd.read_csv('lda_tuning_results.csv')


# Print the data types of the columns before conversion
print("Data types before conversion:")
print(df.dtypes)

# Convert 'Alpha', 'Beta', 'Topics', and 'Coherence' columns to numeric, forcing any non-numeric values to NaN
df['Alpha'] = pd.to_numeric(df['Alpha'], errors='coerce')
df['Beta'] = pd.to_numeric(df['Beta'], errors='coerce')
df['Topics'] = pd.to_numeric(df['Topics'], errors='coerce')
df['Coherence'] = pd.to_numeric(df['Coherence'], errors='coerce')

# Print the data types of the columns after conversion
print("Data types after conversion:")
print(df.dtypes)

# Drop rows with NaN values in 'Alpha', 'Beta', 'Topics', or 'Coherence' columns
df.dropna(subset=['Alpha', 'Beta', 'Topics', 'Coherence'], inplace=True)

# Set the values of Alpha and Beta for filtering
alpha_value = 0.01
beta_value = 0.11

# Filter the DataFrame for the specified Alpha and Beta values
filtered_df = df[(df['Alpha'] == alpha_value) & (df['Beta'] == beta_value)]

# Ensure the filtered DataFrame is not empty and contains no NaN/Inf values in 'Topics' and 'Coherence'
if filtered_df.empty or filtered_df['Topics'].isnull().any() or filtered_df['Coherence'].isnull().any():
    print("No valid data available for the specified Alpha and Beta values.")
else:
    # Plot Topics vs Coherence
    plt.figure(figsize=(10, 6))
    plt.plot(filtered_df['Topics'], filtered_df['Coherence'], marker='o')

    # Set axis limits with a check to avoid NaN/Inf values
    plt.xlim(filtered_df['Topics'].min() - 1, filtered_df['Topics'].max() + 1)
    plt.ylim(filtered_df['Coherence'].min() - 0.1, filtered_df['Coherence'].max() + 0.1)

    plt.xlabel('Topics')
    plt.ylabel('Coherence')
    plt.title(f'Topics vs Coherence (Alpha={alpha_value}, Beta={beta_value})')
    plt.grid(True)
    plt.show()
```

```
#Run again after hypertuning
num_topics = 5

lda_model = gensim.models.LdaMulticore(corpus=corpus,
                                       id2word=id2word,
                                       num_topics=num_topics,
                                       random_state=100,
                                       chunksize=100,
                                       passes=10,
                                       alpha=0.01,
                                       eta=0.9)
```

```
import pyLDAvis.gensim_models as gensimvis
import pickle
import pyLDAvis

# Visualize the topics
pyLDAvis.enable_notebook()

LDAvis_data_filepath = os.path.join('./ldavis_tuned_'+str(num_topics))

# # this is a bit time consuming - make the if statement True
# # if you want to execute visualization prep yourself
if 1 == 1:
    LDAvis_prepared = gensimvis.prepare(lda_model, corpus, id2word)
    with open(LDAvis_data_filepath, 'wb') as f:
        pickle.dump(LDAvis_prepared, f)

# load the pre-prepared pyLDAvis data from disk
with open(LDAvis_data_filepath, 'rb') as f:
    LDAvis_prepared = pickle.load(f)

pyLDAvis.save_html(LDAvis_prepared, './ldavis_tuned_'+ str(num_topics) +'.html')

LDAvis_prepared
```