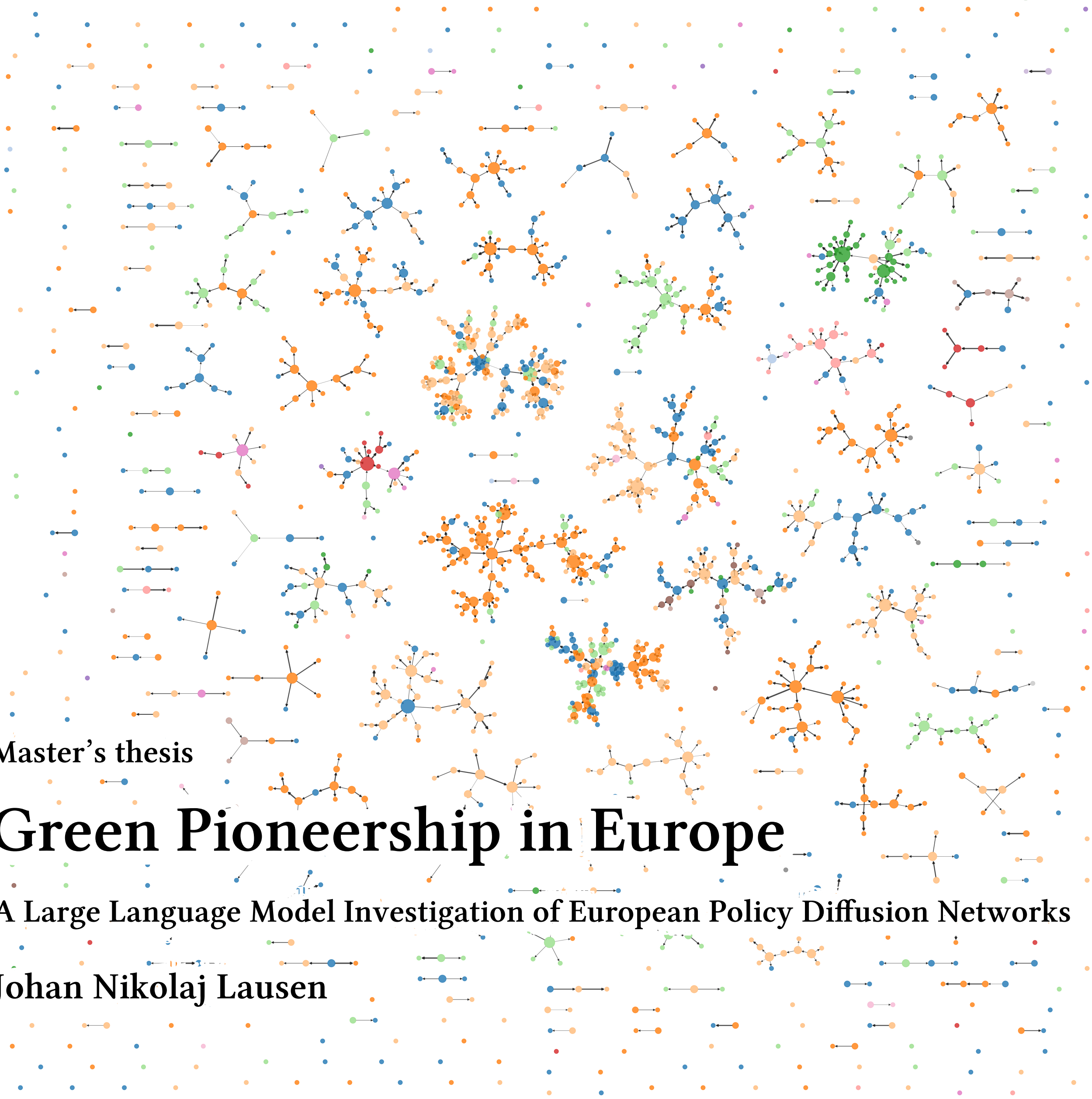
UNIVERSITY OF COPENHAGEN
FACULTY OF SCIENCE



Master's thesis

Green Pioneership in Europe

A Large Language Model Investigation of European Policy Diffusion Networks

Johan Nikolaj Lausen

Advisors: Teis Hansen

June 2nd, 2025

This thesis has been submitted to the Department of Food and Resource Economics (IFRO) of The Faculty of Science, University of Copenhagen

UNIVERSITY OF COPENHAGEN
FACULTY OF SCIENCE

Institute:	Department of Food and Resource Economics
Author:	Johan Nikolaj Lausen
Email:	hgw961@alumni.ku.dk // johan.lausen@gmail.com

Title and subtitle:	Green Pioneership in Europe - A Large Language Model Investigation of European Policy Diffusion Networks
Advisors:	Teis Hansen
Handed in:	June 2nd, 2025
Defended:	June 13th, 2025

Name ____________________

Signature ________________

Date ____________________

Acronyms

ASEAN	Association of Southeast Asian Nations
DK	Denmark
DAG	Directed Acyclic Graph
EU	European Union
EHA	Event History Analysis
FAO	Food and Agriculture Organisation of the United Nations
FAOLEX	FAO law, regulation and policy database
IE	Ireland
JSD	Jensen-Shannon Divergence
LLM	Large Language Model
NAP	National Adaptation Plans
NAS	National Adaptation Strategy
RE	Renewable Energy
T	Temperature (LLM parameter)
UK	United Kingdom

Abstract

The principle of Green Pioneership ('Grønt Foregangsland') has played an important role in Danish public policy and debate in recent years. Motivated by the frequent, but often unsubstantiated claims of (non-)pioneership seen in Danish public debate, this thesis investigates measurable flows of influence between environmental policies in Europe. Specifically, we develop, test and demonstrate the use of a new Large Language Model-based methodology for finding patterns of policy diffusion among legal texts and policy documents. This text-as-data approach uses generative language models to extract information on the regulatory type, objectives, strategy, instruments, targets and implementation mechanisms of policies, before comparing these for similarity using embedding models. Pairwise document similarities are then used to construct a network of policies, which is probed for patterns in characteristics and contents of those documents, who have seemingly influenced others. We test this methodology on four datasets, ranging in size from 10 environmental policy documents to over 1700, and compare to previous policy diffusion studies on those datasets, with mixed results. The methodology shows some success finding large scale patterns, but has difficulties probing the finer contents of policies, nor is it seemingly entirely successful in treating policies in a manner agnostic to the origin, language and format of the documents describing them. Results demonstrating the data processing procedure do not indicate Denmark to be in a distinctive or unique position when it comes to diffusing climate change policies in general, but should be considered tentative at best given the challenges. The methodology shows promise however, as many of these issues seem likely to be improved upon, if not overcome, by improving and fine tuning language model instructions and using more advanced models for the treatment. In particular, the methodology seems especially viable as a tool for case-selection and hypothesis generation in mixed-methods studies, and as a significant development over limited-scope text-as-data approaches available until now.

Acknowledgements

I would like to express deep appreciation to my advisor Teis for his valuable guidance, as well as the rest of the GreenTrac project team for helpful sparring, idea generation and assistance along the way.

Contents

1 Introduction **1**

2 Background **3**
2.1 Leaders, Frontrunners and Pioneers 3
2.2 Policy Diffusion and related fields of research 4
2.3 Natural Language Processing 7
2.3.1 Embeddings 7
2.3.2 Large Language Models (LLMs) 9
2.4 Network Analysis 12

3 Methods **14**
3.1 Datasets and data sources 14
3.1.1 Document Pretreatment 16
3.2 Language Modelling 17
3.2.1 Prompt design and prompting strategy 18
3.2.2 Embedding 19
3.3 Network Analysis 20
3.3.1 Keyword Distributions 21

4 Results **23**
4.1 Directives Test-set 23
4.1.1 Language model outputs 23
4.1.2 Similarity analysis 29
4.1.3 Network Analysis 32
4.2 Diffusion Study Replication 37
4.2.1 Adaptation Strategies 37
4.2.2 Green Taxonomies 42
4.3 FAOLEX 46

5 Discussion **53**

6 Conclusion **63**

References **I**

AI Declaration **XIII**

Appendices **XIV**
A Extended Background XIV
A.1 Policy Diffusion and related fields of research XIV
A.2 Natural Language Processing XVIII
A.3 Embeddings XIX
A.4 Large Language Models (LLMs) XXIV
A.5 Network Analysis XXIX
B Prompt variations XXXIII
B.1 Main prompt XXXIII
B.2 Previous prompt examples XXXIX
C Large Language Model variation XLIII
D Instrument Similarities XLVIII
E Network rules XLIX
F Datasets LIV

1 Introduction

The concepts of leaders, frontrunners and pioneers and have played an important role in in Danish public debate in recent years. Green Pioneership ('Grønt Foregangsland') is an important aspect of both the self- and foreign image of Danish public environmental management, and is increasingly invoked as political justification for environmental action in Danish policy. This is especially true regarding governance of climate change, where green pioneership is included as a guiding principle in the Danish climate law, which states that Denmark must realise this principle by inspiring and influencing climate action in the rest of the world.

Inspiration and influence can both take many forms and have many outcomes. As much is clear from the Danish public debate on green pioneership, where conflicting interpretations of the concept frequently lead to disagreement about the truthfulness of Denmark's self-asserted pioneer status.

One possible interpretation of green pioneership is the tendency for other countries to adopt similar legislation and policies as the pioneer country. The 'diffusion' of (green) policies across geographical and political borders has a long history of study in the field of Policy Diffusion and related branches of scholarship, which has resulted in a rich literature on the topic both theoretical and empirical. Fundamentally however, policy diffusion is difficult to document, as it requires detailed knowledge of policy documents originating in many different geographical areas, hence in many different languages, and from multiple legislative systems.

Researchers on the topic typically resort to one of two toolboxes when attempting to document diffusion. Qualitative researchers conduct limited-scope case studies involving labour intensive manual identification and investigation of policies from different countries, documenting how policies travel between regions and searching for 'smoking-gun' evidence of the pathways through which inspiration or influence occurred. Alternatively, some researchers conduct larger scale quantitative studies which uncover correlations between country action on a certain topic or policy area, and explanatory factors both internal and external to the country. Such quantitative studies can engage with many more policies than case-studies allow, but are however limited in their causal explainability and in uncovering pathways of influence, as they rarely engage with the substantial content, similarities or subtle differences between specific policy documents.

Researchers have attempted to combine the advantages of the two research disciplines through computerised processing of text and network analysis, but have so far faced challenges due especially to the complex nature of language and shortcomings of primitive techniques. Notably, computerised cross-language comparisons of policy content has so far remained unexplored and thereby limited applicability of a text-as-data approach in e.g. European policy diffusion contexts. Thereby, the question of which European countries exert the highest environmental policy diffusion pressures remain severely understudied, and as a result, public debate on the topic in countries such as Denmark revolves around interpretation and gut feeling rather than empirics.

Recently however, the field of Natural Language Processing has achieved improvements in text processing at a dramatic pace, and the recent rise of Large Language Models provides a revolutionary tool for quickly and efficiently parsing large amounts of text, including legislative texts and policy documents, regardless of language and region of origin. This offers an unprecedented opportunity to search quantitively for patterns in the substantive contents of such documents.

Specifically, two emerging tools show great promise for use in policy studies. Generative Language Models have progressed to a point enabling processing of several hundred pages of text at once, and show promise in summarisation and information extraction tasks suitable for extracting the substantial content of policy documents. Furthermore, Embedding Models allow quick context-informed comparisons of longer textual elements. Combined, the two allow pairwise comparisons of the substantial content of policy documents.

This thesis aims to develop, test and explore these new tools and their usefulness for studying Green Pioneership in public management by European states – specifically the dimension of green pioneership involving cross-border policy inspiration. Pairwise relationships will be explored through Network Analysis, which provides a framework and toolbox for investigating the often complex interrelationships between constituents of policy diffusion networks. **Specifically, the thesis will attempt to answer the following research questions:**

1. Can Large Language Models effectively extract information from and process the content and substance of policy documents?

2. Can extracted policy information be used to uncover similarities, correlations and causal links between policies that have diffused across borders?

3. Can a Large Language Model-based investigation replicate the results of previous environmental policy diffusion studies?

4. Given a large-scale dataset of European environmental policy, can a Large Language Model-based investigation help to answer questions such as:

    - Which countries are most successful at influencing and inspiring policy in other European nations?
    - Can Denmark in this context be regarded as a 'Green Pioneer'?
    - Can any general patterns be established regarding what characterises environmental policies that successfully diffuse across national borders?

2 Background

The following subsections provide brief and somewhat simplified overviews of the academic topics relevant to this thesis, excerpts from longer, more technical introductions found in Appendix A. We begin by introducing in more detail the concept of 'Green Pioneership' and weave this into a broader introduction to the scholarship on Policy Diffusion. A brief review of policy diffusion studies follows, with emphasis on the methodological challenges encountered in this field. This leads to a description of how recent advances in computer-based Natural Language Processing may provide a new approach to combining the advantages of distinct methodological approaches while avoiding common challenges. In particular, we describe how Text Embeddings can provide an information-dense and comparison-ready representation of text. After introducing Text Embeddings, we describe how Large Language Models can be used to pre-process documents and prepare them for embedding. The final subsection introduces Network Analytics as a tool for investigating flows of influence and inspiration, including notable measures of importance in networks relevant for policy diffusion.

2.1 Leaders, Frontrunners and Pioneers

The concept of a *Green Pioneer* country ('Grønt Foregangsland', also translated as *Green Leadership* or *Green Frontrunner*) has gained prominence in Danish politics and public debate in recent years. Its importance has been elevated in part as a result of its inclusion as a guiding principle in the Danish Climate Act of 2020 (Klima-, Energi- og Forsyningsministeriet, 2021), albeit usage of the term stretches back several decades (Ritzau, 2009; Rothernborg, 2001). The Climate Act specifies that:

> **§1.** *Stk. 4.*: "1) Denmark must be a front-runner in international climate efforts, inspiring and influencing the rest of the world. Denmark has both a historical and moral responsibility to lead the way."

While the framing of the pioneer role in the Danish Act is strongly normative, there are many reasons beyond the strictly moral for why countries may wish to act as green pioneers (Jänicke, 2005). These include reputational and 'soft power' gains, that may lead to later advantages, both in diplomatic settings or by attracting investment and talent. Likewise, pioneers potentially enjoy first-mover advantages by fostering innovation and gaining competitive edges in global green markets - benefits that may spill over and improve the broader national economy.

The diverse motivations for pioneership contribute to varied interpretations and opinions on *how* a country should embody the role of a pioneer. This diversity is evident Danish public discourse, where the concept has been used in many different, and sometimes conflicting, ways.[1] Some commonly expressed understandings of 'Green Pioneer' countries include:

- Countries with either ambitious (national) environmental goals, strict protection standards and laws, or that have achieved 'success' in the form successful implementation of environmental policies or of beneficial environmental outcomes.
- Countries that have either pledged to, adopted or implemented laws to, or achieved significant (national) reductions in environmentally harmful activities.
- Countries that: have 'green' research and development or strong 'green' industries; invest towards establishing such sectors; have otherwise achieved 'Green Growth'.
- Countries that push for ambitious *global* environmental action, e.g. by inspiring, teaching, coercing or aiding other countries towards environmental action.

Furthermore, each understanding involves choices regarding e.g. indicators or acceptable levels of action to classify as 'ambitious'. Thereby, any given country likely can be interpreted to embody at least one of these roles - why some argue that pioneership involves checking off multiple or even all of these boxes. This has lead to a muddied picture regarding the concept of pioneership, where varying interpretations has led to debate regarding e.g. Denmark's pioneer status, and the evidence supporting it.[2]

This thesis intends to contribute to that debate, by studying a very specific aspect of 'Green Pioneership'. Specifically, we study the dimensions of pioneership which concerns adopting policy that through one way or the other results in similar content being adopted in policies elsewhere. This relates strongly to the academic field of *Policy Diffusion*, that studies the spread of policies between regions, and which we therefore dive further into in the following section.

2.2 Policy Diffusion and related fields of research

The study of if, how and why policies spread between different administrative and geographical areas has a long history under many different headings and fields of academic research. In political science alone, research on the topic has been published under the headings of (policy-) 'diffusion', 'innovation', 'convergence', 'mobilities', 'process tracing', and others (Graham et al., 2013; Oliveira et al., 2023). Throughout these disciplines, scholars vary in how exactly they define diffusion (or related concepts), including notably on whether diffusion occurs over time to subsequent policies in other jurisdictions, or whether the definitions allow "policy choices to be interdependent, simultaneous or anticipatory" (Graham et al., 2013, p. 675).

[1]See e.g. Westersø (2023), Klimarådet (2023), Dansk Industri (n.d.), Udenrigsministeriet (n.d.).
[2]See e.g. Holm (2024), Jørgensen (2022).

Quantitative diffusion studies. The birth of diffusion studies is often attributed to Walker (1969), who introduced 'external factors' as a relevant consideration when analysing policy innovations in American states. Following Berry and Berry (1990)'s seminal study, which introduced 'Event History Analysis' as a quantitative approach to uniting studies of internal and external factors, the 'policy diffusion' subfield has largely been focused on large-N quantitative studies. Event History Analysis (EHA) uses statistics to model the likelihood of adopting a certain (type of) policy and its dependency on variables representing both the internal and external characteristics of the jurisdiction in question, including the influence of pairwise interactions between jurisdictions (Gilardi & Füglister, 2008; Volden, 2006).

Using these techniques, researchers have demonstrated positive correlations between policy adoption likelihoods and indicators of external influence such as: adoption in neighbouring jurisdictions (Berry & Berry, 2018; Linsenmeier et al., 2023), ideological similarities to adopters (Berry & Berry, 2018; Schoenefeld et al., 2022), trading partners and similar trade patterns to other adopters (Steinebach et al., 2021), diplomatic interactions with other adopters (Kammerer & Namhata, 2018), donors of foreign aid (Baldwin et al., 2019) and many more (Berry & Berry, 2018; Graham et al., 2013).[3] Researchers have used the estimated diffusion rates to model the resulting impacts of policy diffusion, e.g. estimating significant global reductions in carbon emissions as a result of carbon pricing policy diffusion (Linsenmeier et al., 2023).

Despite their popularity, EHAs (and other quantitative methods) show several shortcomings. EHA studies rarely engage with the actual text or substance of policies, including their mutual similarities and differences. Instead, policies are often only included so far as indicating whether or not jurisdictions have taken *some* policy action on an issue. This also limits EHAs to studying a single policy area at a time, and thereby not capturing broader diffusion trends or relationships (Oliveira et al., 2023, p.234). The largest methodological issue with quantitative investigations, however, remains causal explainability. Correlations are not to be conflated with causation, and correlations between 'external factors' in adopting jurisdictions does not alone suffice as evidence that diffusion of policies have occurred. For instance, governments might adopt policies within an issue area at similar times, simply in response to facing similar challenges (Volden et al., 2008).

Qualitative diffusion studies. Qualitative methods have been viewed as key to overcoming the challenges of quantitative analysis (Starke, 2013). For instance, the 'policy mobilities' sub-field has applied qualitative methods to follow policies through

[3] Several of these relationships are disputed, or show evolutions in relative importance over time. For instance, the importance of geographic proximity has seemingly diminished over time (Mallinson, 2021). This may be attributable to sometimes questionable choices of indicators for inter-jurisdictional contact relationships (Blatter et al., 2022).

adoption in multiple jurisdictions and study how they evolve along the way (Peck & Theodore, 2010; Temenos & McCann, 2013). Notably, this research tradition finds that policies are re-elaborated and re-invented as they spread, rather than circulating unchanged (Glick & Hays, 1991; McCann & Ward, 2013; Oliveira et al., 2023). Such process tracing is especially fruitful in illuminating causal relationships between policy adoptions (Starke, 2013).

While qualitative methods potentially patch many of the holes of quantitative diffusion studies, they also suffer from the challenges faced by qualitative methods more generally. Namely, that human investigators may arrive at different conclusions when presented with the same evidence, due to different knowledges and implicit biases (Creswell & Poth, 2016; Patton, 2014). Furthermore, humans get tired, make mistakes, or are otherwise not entirely consistent, and rarely possess sufficient language skills and technical knowledge to single-handedly analyse diverse policies from a heterogenous-language region such as Europe. All in all, qualitative processing of policy documents usually require several passes over any given textual source and rounds of consistency checks between multiple more-or-less independent researchers to arrive at accurate and reproducibly results (Saldana, 2021). This is a time- and resource-consuming process, and thus usually limits applications of these methods to small datasets.

Networks and Text-as-Data. More recently, network methods have entered the scene. Networks have been used to model diffusion networks based on ever-growing policy adoption datasets (Boehmke et al., 2020), and infer underlying policy diffusion networks (Desmarais et al., 2015). Garrett and Jansa (2015) use a text-as-data approach to match documents based on similarity and explore diffusion networks, and several more recent studies elaborate on the text-as-data approach in a diffusion context (Abel & Mertens, 2023; Chalmers et al., 2024; Linder et al., 2018). Notably, text-as-data approaches use document similarity as a proxy for diffusion, assuming that similar documents are a result of causal influence by one document on the other. As with adoption patterns in EHAs however, such similarities might be random or spurious. Furthermore, as the sub-field of Policy Mobilities has shown, policies often mutate as they diffuse, which has implications for the validity of the "similarity = diffusion" assumption. Conversely, diffusion studies using text-as-data approaches view these methods as a tool for probing the substantive contents of policies otherwise only achieved through qualitative studies. Thereby these approaches potentially offer the causal explainability of qualitative methods while avoiding the prohibitively expensive manual labour associated with large-scale reproducible qualitative investigation. The following sections will expand upon the text-as-data approach, its weaknesses, and how they might be overcome using recent advances in the computerised processing of text through *Natural Language Processing*.

2.3 Natural Language Processing

The basic barrier to performing computer-based processing of text is the fact that computers operate on *numbers*, while text is a symbolic representation of complex human language(s). Translating just a single human language to numbers is made difficult by the ambiguity of words and context, where nuances are often conveyed by subtle word choices or shared understanding unstated in the text. Metaphors, idiomatic expressions and cultural differences further complicates interpretation, as does the constantly evolving nature of language.

One simple but common way to represent a text document as numbers is to simply count the number of occurrences of unique words. This is often referred to as a 'bag-of-words' approach, and allows for easily interpretable comparisons between texts based on their word-distributions. Such approaches have been used to study policy diffusion, but fail to account for context, paraphrasing, word ordering, document length, and evolving word meanings (Abel & Mertens, 2023; Chalmers et al., 2024; Garrett & Jansa, 2015). For instance, representations for the sentences `"dog bites man"` and `"man bites dog"` would be identical. Essentially, bags-of-words do not carry much information about the substance or meaning present in the 'bagged' text. Moreover, such methods cannot be applied to compare documents in different languages.

Modern approaches for using computers to uncover the complex relationships between words in written language - the field of Natural Language Processing - all involve statistical models achieved through application of machine learning on enormous amounts of textual training data. Recent years have seen colossal progress in using machine learning models for language processing ('language models'), which has spurred much of the 'AI boom' and surrounding hype. The following sub-sections detail two such developments: Text Embeddings as a representation of text, and large language models as a processing tool for documents.

2.3.1 Embeddings

Embeddings allow for machine treatment and comparisons of words and text by creating numerical representations of text in the form of high-dimensional vectors. Here, each vector dimension might be thought of as representing some trait or attributes of the embedded text (Elhage et al., 2022; Templeton et al., 2024).[4] Modern text embedding is done through statistical models that are trained to capture the context and meaning of different strings of text, and output similar vectors when fed similar texts. These models encode information in vector spaces of several thousand dimensions, but dimensions rarely correspond cleanly to a single traits, which limits their direct interpretability (Elhage et al., 2022; Opitz et al., 2025; Templeton et al., 2024).

[4] For details and examples of how this works, see Appendix A.3

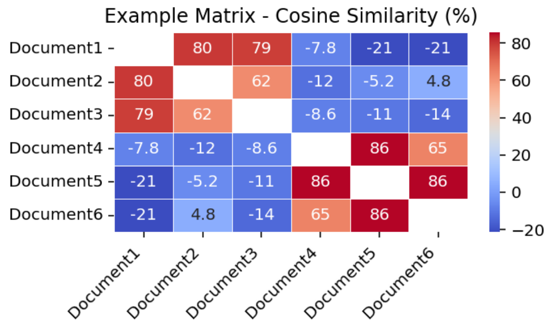


Figure 1: An example of a similarity matrix for a set of documents. Cells show pairwise similarity scores in percentages. Diagonal (self-similarity) scores have been hidden.

Aside from providing a compact and machine-friendly way to represent text, vector representations can be compared for similarity, and embeddings thereby provide an avenue for assessing how similar text snippets are to each other. Many different measures of similarity exist, but in this analysis we will be applying the most commonly used measure 'Cosine Similarity', which measures the angle θ between vector representations of text.[5] Cosine similarity provides information on whether representations of text differ, but notably only on whether the vector representations are more or less aligned, and not on *how* the underlying strings of text differ.

Whatever exact measure one uses to calculate it, similarity is a pairwise relationships between units of analysis (in this case textual documents). A common way to visualise similarities is through a similarity matrix, where rows and columns each correspond to every item in the dataset, and the pairwise similarities between vectors are entries in the corresponding row/column combinations. An example is given in Figure 1, which shows cosine similarity values for embeddings of six example documents. In this example, the data shows clear signs of having two clusters of related documents, as the first three and last three documents respectively show very high within-group similarities and very low out-group similarity.

Text embedding challenges. Currently, very capable general embedding models are commonly available, which allows direct comparison of text similarity despite differences in language, subject and context. The direct applicability of embedding models for text comparison, especially amongst legal and policy documents, is however still limited by especially two challenges:

First, embedding models have an upper limit on the character length of a given input string (termed the 'context window'). For the largest and most advances models, this limit corresponds to $\approx$ 30000 characters ($\approx$ 6000 words $\approx$ 12.5 standard pages) - far from enough to process longer policy and legal documents (Maggiori, 2023).[6]

[5]Henceforth, the term 'similarity' will be used interchangeably with 'Cosine similarity'

[6]More generally, language models process text in the form of 'tokens', which are smaller units of

The second major challenge concerns the inclusion of irrelevant context or information in embeddings. For the task at hand, we are interested in comparing the *substance* of policies and legal acts - rather than the specific language and form with which policies are implemented. However, documents often contain numerous standardised formulations, boilerplate language, and country-specific peculiarities which will influence their embeddings. For instance, Danish laws begin with the text

> "We, Frederik the Tenth, by the Grace of God King of Denmark, do hereby declare: The Parliament has passed and We, by Our consent, ratified the following law: (...)"

Including such country specifics in the text to embed would result in reduced similarity between a Danish law and a substantially identical law from another country (and likely more so if that country is not a constitutional monarchy), while increasing similarity between substantially unrelated Danish laws, based entirely on their common origin. In other words, including these parts of the text will hide and dilute policy substance in the resulting vector representation, and will thereby skew any comparison. This should not be considered a flaw in the embedding model, which is doing its job of embedding similar texts to similar vectors, but rather a result of including unwanted context or information in the text string to embed.

Fortunately, a second major advance in Natural Language Processing provides us with tools to overcome both of these challenges. The following section will detail how Large Language Models can be used for unsupervised directed extraction of information from lengthy documents - information, which can in turn be embedded and compared.

2.3.2 Large Language Models (LLMs)

Large Language Models (LLMs) is the commonly used term for modern natural language processing machine learning models. In the present context we shall use the term specifically to refer to autoregressive text-processing models with *generative* capabilities: models that process text and generate text in response.[7] These models function by iteratively generating a probability distribution for the next word in a sequence, sampling that distribution, updating the sequence with the sampled word, and repeating this process until a certain stop trigger is reached (Brown et al., 2020; Vaswani et al., 2017).[8]

text, acting as the vocabulary of the model. Tokens are typically common sub-elements of words (e.g. prefixes, suffixes, common word parts) (Radford et al., 2019). For instance, the suffix 'ing' could be a token and can follow several different verbs in their base form (such as 'eat', 'jump' and 'sleep') to conjugate these. On average, a token corresponds to approximately 4 characters in English, and the embedding model used in thesis has a context window of up to 8191 tokens.

[7] We'll use the term 'Embedding Models' to cover natural language processing machine learning models that process text and generate high-dimensional vectors.

[8] More accurately, LLMs process and predict *tokens*, see nn. 6. For introductory purposes however, thinking of tokens as individual words is sufficient and intuitive (Sanderson, 2024)

'Largeness' and Prompts. The 'largeness' of language models is not a well-defined measure, but is typically described by their parameter count. Notably, models seemingly exhibit emergent abilities when the number of parameters scale past a certain point (Fu, 2022; Wei et al., 2022). These abilities include the ability for 'common sense' reasoning, but also more complex, multi-step reasoning and problem solving (Kojima et al., 2023; Wei et al., 2023). Furthermore, model parameters store information, and larger models are thereby able to reason with inherent knowledge, without resorting to external information retrieval or fine tuning on a specific knowledge base (Allen-Zhu & Li, 2024; Yu et al., 2023). In combination, reasoning and knowledge allow models to perform language processing tasks which they were not explicitly trained on, due to a general understanding of language and context. Whereas statistical model predictions are generally only effective when based on representative training data, LLM abilities generalise and are robust when faced with inputs that differ significantly from the distribution of its training base (Si et al., 2023). This implies a great potential to use language models as general text processing tools, without specifically training or fine-tuning a model for the task at hand. For instance, even LLMs that are now considered dated due to the rapid pace of development have shown excellent performance in e.g. sentiment analysis, information extraction, deductive coding, text summarisation, translation, and much more (Chew et al., 2023; Chung et al., 2022; Jiao et al., 2023).

Under the right circumstances, language models can perform competitively to humans on a variety of complex language-based tasks, based only on brief instructions in the form of a 'prompt'. Various techniques exist for prompting models effectively and increasing performance (P. Liu et al., 2021), and the prompt is often the single-biggest, but also most easily improved, influence on model performance for a given task (Kojima et al., 2023; Wei et al., 2023; White et al., 2023). Providing examples of desired outputs in the prompt generally improves model performance, especially when faced with ambiguity or unclear tasks, where models can latch onto example formats and receive cues on e.g. labels and classifications (Brown et al., 2020; Min et al., 2022; Radford et al., 2019; Webson & Pavlick, 2022).

LLM challenges. Potential issues LLMs for text processing include the fact, that models are generally eager to follow prompt instructions - sometimes to the point of over-eagerness where they confidently output wrong, untrue or fabricated information, rather than disobeying instructions or indicating that a task cannot be solved given the information present (Ji et al., 2023). Examples include confidently placing samples into ill-fitting categories when no categories fit well, and including (potentially fabricated) information in summaries not present in the source (Kryściński et al., 2019; Maynez et al., 2020). This latter tendency to 'hallucinate' is a major issue for models that is seemingly not improved upon in more recent releases (OpenAI, 2025b).

Related to this, the black-box nature of deep-learning methods, including LLMs, fundamentally limit interpretability of model outputs (Doshi-Velez & Kim, 2017; Templeton et al., 2024). Specifically, it is difficult to probe how the model arrives at any given output, and instead one must rely on prompting the model to also return its reasoning and step-by-step thinking, as well as textual anchors that support its interpretation of text (Wei et al., 2023). Such approaches are unfortunately at best simulations of actual behaviour, and at worst post-rationalisations of potentially hallucinated outputs (Turpin et al., 2023). This fundamentally means, that scepticism is required with respect to model outputs, and that outputs are ideally manually validated, either exhaustively or for larger datasets by sampling (Calderon & Reichart, 2025; Ribeiro et al., 2016). To add to this, models have historically shown trouble processing information from longer texts, and research suggests models preference information at the beginning and end of longer inputs when generating responses (An et al., 2024; H. Liu et al., 2025; N. F. Liu et al., 2023).[9]

Randomness and Temperature. As mentioned, language models output probability distributions over words, which are sampled and updated iteratively to produce the final output. As such, LLMs are inherently stochastic models, which limits the reproducibility of results. Most LLM developers provide options to control inherent randomness through model the 'Temperature' (T) configuration setting, which controls the probability distribution from which the word a model outputs is sampled. Specifically, setting $T = 0$ implies that the most likely word is always selected, which provides a way to increase reproducibility. Temperature, however, does not control *all* stochastic processes in the model, and as such $T = 0$ does not guarantee consistent output. Some providers furthermore give the option to control the random seed of the underlying algorithms, which further increases reproducibility, but still variations can occur between runs.

Training data. Modern Language Models require massive amounts of training data, and are possibly using publicly available text quite exhaustively for this purpose (Villalobos et al., 2024). As a notable consequence, any modern model will almost certainly already have seen many of the public policy document we might show it, and thereby have some 'baked-in' knowledge of related policies. Furthermore, this means that models are trained on, and more-or-less capable in, several hundred languages, but as with any other statistical model, language models reflect the biases of their training data. Given that models are mostly trained on web data, this includes an English language bias, which is likely reflected in the models language abilities (Dodge

[9] It is yet unclear if this persists in the newer models, increasingly optimised for longer contexts and needle-in-a-haystack information retrieval (Kavukcuoglu, 2025; Meta, 2025; OpenAI, 2025a).

et al., 2021; Hu et al., 2025). Biases also include opinions that occur in training data, hereunder potentially racist, sexist and ideological positions (Bender et al., 2021). Notably, models may frame interpretations according to such biases.

In summary, LLMs provide a tool for overcoming the barriers to embedding-based textual comparison. LLMs can be used to write condensed summaries of long documents, and be specifically instructed to only include the context which pertains to the substance of a policy or regulation. Moreover, LLMs can be used to extract information from such documents on characteristics of associated policies, such as policy instruments, targetted actors, regulatory type, etc. This amounts to deductive coding, where models potentially can extract features of documents at speeds and levels of cross-document and cross-language consistency unmatched by human coders (Chew et al., 2023). Policy characteristics can in turn be used to find patterns of common elements in similar policies, potentially enabling conclusions about policy characteristics that might promote policy diffusion. The next section will detail how, given the ability to compare texts for similarity, networks can be used to elucidate and explore the similarity relationships between documents.

2.4 Network Analysis

As the number of sample pairs in a dataset grows quadratically with dataset size, visual representations through similarity matrices quickly become overwhelming and impractical except to read off all but the most superficial patterns in data.

A useful alternative approach is to construct a network graph based on the similarity matrix. Here, units of analysis (in this case policy documents) are represented as nodes, and connections are drawn between nodes, 'weighted' according to their pairwise similarity score. Networks can come in directed forms, where connections have direction, which can correspond to e.g. physical, financial, information or influence flows.

It is often fruitful to limit connections to only those which satisfy some combination of conditions and rules. These conditions represent the assumptions and expectations we have for the system of study, and limit connections to specific aspects we wish to investigate (Cao, 2010). For instance, in a directed network studying policy diffusion, where connections represent influence flows between documents, we might establish causal rules such as only allowing a node to receive connections (and thereby influence) from documents older than itself. Furthermore, we might only expect direct influence to be received from the most similar documents and/or for there to be some minimum threshold of similarity for us to consider it likely that diffusion has occurred. Notably, networks where connections only point forward in time do not have cycles, and thereby correspond to an important subset of networks called 'Directed Acyclic Graphs' (DAGs) for which the concepts of 'descendants' becomes meaningful.

Network statistics. Several measures of node importance (often called 'centrality'), useful in characterising those nodes which most likely exert or receive policy influence, have been devised for DAGs (Newman, 2010; Wasserman & Faust, 1994). Common measures include 'out-degree centrality' (the number of direct out-going connections - 'children' - for each node), and number of total descendants. Both of these have intuitive appeal, as having many children or descendants implies that one's influence has propagated to many others. The simplicity implies disadvantages however, respectively not accounting for the importance of a node's descendants (a 'dead-end' child confers the same importance as a high out-degree child) or not differentiating between close- and far-removed descendants. Instead, the primary measure of importance used here will be 'Katz centrality' - a more complicated measure commonly used in analysis of Directed Acyclic Graphs, which effectively mixes considerations of immediate connections to nodes with considerations of descendant importance.

Katz centrality can be considered a generalisation of degree centrality, at it computes the number of nodes that connect to a given node directly or through other nodes, but attenuates the contributions from further away nodes by a factor $\alpha^k \in]0,1[$, where k is the number of steps separating nodes. The choice of α is arbitrary but important, as it determines how many steps removed a node can be from its descendants, while still gaining significant 'credit' for their influence. As an example, for $\alpha = 0.35$ it takes 3 steps for the aggregated weight of a descendant to drop below 10%, 4 steps for $\alpha = 0.5$, and 6 for $\alpha = 0.65$. When calculating, nodes are given a baseline level of influence determined by the parameter β, and the influence flowing through a connection is also commonly weighted by connection strength (e.g. document similarity).

These measures represent node-level statistics, but can be aggregated based on e.g. document authorship to arrive at country-level measures of influence. This enables conclusions as to which countries are more or less successful in influencing other countries' policy. Furthermore, importance can be used to weight contributions of individual documents when studying patterns in policy characteristics such as applied instruments and targetted actors. Hereby we can determine whether some characteristics are more or less common among policies that diffuse.

Finally, the networks graphs themselves admit some observations regarding the structure of policy diffusion. Notably, we will measure the tendency of nodes to connect to other nodes with similar attributes - what's termed 'assortativity', and is measured with respect to some node attribute such as country of origin (Newman, 2010).

All in all, networks provide a useful tool for probing the contents and substance of diffusing policies, both individually and in the aggregate. With this in mind, we move on to clarifying more precisely the methods used to construct and study diffusion networks.

3 Methods

The following section describes the extensive data pipeline through which environmental policy documents are processed in the analysis to come. The pipeline consists of document pre-processing, LLM-processing, embedding, and finally 3 post-processing steps involving similarity and network analysis. The full pipeline is shown in a flowchart in Figure 2. We explain each step in full in the following subsections. We begin, however, by describing our data sources and the datasets investigated.

3.1 Datasets and data sources

The investigation in this thesis builds on analysis of four different varyingly-sized datasets of environmental policy, primarily (but not exclusively) from Europe. These include:

- The **Directives Test-set** of select EU-directives and their associated implementations in two countries. The dataset is used to test information-extraction capabilities of LLMs, the pattern-finding viability of this text-as-data approach, and as an illustration the data-processing pipeline on a familiar set of policies.

- Two datasets of policies also studied in peer-reviewed qualitative studies of policy diffusion, each used to test mostly distinct parts of the text-as-data methodology. The **Adaptation Strategies** dataset is used to validate machine-extracted textual descriptions of policies. The **Green Taxonomies** dataset is used to validate diffusion patterns.

- The **FAOLEX** dataset, representing large-N dataset of European climate-change policy documents and legislative texts, intended for large-scale investigation of diffusion patterns in European environmental policy.

The following describes how each dataset was constructed, including pretreatment and cleaning. Full tabular overviews of the first three datasets are available in Appendix F.

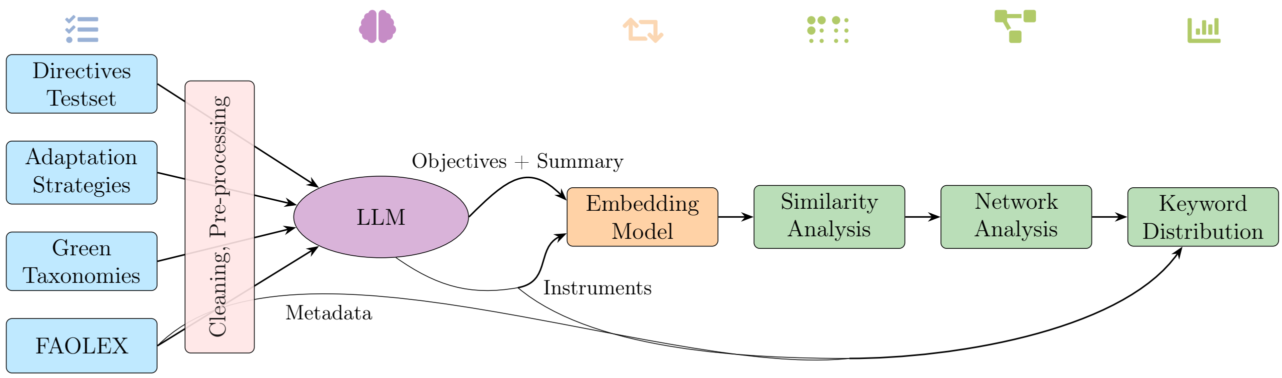


Figure 2: Flowchart showing the data processing pipeline

The Directives Test-set includes 23 documents, divided into four categories (single-use plastics, waste, biodiversity and renewable energy,) each corresponding to an environmental policy area in which the EU has taken action. The dataset was constructed by selecting four EU-directives (Single-use Plastics Directive; Landfill of Waste Directive; Habitats Directive; Renewable Energy Directive II), and for each directive, selecting a few legal texts transposing and implementing this directive from both Ireland and Denmark, starting from the lists of national transposition measures listed on EUR-Lex for each directive. The directives were chosen such that some policy areas (plastics and waste) were overlapping while others were distinct. National policies were selected to include the most relevant national implementations of a given directive. Documents were labelled into one of the four policy area categories depending on which directive they were associated to. Documents associated to multiple areas were given multiple labels. Ireland and Denmark were chosen as countries based on language, author familiarity, and to represent different legal systems and thereby provide a more diverse test-set.

Furthermore, the dataset includes 7 documents not related to any particular EU directive, but instead included to further diversify the database by including two legal texts from outside the environmental domain as well as non-legal texts with varying relation to the environmental domain.

Documents were dated according to their date of adoption. Several of the texts included in the dataset are consolidated versions of legal texts incorporating in some cases decades of changes and updates to the legal act in question - but all dated as the date of the most recent update. Furthermore, the dataset is quite small, and does not very well capture all possible sources of policy inspiration. Thereby, the dataset provides a relatively bad test for *diffusion* and causal relationships.

The Adaptation Strategies dataset is based on six documents studied in detail in Jensen et al. (2023): Climate-change adaptation strategies from Denmark (2008 and 2012), the UK (2013, 2018) and the EU (2013, 2021). The Danish 2nd generation National Adaptation Strategy was not included, despite being in the study, as no collected supporting documents could be found. Furthermore, several supporting EU-documents referred to, but not analysed, in the Jensen study, were included in the dataset, as well as the UK's 2023 adaptation plan, released after the original study.

The Green Taxonomies dataset contains 'green' and 'sustainable' financial activity taxonomies from countries, regional groupings and organisations spanning all continents. The dataset encompasses taxonomies adopted, released or drafted by countries or multilateral organisations, based on those listed and studied by Natixis (2023). In cases where both a draft and adopted version of a taxonomy could be found, the final version was preferred. When taxonomies consisted of several documents, or included appendices to a primary report, the PDF versions of all relevant documents were merged and processed as a single file.

The FAOLEX dataset is a subset of the FAOLEX Complete Collection Open Data set, one of the world's largest databases of environmental legislation, policy and related documents. The full database totals 212, 000 entries, each with associated descriptions and information on dates of enactment, subjects, policy domain, and keywords (FAO, 2025). The subset treated here is limited to those 1690 entries originating from the European geographical area,[10] which were labelled as 'Policies', 'Legislation' and/or 'Regulation',[11] and which had the associated keyword 'climate change'. The dataset of 1690 was further cleaned for entries without an associated document or without an associated date. For database entries representing updated policies, which contained both a date of original policy adoption and a date of policy updates, documents were dated according to the most recent date noted in each entry. For some entries this will have caused errors however, as the database in certain of such cases only contains the files for the original text, which would thereby mistakenly be labelled as younger than they are.[12] Some entries contained multiple documents. If any such additional documents could be identified as annexes to an original file, they were merged with the original document to a single file.[13] In other cases (e.g. documents containing different language versions), each document was treated independently. Exact duplicate files were excluded when identified, but not exhaustively or systematically.

3.1.1 Document Pretreatment

To prepare dataset the for analysis, documents were collected in PDF format, then loaded and processed in python using the `PyMuPDF` library (Software, 2025, p. v. 1.25.5). Documents were pre-processed to clean them and prepare them as input to language models. Several pre-processing methods were tested on the Directives Test-set, varying from no treatment (raw text), through automatic structuring into markdown (an easy-to read markup language LLMs are commonly trained on) to removing footnotes, headers and even appendices through semi-automated detection. These methods showed similar results, so automated preprocessing into markdown was settled on to structure the documents without the need for (semi-)manual cleaning. Specifically, documents were

[10]Defined as countries listed in Europe under the UN M49 standard (UNSD, n.d.), as well as the EU. Documents authored by multiple states were included if all were from this region.

[11]Thereby excluding the labels 'Constitutions', 'International Agreements' and 'Miscellaneous'. Entries with multiple labels were included if all type labels were one of the accepted three.

[12]It is unknown, but would be unsurprising, if the database contained reverse cases, where the date a text updated was missing but a new document version was included.

[13]Annexes were identified as those filenames with `'anx'` or `'annex'` appended.

preprocessed to markdown format using the `pymupdf4llm` package (Artifex Software, 2025, p. v. 0.0.21).

A few handfuls of documents were during this process excluded from the FAOLEX dataset, as they did not preprocess successfully. This was typically as a result of a) PDFs files created by scanning physical documents, and thus having no associated machine-readable text, or b) PDFs being malformed, and having machine-readable text only in strange character sets and formats. Furthermore we discarded those (≈ 35) documents from the FAOLEX database which processed into 300 tokens or less due to errors in the preprocessing which left most error-ridden.

3.2 Language Modelling

Following pre-processing, documents were processed and their contents analysed using Large Language Models (LLMs). A host of models are available from many different providers, so here we list the considerations made for choosing among them. These included:

- **Price and rate limits**: The model had to be cheaply or freely available, with sufficient request and token allowances to process the described data sets.
- **Context window**: The model had to accept input context of considerable length, and a sufficiently large maximum output length to contain all requested outputs (see the following subsection).
- **Accessibility**: The model had to be generally publicly available to allow reproduction of efforts, preferably without hardware requirements. The model had to be a stable release, not in experimental testing, as to avoid buggy interactions secure increase reproducibility of the results.
- **Performance**: The model had to be as advanced and well-performing as possible under while adhering to the above restrictions.

Only a few models satisfied these criteria at the launch of this project. Several of Google's Gemini models were freely available with sufficiently large rate limits to allow analysis. Unless otherwise noted, this study uses 'Gemini 2.0 Flash', which admits inputs up to 1 million tokens and outputs of 8192 tokens, which was deemed sufficient seeing as longer outputs wouldn't fit in the context window of embedding models (see section 3.2.2).[14] Alternative options include 'Gemini 2.5 Flash Preview', which should show better performance and an even longer output length but is not a stable release and often showed issues in testing when processing some documents. OpenAI's 'GPT 4.1-mini/nano' models which allow million-token inputs and longer outputs were also

[14]see nn. 6 for an explainer on tokens

considered, but Gemini 2.0 Flash was favoured due to the costs. All of these however, and others, show similar results when tested and compared on the Directives Test-set (see Appendix C).

In all cases, models were run with temperature $T = 0$ to increase reproducibility. Even so, outputs were not entirely consistent between runs, with slight variations occurring in model output. Models were accessed in python using the Gemini API and OpenAI chat completions API, respectively, through their official python SDKs (Google, 2025c, p. v. 1.11.0; OpenAI, 2025c, p. v.1.75.0).

Given the context window limits of roughly a million tokens, all but the very longest documents in all datasets were admittable. Those few that exceeded this limit (all from the FAOLEX dataset - usually a result of merging several very long appendices with a long main body file) were truncated to match the context window.

3.2.1 Prompt design and prompting strategy

To investigate the contents of policies, the language model was prompted to extract policy information from documents. All model runs were performed using the same prompt, which due to its length is displayed in Listing 1 in Appendix B. The prompt was introduced to the model through the API as a system instruction before including any document text as user content. The prompt was designed to generate comparable descriptions of policy documents by summarising objectives and substantial content of longer texts while removing boilerplate legal text and unsubstantial descriptions specific to a document's particular geographic context. Furthermore, the prompt was designed to extract information on the policy, including any policy instruments it implements. This included instructions for identifying instruments in the text, including naming and summarising each instrument while also classifying them according to their type (one of 'regulatory', 'economic' or 'soft') and extracting information regarding the actors targetted by the policy (such as 'producers', 'consumers' etc.) , the products, entities and objects affected by intervention (e.g. 'plastic waste', 'migratory birds') and the mechanisms and methods used for implementation (e.g. 'permitting', 'reporting'). Furthermore, for each instrument the prompt also instructs the model to extract textual anchors consisting of citation-quote pairs of relevant supporting passages in the text, to support validation of outputs.

Outside of instruments, the prompt also contains instructions to extract the title, date and classify the policy domain of the document, each intended for validating model output. Policy domain refers to one of 16 domains listed in the FAO dataset.

The prompt was developed through an iterative process, gradually adding and improving upon the prompt according to results of processing the Directives Test-set.

The model was in all cases prompted using 'Structured Outputs' which forces the model to adhere to a certain (JSON) schema when responding to a prompt (Google, 2025b). The schema includes strict formatting requirements for each field, including the output datatype, mandatory/optional fields and limiting response options for certain fields to a predefined list, e.g. limiting the instrument type to one of 'regulatory', 'economic', 'soft' (or 'none', if no type can be determined). For targetted actors, intervention objects, and implementation mechanisms, the schema does not restrict the allowed keywords, allows listing several keywords if applicable, and allows listing no keywords if the category is not applicable to a specific instrument. Furthermore, the schema includes descriptions for each field to further reinforce LLM behaviour. Schemas were defined using python's `Pydantic` library and fed directly to the model call using the respective model SDK (Colvin et al., 2025, p. v. 2.11.3). The `Pydantic` schema used is displayed in Listing 2, also in Appendix B.

Most documents process unproblematically but in few cases, the model's textual completion exceeds its maximum output length before completing the provided schema, and processing thereby fails. In all such investigated cases, this occurred when the model, attempting to produce textual anchors for instruments, exceeds its context window - either by finding too many (sometimes repeated) instruments, each with lengthy anchors; by finding too many relevant quotes for a single instrument (at times attempting to quote the entire text); or by otherwise getting stuck during the process (sometimes citing the same passage several times). This was mostly an issue for non-English documents, and especially those in non-Roman alphabets. This likely relates to the fact these languages and alphabets, and thereby direct textual quotes, are much more token-intensive than standard English.[15] The final prompt was designed in part to limit such errors by reducing the length of textual quotes and numbers of instruments. When processing the FAOLEX dataset, in case of such errors, documents were attempted to be processed using a different language model ('Gemini 2.5 Flash Preview 04-17' with a much longer maximum output length). If this still did not result in valid output, the document was excluded from the analysis.

3.2.2 Embedding

To allow directed comparison of the model-processed documents, the generated descriptions were represented as vectors using using OpenAI's 'Text-embedding-3-small' embedding model, which accepts input of length up to 8191 tokens and produces embeddings in a 1536-dimensional vector space. This model was chosen to balance costs and maintain a context window sufficiently large to embed a full model response in one model call, without hitting rate limits during testing (when performing many thousand embeddings).

As the generative model's output object contains several fields of information, includ-

[15]For instance, "Woman" written in Telegu requires 18 tokens (Maggiori, 2023).

ing objectives (in the field 'policy_objectives'), summaries (in 'policy_summary'), and a list of policy instruments each with corresponding information fields, documents can be compared based on multiple combinations of these informations. The main analysis proceeds by performing pairwise document comparison based on the two document-level descriptive fields 'policy_objectives' and 'policy_summary'. These were combined in that order to a single string, embedded, and compared to corresponding descriptions of other documents. Furthermore, for each extracted instrument, a descriptive string was created which joined the generated title ('instrument_title'), description ('summary_description'), type ('instrument_type'), targets ('target_actors'), objects of intervention ('intervention_objects') and implementation mechanisms ('implementation_mechanisms'). This string was also embedded, to allow direct comparison of individual instruments.[16]

Document and instrument embeddings are compared for similarity using cosine similarity (Eq. 2, Appendix A.3), calculated in python using the scikit-learn implementation (Pedregosa et al., 2011, p. v. 1.6.1). All data visualisations are created using the seaborn (Waskom, 2021, p. v. 0.13.2) and Matplotlib libraries (Hunter, 2007, p. v. 3.10.1). For each dataset, the distribution of similarity values is investigated by fitting and comparing to normal distributions. Embedding- and similarity-related distributions are computed using seaborn's histplot and normal distributions were fitted to a smoothed Kernel Density Estimate also output by these functions. Normal distributions were generated and fitted using SciPy (Virtanen et al., 2020, p. v. 1.15.2).

3.3 Network Analysis

To probe document interrelationships, influence flows, and document substance more selectively, we construct networks from document similarity data. Networks were constructed using a flexible algorithm that accommodates a variety of node-connection rules and combinations thereof. Node connections are restricted such that nodes can only receive connections - and thereby modelled influence - from older policies. In particular we restricted connections to policies adopted in previous years, due to the frequently incomplete information on adoption dates in the FAOLEX dataset, and maintain this consistency across datasets.

In addition to temporal restrictions, network construction allows varying rules for connecting nodes based on pairwise document similarity. The primary ruleset followed Garrett and Jansa (2015), where each node is connected to the previous node with which it has the highest pairwise similarity. Variations were performed on this where connections are further restricted such that the document pairwise similarity score must also exceed a certain similarity threshold to allow connections.

[16]The exact format of the embedded instrument text string can be seen in the later Table 4

We measure node importance using Katz centrality, but also calculate out-degree centrality and number of descendants, both as absolute values and as fractions of potential connections to nodes from subsequent years.

Networks are created, manipulated and studied in python using the `networkx` library (Hagberg et al., 2008, p. v. 3.4.2) and statistics computed using that same library's built-in functionalities. Katz centrality is computed using $\alpha = 0.65$, $\beta = 1$ for all nodes, and weighted by connection similarity.[17] Katz-centralities are expressed as multiples of the lowest-centrality document in each dataset, by dividing the Katz centrality for all documents by the lowest found centrality. This process renders the choice of β inconsequential. Assortativity, which describes the tendency of nodes to connect to other nodes with similar attributes, and ranges in values from -1 (disassortative; nodes connect to nodes with dissimilar attributes) to 1 (perfectly assortative), is computed using the `nx.attribute_assortativity_coefficient` method. Networks are also plotted through `networkx`, using Graphviz (Gansner & North, 2000) interfaced to python via `PyGraphViz` (Hagberg et al., 2024, p. v. 1.14) to compute node positions.

3.3.1 Keyword Distributions

Katz centrality scores as a measure of 'influence' are used to investigate if the properties of documents and policies differ between 'influential' documents and the general dataset.

For each keyword field in both model-extracted policy instruments and database policy information, a frequency distribution is constructed of keywords found across the dataset. Similar distributions are constructed for only 'influential' documents, by weighing the counts of keywords found in these documents by the documents' Katz centrality. Specifically, the weights are calculated from the Katz scores (expressed as multiples of the lowest scoring documents) by subtracting 1 (thereby assigning zero weight to documents with no outgoing connections). Notably however, the weighted and unweighted distributions are not independent since they stem from the same documents. Consequently, standard parametric tests such as χ^2, designed for independent samples, aren't suitable to differentiating between them.

Instead, we quantify the difference between weighted and unweighted keyword distributions using Jensen-Shannon divergence (JSD), which measures dissimilarity of probability distributions (J. Lin, 1991). The JSD, computed using `SciPy`'s python implementation with base logarithm e, is used as a test statistic for a permutation test to probe the significance of distributions' dissimilarity.

The permutation tests generates a range of JSD dissimilarity values that represent the values expected if the document weights were assigned independently of document content. It does so by repeatedly shuffling document weights and calculating the

[17] `nx.katz_centrality` computes in-flowing influence, so we compute on the reversed graphs

dissimilarity between the resulting (shuffled) keyword distribution and the unweighted keyword distribution. If the 'shuffled' dissimilarities only rarely exceed the JSD of the original weighted/unweighted distribution pair, it suggests a more extreme dissimilarity between the original pair than what random variation can explain (Efron & Tibshirani, 1994). If this is the case, it suggests that the weights capture a meaningful, non-random difference in keyword focus.

Formally, the null hypothesis of the test is that document weight (i.e. 'influential status') is unrelated to the keyword distribution. The hypothesis is assessed based on the p-value, which is the proportion of permuted JSD-scores greater than or equal to the observed score between the original distributions. Since the number of weight permutations are typically *much* too large to sample exhaustively, we take a Monte Carlo approach and sample 10000 shuffled permutations (Efron & Tibshirani, 1994).

4 Results

The following subsections walk through the textual processing pipeline for each of the four datasets in turn. First, language and embedding model outputs are studied in detail for the Directives Test-set, to explore the first and second research question, and to demonstrate the methodology for exploring the third and fourth. We further investigate the first, second and third research question by processing the Adaptation Strategies and Green Taxonomies datasets, and comparing results to other research on these documents, highlighting the usefulness and limits of the model in replicating case studies and investigation policy diffusion. Finally, the fourth research question is explored by use of the FAOLEX dataset, with a focus on network representations.

4.1 Directives Test-set

The Directives Test-set consists of 30 documents. 25 of these are legal texts originating from either Denmark (DK - 8), Ireland (IE - 12) or the EU (5) - whereof 23 are related to one of four policy areas: Biodiversity (6), Plastics (3), Waste (5), Renewable Energy (7), with two documents associated to multiple directives. The remaining 7 documents are classified as 'Other' or 'the unrelated test-set' and consist of documents more-or-less unrelated to environmental policy with the intention of testing our methodology outside of this policy domain or with non-legal texts. Of the 25 legal texts, 13 are consolidated versions of larger laws amended after their original adoption. The datasets consists of 4 EU directives, 10 parliamentary legal acts, and 10 delegated acts. Furthermore, the dataset contains an EU-regulation, a research paper, two news articles (on competitive cycling and the EU Clean Industrial Deal), a fiction book chapter and a food recipe. A full overview of the database is available in Appendix F.

We proceed to detail each step of the document processing pipeline. We first show examples of outputs after language model processing of the dataset documents, as well as the results of sample spot-checks. We then study the embeddings of language model outputs, and show that these conform to the expectations for the dataset on a macroscopic scale. Finally, we illustrate the use of network methods on the dataset.

4.1.1 Language model outputs

Table 1 shows examples of language model outputs for seven documents in the dataset. From the descriptions, it is clear that the prompt used is not entirely effective in removing all references to document origin, as all descriptions of objectives and summaries begin with referencing document type ("This directive/law/article ...") and in very few cases (e.g. *DK Miljøbeskyttelsesloven*) even references the name of the act, contrary to instructions. Minor related issues are single mentions of relevant agencies

Table 1: Examples of model-generated descriptions of policy objectives and summaries for documents in the Directives Test-set

Policy	Objectives	Summary
EU Single-use plastics directive	The directive aims to reduce the impact of certain plastic products on the environment and human health, promoting a transition to a circular economy by fostering sustainable business models, products, and materials.	This directive addresses the environmental impact of single-use plastics, oxo-degradable plastics, and fishing gear containing plastic. It focuses on reducing marine litter and promoting a circular economy. The strategy involves consumption reduction targets for specific single-use plastics, bans on certain products, design requirements, marking requirements, and extended producer responsibility schemes. It also includes separate collection targets for beverage bottles and awareness-raising measures for consumers. The directive targets producers, importers, and consumers of single-use plastic products and fishing gear, as well as Member States who are required to implement the measures and report on progress.
DK Miljøbeskyttelsesloven	The law aims to protect nature and the environment, promote sustainable societal development, prevent and combat pollution, establish hygiene standards, limit resource waste, promote cleaner technology, and encourage recycling and waste management.	The law on environmental protection aims to safeguard nature and the environment while promoting sustainable development. It addresses pollution prevention across various domains (air, water, soil), establishes hygiene standards, and emphasizes resource conservation. The strategy involves regulating polluting activities, promoting cleaner technologies, and implementing waste management measures. Key instruments include setting pollution limits for businesses and transport, establishing environmental quality standards, mandating environmental impact assessments, implementing extended producer responsibility schemes, and promoting recycling. The law targets a wide range of actors, including businesses, producers, importers, public authorities, and consumers, and focuses on various sectors such as industry, agriculture, waste management, and transportation.
IE Waste Management Act	The act aims to prevent, manage, and control waste, implement relevant European Community acts, and amend existing environmental legislation.	This legislation provides a framework for waste management, encompassing planning, reduction, recovery, and disposal. It establishes responsibilities for waste holders, collectors, and producers, and introduces permitting and licensing systems for waste facilities and collection services. The legislation promotes waste prevention and recovery through various measures, including waste management plans, economic instruments, and public awareness campaigns. It also grants powers to local authorities and the Environmental Protection Agency (EPA) for monitoring, enforcement, and remediation of environmental pollution caused by waste. The act addresses specific waste streams, such as hazardous waste, end-of-life vehicles, and waste electrical and electronic equipment.
DK Habitatbekendtgørelsen	The order designates international nature protection areas and establishes rules for the administration of the areas. It also lays down rules for the protection of animal and plant species.	This legislation designates and administers international nature protection areas, including bird protection areas, habitat areas and Ramsar areas. It sets binding regulations for authorities regarding planning and administration that affect these areas, in accordance with international agreements. The legislation specifies considerations authorities must take into account regarding raw material planning and general administration within these protected zones. It mandates assessments of projects that could significantly impact Natura 2000 areas, requiring impact assessments and disallowing permissions for projects that would harm the integrity of these areas. The legislation also addresses the protection of animal breeding and resting areas, as well as plant species, and outlines conditions for potential exemptions, including considerations of public interest and environmental impact.
IE Renewable Energy Regulations 2020	The regulations aim to implement articles of a European Union directive regarding support schemes for renewable energy and the stability of financial support for renewable energy projects.	These regulations outline the application of support schemes for renewable energy, ensuring incentives for market integration, responsiveness to market conditions, and avoidance of market distortions. The support schemes are designed to maximize the integration of electricity from renewable sources and ensure producers respond to market price signals. Support is granted through open, transparent, competitive, and non-discriminatory processes, with considerations for small-scale installations and regional diversification. The regulations also address the stability of financial support for renewable energy projects, preventing revisions that negatively affect the rights and economic viability of supported projects. A long-term schedule for support allocation is to be published, along with periodic assessments of the effectiveness and distributive effects of the support schemes.
Fiction chapter - BlackCompanyVol1Chap4	The policy objectives are to eliminate a key enemy commander and disrupt enemy operations, gather intelligence, and maintain control over strategic locations.	This chapter describes a military unit's operations in a war-torn region. The unit stumbles upon a Rebel base camp, leading to the capture of crucial intelligence. This intelligence reveals the enemy's plans and the presence of a high-ranking enemy general. The unit is then tasked with eliminating this general and disrupting enemy operations. The chapter also explores the moral complexities of war and the psychological toll it takes on soldiers. It highlights the strategic importance of key locations and the constant struggle for control in a conflict-ridden environment. Ultimately, the unit succeeds in its mission, but the long-term consequences of their actions remain uncertain.
News - Sports	The article reports on a cyclist winning a point-based competition in a cycling race.	The article reports on a cyclist's performance in a multi-stage cycling race, specifically highlighting their success in securing the green jersey, which is awarded based on points accumulated throughout the race. The cyclist demonstrated strong performance by initiating attacks and maintaining a competitive position, ultimately winning the jersey.

(e.g. 'Environmental Protection Agency (EPA)' under *IE Waste Management Act*) and objectives listing the intent to implement EU legislation. Otherwise, the model generates descriptions using neutral, generic language and neither policies nor region of origin are directly identifiable from the generated descriptions.

Notably, the model generates objectives and summaries even for texts with no policy content, such as the fantasy chapter and the cycling report. This illustrates well the model's eagerness to solve the task given, even when the data does not correspond to that task.[18] The descriptions for these documents did however vary more given slight perturbations to the prompt or different model choices, likely owing to how ill-fitting documents are to the prompt (see Appendix B, C). Outside of the generous interpretations of 'policy' described above, we did not find false information or hallucinations in outputs. Checks were however concentrated on shorter, Danish and/or edge-case documents familiar to the author. Thus, checks were not performed systematically on a representative sample, and in particular did not focus on the potential issues of processing longer texts described in Section 2.3.2.

Instruments. To further probe to content and substance of policies, the model was also prompted to extract information on the policy instruments implemented in each document. This resulted in a total of 255 instrument descriptions. Decidedly factual errors or hallucinations weren't found in those instrument descriptions manually inspected, but interpretations were at times questionable. For instance, Table 2 shows instruments found when processing the Danish Nature Protection Act. The model classifies the land-use agreements with private land-owners (no. 11) as a regulatory instrument, despite the (possible) economic nature of such agreements. In fact, the table shows no economic or soft instruments whatsoever. This dominance of regulatory instruments is true for all documents in the dataset, where 219 of the 255 instruments found in total were classified as regulatory. Economic instruments follow with 27, while 'soft' instruments only number 7 (2 instruments were left unclassified). This indicates that the model greatly favours the regulatory classification.

How the model grouped and differentiated between instruments was inconsistent between documents and between different treatments of the same document. Seven rows in Table 2 are highlighted, showing four instruments titled 'Area Restriction' and three titled 'Building Restriction'. These refer to the law's various building- and protection lines around churches, ancient monuments, lakes, streams, coasts, dunes and forests, as well as general restrictions to protect the state of nature in these and other areas. Such instruments could arguably have been grouped into just one or two instruments, as they indeed were in previous prompt versions and by different models.[19]

[18] Other models noted the lack of policy contents, albeit not consistently. See Appendix C.

[19] See Appendix B, C. While different models grouped these specific instruments, they split others.

Table 2: Instruments found in the Danish Nature Protection Act (LBK 927, 28/06/2024). Highlighted rows correspond to instruments with similar titles. Textual anchors and type evidence phrase are omitted from the table but in all cases point to relevant supporting passages in the text.

	Instrument Title	Description	Type	Target actors	Intervention Objects	Implementation Mechanisms
1	Area Restriction	Prohibits changes to the condition of protected nature types, including lakes, streams, heaths, bogs, salt marshes, fresh meadows, and biological grasslands, exceeding 2,500 m2. Restrictions also apply to smaller bogs connected to protected lakes or streams.	Regulatory		'Lakes', 'Streams', 'Heaths', 'Bogs', 'Salt marshes', 'Fresh meadows', 'Biological grasslands'	
2	Activity Restriction	Prohibits spraying, fertilization, conversion, or soil treatment on protected areas, with exemptions for military activities and special cases approved by the environmental minister.	Regulatory		'Protected areas'	
3	Area Restriction	Prohibits changes to the condition of dune-protected areas, including construction, planting, terrain changes, fencing, camping, and land division.	Regulatory		'Dune protected areas'	
4	Land Use Prohibition	Allows the environmental minister to issue orders regarding the use of dune-protected areas and prohibit certain forms of use to prevent sand drifting.	Regulatory		'Dune protected areas'	
5	Area Restriction	Prohibits changes to the condition of beach areas and other areas between the beach and the coastal protection line, including construction, planting, terrain changes, fencing, camping, and land division.	Regulatory		'Beach areas', 'Coastal areas'	
6	Building Restriction	Restricts the placement of buildings, camping wagons, and similar structures, as well as planting or terrain changes, within 150 meters of lakes (at least 3 ha) and streams registered with a protection line.	Regulatory		'Lakes', 'Streams'	
7	Building Restriction	Prohibits the placement of buildings, camping wagons, and similar structures within 300 meters of forests that are privately owned only if the area constitutes at least 20 ha of continuous forest.	Regulatory		'Forests'	
8	Area Restriction	Prohibits changes to the condition of the area within 100 meters of ancient monuments protected under the Museums Act, including establishing fences or placing camping wagons.	Regulatory		'Ancient monuments'	
9	Building Restriction	Restricts the construction of buildings taller than 8.5 meters within 300 meters of a church, unless the church is surrounded by urban development throughout the protection zone.	Regulatory		'Churches'	
10	Activity Notification	Requires written notification to the municipal council before initiating activities listed in Annex 2 within international nature conservation areas, to assess the impact on the area's conservation objectives.	Regulatory		'International nature conservation areas'	'Notification', 'Impact assessment'
11	Land Use Agreement	Allows municipal councils to enter into agreements with property owners or users in international nature conservation areas regarding operations or other measures to realize the Natura 2000 plan.	Regulatory	'Municipal councils', 'Property owners', 'Property users'	'International nature conservation areas'	'Agreements'
12	Land Use Mandate	Requires municipal councils to mandate property owners in or outside international nature conservation areas to undertake operations or other measures necessary to realize the Natura 2000 plan if agreements cannot be reached.	Regulatory	'Municipal councils', 'Property owners'	'International nature conservation areas'	'Mandates'
13	Public Facility Planning Mandate	Requires public facilities in open land to be located and designed to consider landscape values and other interests outlined in § 1.	Regulatory		'Public facilities'	'Planning', 'Design'
14	Advertising Restriction	Prohibits the placement of posters, images, freestanding signs, light advertisements, and other devices for advertising and propaganda purposes in open land.	Regulatory		'Advertisements'	
15	Waste Disposal Prohibition	Prohibits the disposal or placement of waste or similar items on someone else's property without the owner's permission.	Regulatory		'Waste'	
16	Species Disturbance Prohibition	Prohibits the intentional disturbance of animal species listed in Annex 3, with harmful effects on the species or population, across all life stages.	Regulatory		'Animal species'	
17	Habitat Damage Prohibition	Prohibits damaging or destroying breeding or resting areas for species listed in Annex 3.	Regulatory		'Habitats', 'Breeding areas', 'Resting areas'	
18	Species Management Plan	Allows the environmental minister to develop management plans and implement other measures, including subsidies, to conserve species or populations listed in Annex 3.	Regulatory	'Environmental Minister'	'Animal species'	'Management plans', 'Subsidies'
19	Species Exploitation Regulation	Allows the environmental minister to establish rules to protect or regulate the exploitation of wild animal and plant species, including conservation, registration, and marking.	Regulatory	'Environmental Minister'	'Wild animal species', 'Wild plant species'	'Conservation', 'Registration', 'Marking'
20	Species Introduction Prohibition	Prohibits the release of non-native animal species into the wild without the environmental minister's permission, including the sea and fishing territory.	Regulatory		'Non-native animal species'	'Permitting'
21	Reed Cutting Restriction	Restricts reed cutting during the period from March 1 to October 31, requiring permission from the environmental minister.	Regulatory		'Reeds'	'Permitting'

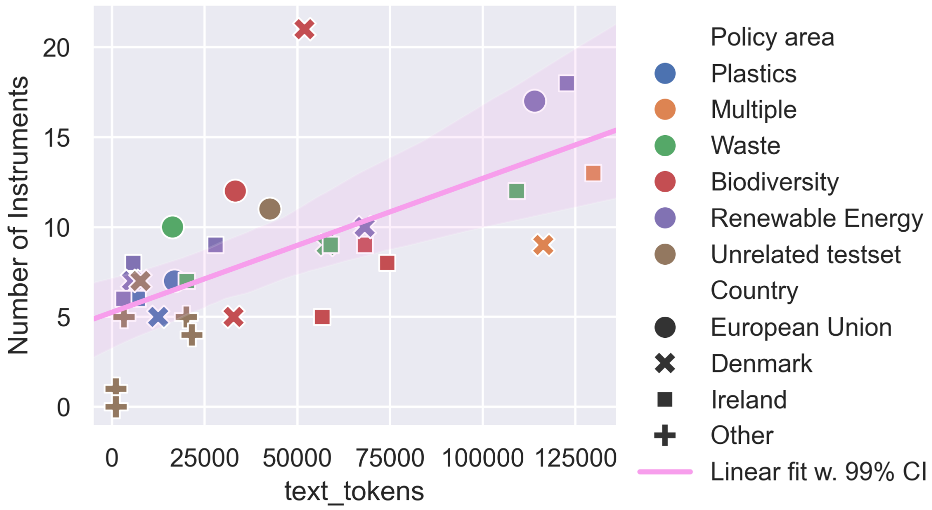


Figure 3: Scatter plot showing document length and number of instruments extracted. Documents are coloured by policy area, shaped according their origin, and have been fit with a linear function. The red outlier corresponds to the Table 2 instruments.

The variance displayed by model-found instruments also extended to omissions, where the model varies significantly in what instruments it leaves out. In the case of the Danish Nature Protection Act, neither the instruments in Table 2 nor textual anchors make reference to conservation areas ('fredning') or conservation boards ('fredningsnævn'), despite their important role in Danish conservation law and their central (both figurative and literal) role in the legal act. Different model- and prompt treatments of the document did however pick up on these instruments, and the model did find other instruments from near the middle of the text, suggesting that the issue cannot be attributed solely to language models struggling with extracting information from the middle portions of longer texts. Generally, the exact wording of instruments, the delineation between instruments, and omissions were quite sensitive to e.g. the wording of the prompt (Appendix B), model choice (Appendix C) and changed even between model runs seemingly controlled for random factors through setting the temperature. This is in contrast to the document-level summaries and especially objectives which were much more consistent in face of varying prompts and models.

Overall, more instruments were found in longer texts, as displayed in Figure 3. This is reassuring, as it indicates that the model does not aim for finding a certain number of instruments, whereby it might artificially split up instruments (for shorter texts) or group instruments (for longer documents) more aggressively than a neutral treatment would warrant. It is also reassuring that for two documents without any policy relevance, the recipe and sports news article, only a single instrument is found between them (relating to a food carbon emissions calculator on the recipe website). On the other hand, the four instruments found in the fantasy chapter show once again

Table 3: Instruments found in the fantasy book chapter (Cook, 1984, chap. 4).

	Instrument Title	Description	Type	Target actors	Intervention Objects	Implementation Mechanisms
1	Military Action	Describes military actions, including reconnaissance, ambushes, and battles, aimed at eliminating enemy forces and disrupting their plans. Actions target enemy camps and key personnel.	Regulatory	'Military Units', 'Enemy Forces'	'Enemy Camps', 'Key Personnel'	'Reconnaissance', 'Ambush', 'Elimination'
2	Intelligence Gathering	Involves collecting and analyzing enemy documents and information to understand their strategies and plans. This includes deciphering orders, identifying key personnel, and uncovering hidden agendas.	Regulatory	'Military Units', 'Intelligence Personnel'	'Enemy Documents', 'Strategic Plans'	'Deciphering', 'Analysis', 'Identification'
3	Strategic Planning	Focuses on developing and adapting military strategies based on gathered intelligence and changing circumstances. This includes identifying key objectives, allocating resources, and coordinating troop movements.		'Commanders', 'Military Strategists'	'Troop Movements', 'Resource Allocation'	'Coordination', 'Adaptation', 'Objective Setting'
4	Coercive Transformation	Involves the forced conversion of an enemy leader into a loyal subject through a dark ritual, altering their allegiance and granting them new powers.	Regulatory	'The Lady', 'Whisper'	'Loyalty', 'Allegiance'	'Ritual', 'Mind Control'

Table 4: Side-by-side comparison of the exact text (including field labels) embedded for the two most similar instrument pairs between the European Single-Use Plastics Directive and the Danish *Engangsplastbekendtgørelse*

EU Single-Use Plastics Directive	***DK Engangsplastbekendtgørelse***
TITLE: **Market Restriction** DESCRIPTION: Prohibits the placing on the market of specific single-use plastic products, such as cutlery, plates, straws, beverage stirrers, and expanded polystyrene food and beverage containers, as well as products made from oxo-degradable plastic. TYPE: Regulatory ACTORS: ['Producers', 'Distributors'] OBJECTS: ['Single-use plastic products', 'Oxo-degradable plastic products'] MECHANISMS: ['Prohibition of sale']	TITLE: **Product Ban** DESCRIPTION: Prohibits the marketing of specific single-use plastic products, including cotton buds, cutlery, plates, straws, stirrers, balloon sticks, and food and beverage containers made of expanded polystyrene. Also bans products made of oxo-degradable plastic. TYPE: Regulatory ACTORS: ['Producers', 'Importers', 'Economic operators'] OBJECTS: ['Single-use plastic products', 'Oxo-degradable plastic products'] MECHANISMS: None
TITLE: **Product Design Requirement** DESCRIPTION: Requires that single-use plastic beverage containers with plastic caps and lids can only be placed on the market if the caps and lids remain attached to the containers during use. Also sets minimum recycled plastic content for beverage bottles. TYPE: Regulatory ACTORS: ['Producers'] OBJECTS: ['Single-use plastic beverage containers'] MECHANISMS: ['Standardization', 'Minimum recycled content']	TITLE: **Product Standard** DESCRIPTION: Sets requirements for single-use plastic beverage containers, mandating that caps and lids remain attached during the use phase. Also mandates minimum recycled plastic content in single-use plastic beverage bottles made of polyethylene terephthalate. TYPE: Regulatory ACTORS: ['Producers', 'Importers', 'Economic operators'] OBJECTS: ['Single-use plastic beverage containers', 'Single-use plastic beverage bottles'] MECHANISMS: ['Documentation', 'Calculation', 'Reporting']

that the model is eager to satisfy instructions and will stretch the meanings of source material to do so (see Table 3). It should be noted however, that far from all document instruments are untrustworthy or erroneous. Several documents show very reasonable and consistent instruments, such as those found in Table 4, which shows a comparison of two instruments each from closely related single-use plastics documents.

This highlights some of the issues of model-extracted instrument descriptions, and thereby some major challenges for probing the substantive content of policies, which has obvious implications for the answer to our first research questions. As a result, in the following sections we turn our main attention to the comparison of the more consistently worded objective and summary descriptions, and investigation of the second research question.

4.1.2 Similarity analysis

Generated descriptions of objectives and summaries were combined and embedded as a single string. The resulting vectors were then compared pairwise by cosine similarity to arrive at the similarity matrix displayed in Figure 4. The data has been sorted by help of an agglomerative clustering algorithm, the dendrogram of which is also displayed. Higher similarity values are are shown in red, and patches of red are apparent in the matrix, highlighting groups of mutually similar documents. Clusters clearly correspond well to the respective document policy areas, with the documents not representing environmental policy being spread around somewhat. This is encouraging, as the policy area information is not something the language model (or clustering algorithm) has been exposed to, and thereby the cluster structure emerges entirely from the language model's textual interpretation in the form of generated objectives and summaries. The 'Biodiversity' and 'Renewable Energy' policy areas are clearly distinguishable from the 'Plastics' and 'Waste' areas, which are more closely related, yet not entirely mixed. Again, this conforms to expectations and corresponds exactly to the logic applied during dataset construction.

Three documents with no policy relevance are separated out from the rest, as shown in rightmost and bottommost rows of the similarity matrix. The embeddings of the *GDPR* and the Danish *Ligestillingsloven* are also quite different to the rest of the dataset, across the board. These two show neither particularly strong or weak signals to any other items in the dataset. Notably though, they are more similar to every other legal text than to the rest of the 'unrelated' subset of documents. This indicates that some of their legal-text nature is shining through the descriptions and resulting in higher similarities to other legal texts across the board. It also indicates that we should expect some baseline similarity between documents, with variance around this mean depending on the specific contents and types of the documents. The 'shining through' can partially be explained by the previously noted mentions of document type in objectives and summary descriptions, but is also reflected in the substance of descriptions.

In similar fashion, the non-legal nature of the *EU Green Politics* news article and the *Vietnam RE* research paper likely dampen their similarity to legal documents. In a reflection of common subject area, they are clustered with the renewable energy documents, but are notably peripheral to the cluster - possibly due to their non-legal nature. Still, their scores are not *much* lower than e.g. the Irish *Electricity Regulation Act*, which *is* a legal document, so evidently legal status alone is not sufficient to secure high similarity scores. In fact, only a single document pair from the RE cluster has a very high similarity score that truly stands out, which may indicate greater variance amongst documents in this cluster compared to e.g. the biodiversity or plastics clusters.

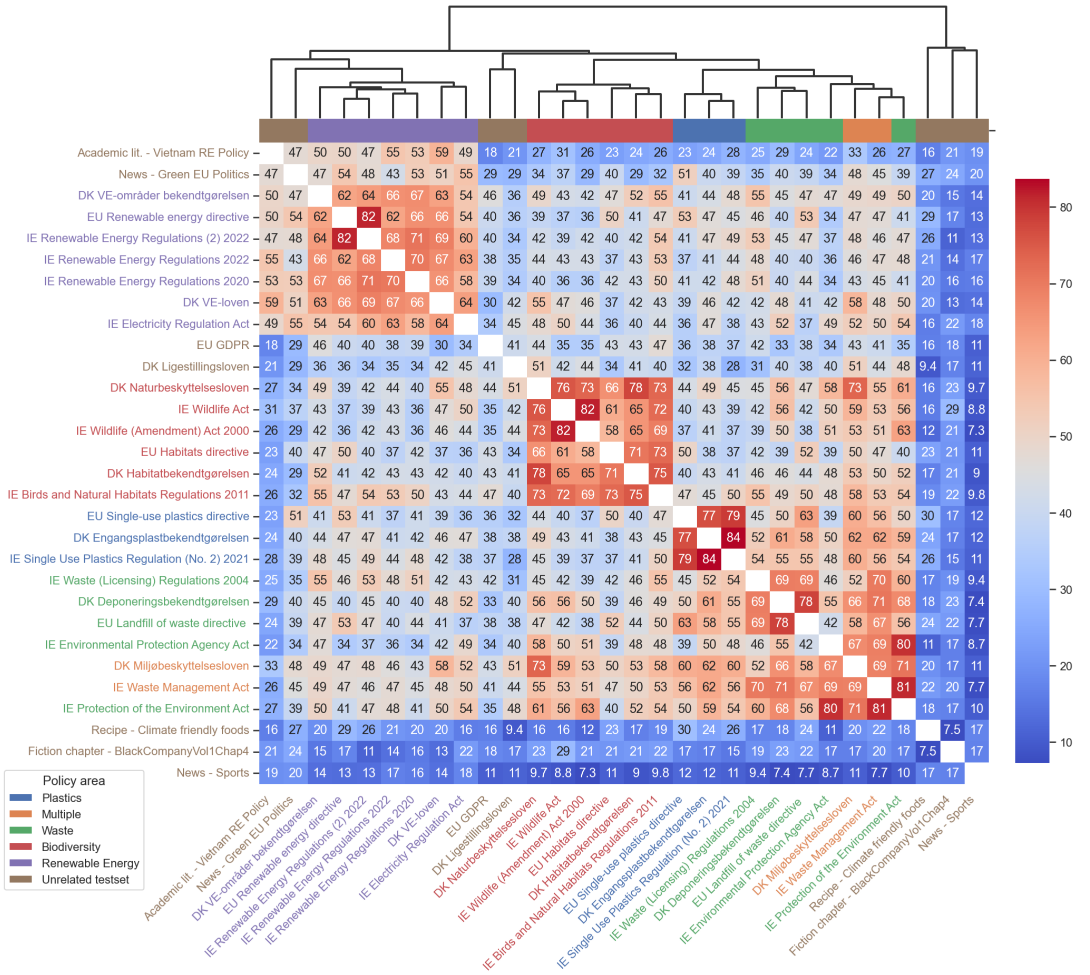


Figure 4: Similarity matrix of document embeddings for the Directives Test-set. Each matrix cell shows the cosine similarity (in %) of the corresponding document pair. Above the plot is shown an agglomorative clustering dendogram generated using `seaborn`'s `clustermap` function (UPGMA and euclidian distance) (Waskom, 2021). Rows and columns have been sorted according to the clusters. The topmost matrix row repeats policy area colour codes to illustrate the cohesiveness of policy areas.

As expected, the plastics documents show low-to-average similarities with the biodiversity and RE documents, but higher-than average similarities with the waste-labelled documents. Notably however, most of the biodiversity documents also show stronger-than-average similarities to the multi-tagged documents, as well as the Irish *Protection of the Environment Act*. This is not a big surprise, given that these documents represent big, broad laws, which dip their toes in multiple policy areas, including conservation efforts, but it does illustrate that the model-generated descriptions capture multiple facets of the laws in question. The final result is that the biodiversity cluster is linked more closely to the plastics-waste cluster than to the RE-cluster.

Similarity analysis could in principle also be applied to the *instruments* extracted from the dataset documents, but given the limited success of the instrument extraction, we relegate the demonstration of how to Appendix D.

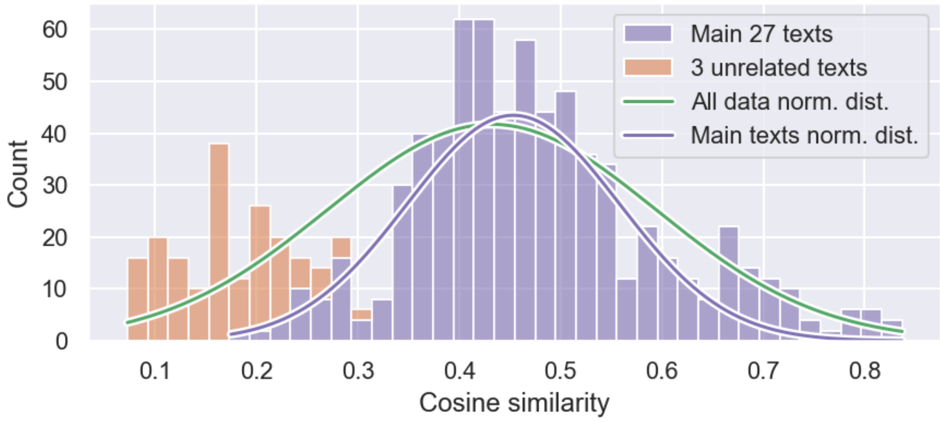


Figure 5: Histogram showing the distribution of pairwise similarity scores from the similarity matrix (Figure 4). The histogram shows stacked counts of two categories, which separate out the three lowest-scoring texts identified in Figure 4 from the rest of the dataset. Alongside is plotted the best-fitting normal distribution for both the entire dataset (green), and the dataset when trimmed of outliers (purple). These were fitted to a Kernel Density Estimate of the underlying similarity scores.

Similarity distribution. Visually inspecting the similarity matrix indicates that a pairwise similarity score approaching 60 is, seemingly, enough to indicate a common cluster membership for this dataset. We can quantify this observation somewhat by inspecting the distribution of similarity values in the matrix. Figure 5 shows the histogram of similarity scores, where the scores from rows and columns corresponding to the three outlier texts with no policy relevance have been coloured differently from the 'main' set. Alongside are plotted two normal distributions, fitted to the whole dataset and the main subset respectively. These illustrate what the histogram would resemble, if the similarity scores were normally distributed around the dataset's mean.

The full histogram shows a rough resemblance to the (green) normal distribution, albeit with gaps and a heavy tail of towards lower scores. This tail however corresponds to the many cells of the matrix with low pairwise scores between legal documents and the three 'unrelated' documents. When these are omitted from the analysis, the distribution is clearly closer to the refitted normal distribution (purple). However, a (somewhat less dramatic) tail appears on the high-similarity end of the histogram.

Whereas the low-score tail is caused by deliberate inclusion of badly-matching documents, this tail is explained by document-pairs that match *better* than random variance would explain. Some effect is seemingly boosting certain document pair similarities, resulting in a higher-than-normal frequency of high-similarity pairs. While any single pair could score highly due to random variance, the abnormality of the distribution makes it unlikely that *all* of them do so. In fact, reinspecting the similarity matrix shows that the very highest similarity values are *only* found between a) closely related laws from the same nation) b) direct, faithful transpositions of EU-directives.

Seemingly, common influence results in skewing the similarity distribution towards higher values.

This thereby gives us an idea of what to look out for when searching for evidence of policy diffusion. Though far from providing a rigorous test or evidence for diffusion, the skewed distribution certainly suggests that some effects worth investigating are influencing the document pairs with highest similarities. That the histogram counts begin to exceed the fitted distribution when similarity approaches 0.65, may indicate 'activity' occurring here. We might suspect this represents a sort of 'minimum similarity threshold' a given document pair must surpass for us to form a suspicion of diffusion.

Different treatments of the dataset documents leads to slight differences in the resulting embeddings, as demonstrated in Appendices B and C. This leads to variations in the pairwise similarities, and thereby in principle the clustering, as well as the similarity distribution and the implied similarity threshold. The larger-scale patterns such as clusters and distribution shape are overall robust to these perturbations, however, and the data shows the same skewing towards higher similarity scores across treatments. Encouraged by this apparent, or at least potential correlation between document-pair similarity and policy inspiration, we move on and illustrate how pairwise similarities can be used to draw networks and investigate central policies.

4.1.3 Network Analysis

Armed with pairwise similarities, we can construct a network by creating nodes corresponding to each document, and connecting them based on similarity. Following Garrett and Jansa (2015) we construct a directed network where each node receives a connection from most similar node *older* than itself. Note that this particular dataset is a relatively bad example for investigating the temporal aspect of diffusion, as it mostly contains consolidated laws, where the adoption timestamp doesn't match the actual time of policy innovation. Nonetheless, the dataset provides a good illustration of the procedure. The resulting network is visualised in the topmost pane of Figure 6.

Immediately apparent is that once again, nodes are clearly clumped according to the policy area divisions. However, the graph also displays connections that are clearly artefacts of simply connecting to the most similar candidate. These include the connection from the fiction chapter (the oldest text) to the *Habitats Directive*, the connection from the *Habitats Directive* to the *Landfill of Waste Directive*, and in fact most of the connections to the unrelated subset. We don't expect these connections to actually represent influence or policy inspiration, especially given their very low similarity scores (compare Figure 4). A way to incorporate this intuitive expectation is mandating a minimum similarity threshold to allow connections, for instance using the similarity threshold as estimated in the previous subsection. The exact threshold we choose as $Similarity > 0.65$, which corresponds to the jump in the Figure 5 histogram,

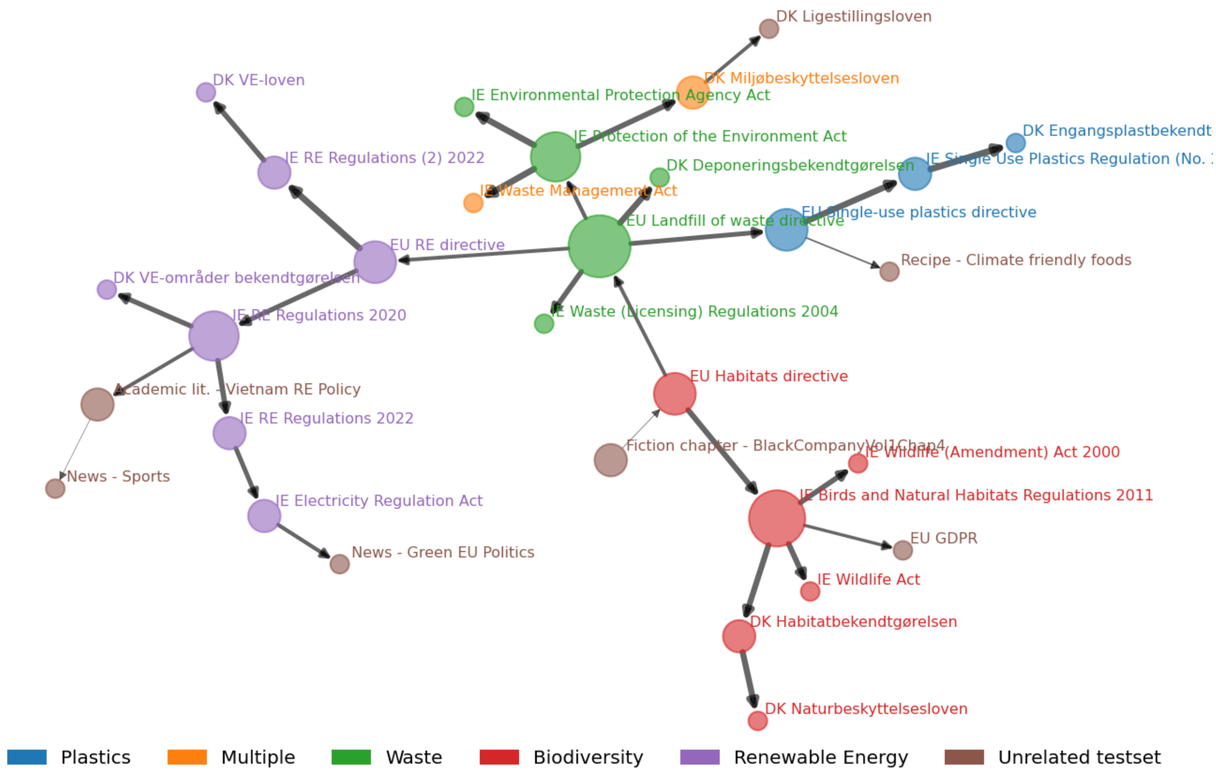
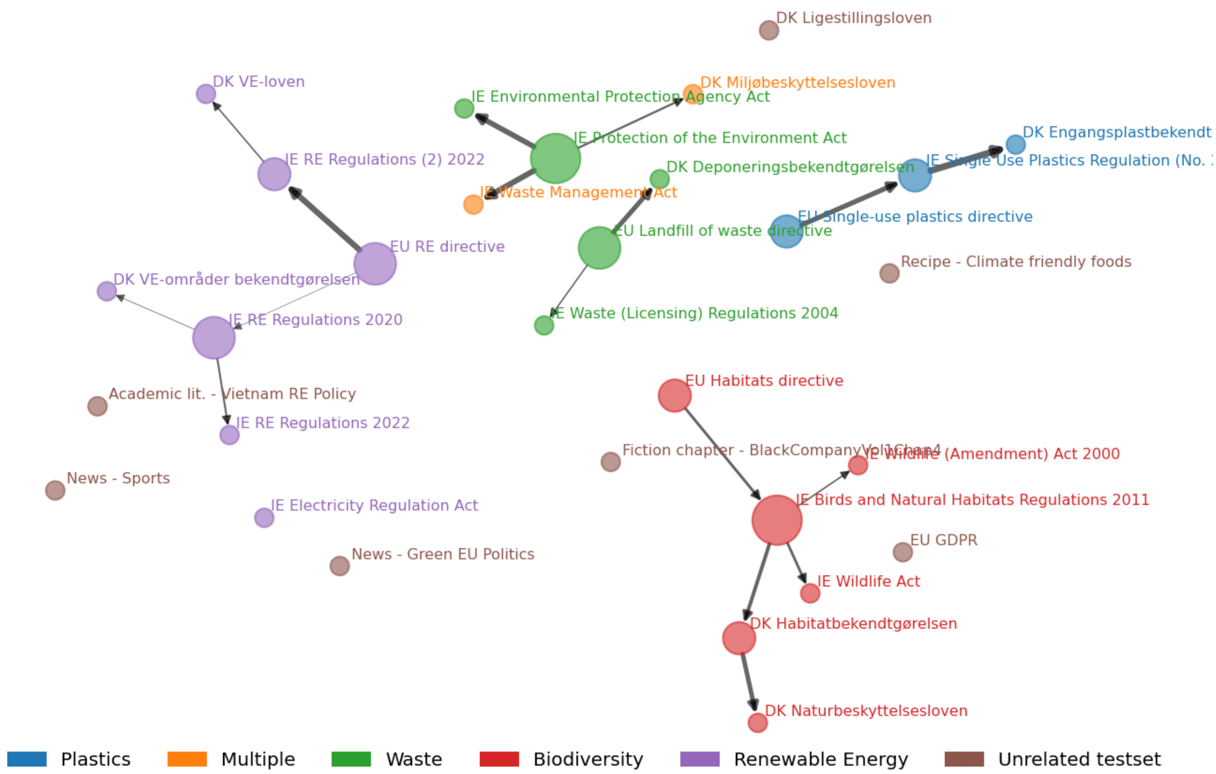


Figure 6: Network drawing of the Directives Test-set with connections drawn to the most similar previous policy (**above**) and restricted to connections with pairwise similarity above 0.65 (**below**). Nodes are sized according to their out-degree, and thicker connections indicate higher pairwise similarity (scaled visually to those connections present in the graph).



and the results of applying this threshold are displayed the lower pane of Figure 6. What was previously connected as a single network splits into several sub-graphs. These correspond roughly to the policy area divisions, but especially the waste/plastics-area has been fragmented into fairly specific branches of policy. Furthermore, the minimum similarity restriction leads to several isolated nodes, corresponding mostly to those documents with low across-the-board similarity scores, but notably not only these.

Network Statistics. We can quantify our observations regarding the network using statistics. At the macro level, the similarity-restricted network is strongly assortative by policy area (0.85)[20] but non-assortative by type (0.10) and country (-0.03). Visually inspecting the network, we'd expect nodes such as the *EU RE-Directive* and the Irish *Birds and Natural Habitats Regulations* to be the most 'important' or 'influential' nodes in each sub-graph, and isolated nodes to be the *least* influential. To quantify this, Table 5 displays an overview of network statistics for nodes in the network with at least one outgoing connection. The different centrality measures are fairly well-correlated, except that the measures as fraction of younger nodes are susceptible to outliers in the form of recent policies with connections. The ordering of nodes, sorted by 'importance' as given by Katz centrality score, aligns well with expectations.[21]

Table 6 displays the statistics when averaged across country, policy areas and document type. The standard deviation among the group is given in parenthesis. Clearly, the EU is exerting more influence than Irish and Danish documents, also reflected in document type, where Directives (exclusive to the EU) score highest. This

Table 5: Statistics for nodes in the *Similarity* > 0.65 network with at least one outgoing connection, sorted by Katz Centrality. Katz-centralities are displayed as multiples of the lowest-scoring document.

	Out-degree	Out-degree as fraction of younger nodes	Total descendants	Descendants as fraction of younger nodes	(Out-going) Katz centrality	Node weight
Birds and Natural Habitats Regulations 2011	3	0.13	4	0.17	2.66	1.66
Renewable energy directive	2	0.09	5	0.23	2.58	1.58
Protection of the Environment Act	3	0.12	3	0.12	2.51	1.51
Habitats directive	1	0.04	5	0.18	2.26	1.26
Landfill of waste directive	2	0.07	2	0.07	1.96	0.96
Renewable Energy Regulations 2020	2	0.11	2	0.11	1.89	0.89
Single-use plastics directive	1	0.05	2	0.10	1.79	0.79
Single Use Plastics Regulation (No. 2) 2021	1	0.06	1	0.06	1.54	0.54
Habitatbekendtgørelsen	1	0.08	1	0.08	1.51	0.51
Renewable Energy Regulations (2) 2022	1	0.07	1	0.07	1.45	0.45
All other nodes	0	0.00	0	0.00	1.00	0.00

[20] Skewed down from one by the somewhat artificial division of the 'waste'/'multiple' policy area

[21] The absence of Danish documents is likely due to legal consolidation, resulting in the included documents being dated very recently and thus having few documents to 'exert influence onto'.

Table 6: Node statistics averaged by country, policy area, and document type (standard deviation in parenthesis).

	Out-degree		Out-degree as fraction of younger nodes		Total descendants		Descendants as fraction of younger nodes		(Out-going) Katz centrality		Node weight	
European Union	1.2	(0.84)	0.05	(0.04)	2.8	(2.17)	0.12	(0.09)	1.92	(0.6)	0.92	(0.6)
Ireland	0.83	(1.19)	0.04	(0.05)	0.92	(1.38)	0.04	(0.06)	1.42	(0.62)	0.42	(0.62)
Denmark	0.12	(0.35)	0.01	(0.03)	0.12	(0.35)	0.01	(0.03)	1.06	(0.18)	0.06	(0.18)
Other	0		0		0		0		0		0	
Biodiversity	0.83	(1.17)	0.04	(0.05)	1.67	(2.25)	0.07	(0.08)	1.57	(0.73)	0.57	(0.73)
Waste	1.0	(1.41)	0.04	(0.05)	1.0	(1.41)	0.04	(0.05)	1.49	(0.7)	0.49	(0.7)
Plastics	0.67	(0.58)	0.04	(0.03)	1.0	(1.0)	0.05	(0.05)	1.45	(0.41)	0.45	(0.41)
Renewable Energy	0.71	(0.95)	0.04	(0.05)	1.14	(1.86)	0.06	(0.09)	1.42	(0.62)	0.42	(0.62)
Multiple	0		0		0		0		1		0	
Unrelated testset	0		0		0		0		1		0	
Directive	1.5	(0.58)	0.06	(0.02)	3.5	(1.73)	0.14	(0.07)	2.15	(0.35)	1.15	(0.35)
Delegated Act	0.8	(1.03)	0.04	(0.05)	0.9	(1.29)	0.05	(0.06)	1.41	(0.54)	0.41	(0.54)
Law	0.3	(0.95)	0.01	(0.04)	0.3	(0.95)	0.01	(0.04)	1.15	(0.48)	0.15	(0.48)
Others	0		0		0		0		1		0	

is unsurprising, given the dataset was sourced mostly from a database of documents explicitly reported by countries as influenced by EU directives. The scores across policy areas is very even, except for the two categories showing no exerted influence. It's important to note that none of these figures are significantly different from one another, owing to the small sample sizes and large variances, but the numbers do support both our expectations and intuition from visually inspecting the network.

Keyword Distributions. Finally, we investigate the distribution of keywords in the instruments found across the dataset as a whole compared to the distributions of keywords when weighing documents according to their Katz-centrality. This allows investigation of whether any keywords particularly characterise those documents that exert more influence. For each of the `'instrument_type'`, `'implementation_mechanism'`, `'target_actors'` and `'intervention_objects'` fields found in model-extracted instrument descriptions, we construct frequency distributions by counting the occurrences of unique words. We then construct similar weighted distributions of relative frequencies, by repeating this procedure, except now multiplying word counts for each document by the corresponding document weight. Given that we do not wish to assign weight to policies which exert no influence, we subtract one from the column of Katz Centrality in Table 5 to arrive at our weights. Notably this sets the weight to zero for keywords found in isolated nodes and those without outward connections. The corresponding distributions are illustrated in Figure 7 for the most common words found in the `'target_actors'` field.

The hypothesis that the keyword distributions are unrelated to the document weight

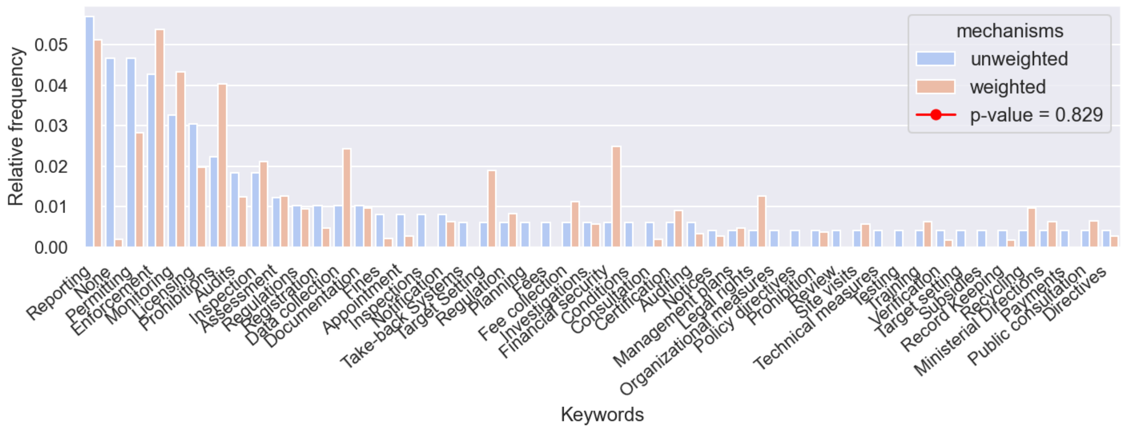


Figure 7: Bar-plot showing relative frequencies of the 50 most common keywords in the `'implementation_mechanism'` field of Directives Test-set instruments (blue). Orange bars show corresponding relative frequencies for those keywords when occurrences are weighted according to node importance. The legend shows the estimated p-value of the null hypothesis that keyword distribution is unrelated to the document weights. In this case, the large p-value means the null hypothesis *cannot* be rejected.

is investigated by 10000-iteration permutation tests. For all four instrument information fields, we find large p-values (0.69-0.80), indicating that the keywords used in influential nodes are not significantly different from those in the dataset more broadly. More importantly, this indicates that the contents of the model-extracted instruments does not very well capture *why* some documents are more similar to others.

Following all this, the reader will hopefully have gained an impression of the some strengths and weaknesses of the data processing pipeline applied to the Directives Test-set. It's worth reiterating that the Directives Test-set is a pretty bad example of a diffusion network, due to the limited amount of policies in each policy field (missing, e.g. transpositions from 25 other EU countries, and many more contextual documents) and the timestamp issues of consolidated policies (i.e. the Irish Birds and Habitats regulations were more likely influenced by the Irish Wildlife act, than the other way around). The node-importance results should, in this case, not be interpreted as diffusion in most cases, but simply as an illustration of how intuitive hunches of node importance can be measured quantitatively, and investigated further. To cover for this weakness of the dataset, we continue by investigating two datasets where peer-reviewed policy diffusion research *has* identified diffusion. By applying the same approach used here to those two datasets, and comparing the results with those of the corresponding studies, we can assess the data-processing approach presented here, and validate against a ground truth.

4.2 Diffusion Study Replication

This section shows the results of processing the *Adaptation Strategies* and *Green Taxonomies* datasets. Results are evaluated by comparison to the findings of the respective studies that inspired each dataset. This will highlight not only whether our methodology can replicate diffusion study results, but also whether the model and prompt used extract the same information that actual researchers have found relevant. Analysis of *Adaptation Strategies* will mainly focus on the first steps of the data pipeline: language model summarisation and instrument extraction, though document similarities will be used to test for divergence between documents. Analysis of the *Green Taxonomies* dataset will mainly be focused on larger-scale network patterns.

4.2.1 Adaptation Strategies

Jensen et al. (2023) study adaptation strategies based on National Adaptation Plans (NAPs) and Strategies (NAS) from Denmark and the United Kingdom (UK). The study analyses country planning individually before comparing plans and drawing conclusions regarding diffusion. We will follow the same formula, by reviewing model-generated outputs for the four documents the study considers in detail.[22]

Denmark Climate Change Adaptation Strategies and Plans. According to Jensen et al. (2023), Denmark's 2008 'Strategy for Adapting to Climate Changes in Denmark' has a broad scope and covers many sectors, and the plan "emphasizes risk assessment and early action, though without committing to concrete actions."

By comparison, the 2012 NAP 'How We Manage Cloudbursts and Rains' narrows down targetted sectors and focuses on water management. Specifically, Jensen et al. (2023) list the targetted sectors in the 2012 plan as 'Water', 'Coastal protection', 'Transport' and 'Agriculture'. The plan considers climate adaptation as a policy issue across administrative levels, but is heavily geared towards local actions, mandating municipal climate adaptation plans and implementing national mechanisms to support this. Jensen et al. (2023) list the instruments applied in the 2012 plan as 'Mandatory local adaptation plans', 'Knowledge platform', 'Task force', 'Risk assessment', 'Sectoral plans', 'Coordinating networks' and 'Cross-ministerial committee' (but do not produce a similar list for the 2008 strategy).

Table 7 shows model-generated objectives, summaries, and domain categorisation, where several words have been bolded (our emphasis). The model picks up on the broad focus of the 2008 strategy and the comparatively narrower focus of the later NAP, as evident by the given domain, the water-related keywords in the 2012 summary,

[22]The study also included a fifth document, the '2021 Danish NAP', which was in preparation at the time. It seemingly still is, and we were not able to find a suitable PDF to include in our dataset.

Table 7: Side-by-side comparison of the objectives and summaries generated for the Danish Adaptation Strategies. Some words are bolded (author emphasis).

2008 National Adaptation Strategy	***2012 National Adaptation Plan***
DOMAIN: General OBJECTIVES: The strategy focuses on national adaptation to climate change, integrating climate change considerations into planning and development to enable authorities, businesses, and citizens to respond **proactively** to the challenges posed by climate change. SUMMARY: The strategy outlines the need for national adaptation to climate change, emphasizing the integration of climate considerations into planning and development **across various sectors**. It highlights the importance of ongoing adjustments by authorities, businesses, and citizens within existing legal and economic frameworks. The strategy identifies key **sectors** vulnerable to climate change, including **coastal management, construction, water supply, energy, agriculture, fisheries, nature conservation, and health**. It proposes initiatives such as establishing a climate adaptation portal, promoting research, and creating a coordination forum to ensure a coordinated response to climate change challenges. The strategy also addresses the role of the rescue services and insurance aspects in managing climate-related risks.	DOMAIN: Water OBJECTIVES: The action plan aims to ensure that the country is robust against climate change by improving the framework for climate adaptation, creating a new **knowledge base**, **strengthening cooperation and coordination**, promoting green conversion, and addressing **international climate adaptation**. SUMMARY:The document outlines a national action plan for climate adaptation, focusing on managing **cloudbursts and rainwater**. It emphasizes the **shared responsibility of government, municipalities, businesses, and citizens** in addressing climate challenges. The plan aims to improve the framework for climate adaptation through updated legislation and regulation, establish a common **knowledge base and advisory services**, and strengthen **coordination across sectors**. It also seeks to promote green conversion by developing innovative solutions and addressing international climate adaptation efforts. Key areas of focus include **managing increased rainfall**, incorporating climate adaptation into **planning and construction**, and promoting collaboration among stakeholders. The plan includes initiatives such as modernizing watercourse and water supply laws, **risk mapping**, and establishing a **task force** for climate adaptation.

and the long list of targetted sectors in the 2008 summary which is missing from the 2012 one. However, the 2012 description does not emphasize municipal actors to the expected degree, makes no mention of Transport or Agriculture, and furthermore both plans' summaries include several points that Jensen et al. (2023) do not emphasize.

Some of the expected emphasis can be recovered by inspecting extracted instruments for each document. Municipalities are by far the most frequently targetted actor among the 2012 instruments, much more so than in the 2008 strategy. On the other hand, emphasis on Transport and Agriculture cannot be found in the 2012 instruments either, albeit several other instruments that Jensen et al. (2023) highlight can be found here or in the summary, such as the municipal adaptation plans, knowledge sharing platform and task force, risk assessments, and wastewater company project co-financing. While Jensen et al. note a lack of concrete actions in the 2008 strategy, the model generates almost twice as many policy instruments here as compared to the 2012 plan, though this is in line with the pattern of Figure 3 as the earlier document is longer.[23]

[23]As illustrated by Table 2 however, the number of instruments can be misleading.

Table 8: Side-by-side comparison of the objectives and summaries generated for the UK 2013 and 2018 Strategies. Some words are bolded (author emphasis).

2013 National Adaptation Programme	***2018 National Adaptation Programme***
DOMAIN: Environment OBJECTIVES: The National Adaptation Programme aims to build the UK's resilience to climate change by **addressing risks** and opportunities across various sectors, including the **built environment, infrastructure, health, agriculture, and business**. It seeks to increase **awareness**, enhance resilience to extreme weather, take timely action for long-lead time measures, and address major evidence gaps. SUMMARY: The National Adaptation Programme outlines a comprehensive strategy to enhance the resilience of various sectors to climate change impacts. It emphasizes **collaboration between government, industry, and local communities** to address risks and capitalize on opportunities. Key areas of focus include flood and coastal erosion risk management, spatial planning, **infrastructure resilience**, and community preparedness. The program promotes sustainable development, water efficiency, and the integration of climate considerations into regulatory frameworks and business practices. It also aims to protect vulnerable groups, enhance ecosystem resilience, and improve the evidence base for informed decision-making. The strategy addresses international threats and opportunities, emphasizing the importance of global stability and sustainable development. The programme includes measures to improve water management in **agriculture**, enhance **forestry** resilience, and manage **pests and diseases**. It also promotes innovation and the development of climate-resilient technologies.	DOMAIN: General OBJECTIVES: The document outlines the government's strategy for adapting to the impacts of climate change. It assesses climate change risks and sets out actions to **address** these **risks**, improve resilience, and promote **collaboration across various sectors**. SUMMARY: This document presents the second National Adaptation Programme, detailing governmental actions to address climate change risks and opportunities as identified in the second Climate Change Risk Assessment. It emphasizes a flexible, iterative approach, **building upon previous adaptation efforts**. The program focuses on priority areas including managing risks to **communities, businesses, and infrastructure** from flooding and coastal changes; addressing water supply shortages; **protecting natural capital**; and ensuring food production and trade. It promotes ecological resilience, sustainable resource management, and collaboration across sectors. The document also outlines plans for the third cycle of **adaptation reporting**, encouraging voluntary participation from public bodies and infrastructure operators to enhance resilience and **inform future policy development**.

United Kingdom Climate Adaptation Programmes. Jensen et al. (2023) note that the UK Climate Adaptation Programmes target similar sectors ('Health', 'Agriculture and Forestry', 'Built environment', 'Infrastructure', 'Natural environment', 'Business', 'Local government') and actors ("National, regional and local administrators, private sector and 3rd Sector"). These sectors are very well reflected in the 2013 model-generated descriptions, but not emphasized in the 2018 description (displayed in Table 8). One hypothetical explanation is that the 2018 programme serves as an update of the 2013 programme, and that the document thereby more heavily emphasises *changes* from the first, rather than overlaps. This hypothesis is somewhat supported by the 2018 description that the programme "builds upon previous adaptation efforts". Jensen et al. (2023)'s descriptions of the programmes differ mainly by the wider range of instruments employed in the later programme, and they further mention the 2018 plan strongly emphasizing natural capital for adaptation. This is mentioned in the description and in

Table 9: The amount of instruments classified by type, for each strategy.

	DK 2008	DK 2012	UK 2013	UK 2018	Total
Economic	1	3	2	1	**7**
Regulatory	5	2	11	8	**26**
Soft	5	1	1	1	**8**
Total	**11**	**6**	**14**	**10**	

several of the found instruments (relating to woodland planting, peatland restoration and fisheries), but as in the Danish case the model finds significantly more instruments in the first plan, though once again this can also be explained by the lengths of the plans and does not necessarily reflect substance very well.

Cross-country comparison. Jensen et al. (2023) highlight the economic rationales for adaptation policy adoption present in all four climate plans, such as co-benfits, cost-avoidance and risk-reductions. They find however, that neither country relies strongly on economic incentives in their strategies, though noting that the UK does so to a higher degree. Likewise, Jensen et al. (2023) find that "hierarchical instruments", i.e. rules, orders and stringent regulation, only plays a very limited role in the strategies, noting though the Danish mandate on municipal planning. Instead, they find that knowledge-building and -sharing as well as procedural instruments to support collaboration have been the primary instruments for both countries' strategies, and that the strategies between the two countries are very similar in this dimension.

The topic of these observations align nicely with categories of the 'type' field in each instrument, where the 'soft' category instructions corresponding to what Jensen et al. (2023) term "knowledge-building" and "procedural" instruments, the 'regulatory' corresponding to "hierarchical instruments", and likewise the 'economic' category obviously fitting well. We can thereby investigate if the language model outputs agree with Jensen et al. (2023)'s findings.

Table 9 shows the amount of model-extracted instruments classified in each of the three 'type' options, for each document. Overall, the numbers presented there do not agree very well with the findings of Jensen et al. (2023). The distribution between types is not at all similar for the two countries, nor between the Danish documents. Like for the Directives Test-set, we observe the same heavy emphasis on regulatory instruments, which explains much of the discrepancy. Worth noting however, is that the distribution here leans *less* towards regulatory instruments than for the previous dataset. Assuming that the model does show some bias towards the regulatory category, it would require relatively more emphasis to show a strong response in a non-regulatory category. Thereby the large amounts of soft instruments in the Danish

2008 strategy, and the majority of economic instruments in the 2012 plan, indicate *very* heavy emphasis on these types of instruments in the respective documents. For the 2008 strategy, this aligns with Jensen et al. (2023)'s findings, but less so for the 2012 plan. In sum, the model-found instruments have once again not particularly impressed, and the findings here support the impression left by Directive-testing that the instrument-finding is somewhat spurious.

Policy divergence. The final few findings of Jensen et al. (2023)'s that we will attempt to replicate concern the evolution, convergence and divergence of the investigated plans over time. Jensen et al. (2023) suggest that policy diffusion played a factor in the evolution of the adaptation strategies, including the design of the EU's Adaptation Strategy in both 2013 and 2021. Furthermore, they find that divergences began to appear between Danish and UK strategies, especially in leading up to and following 'Brexit'.

Convergences and divergences should be reflected in the similarities between documents, as diverging policy designs between countries would lead to falling pairwise similarity scores between their documents over time. To investigate, we show in Figure 8 the similarity matrix for embeddings of the investigated strategies. It includes several other documents to those manually investigated, including the most recent UK programme and several related EU documents, including the 2013 and 2021 EU Climate Adaptation Strategies and two documents from 2007 and 2009 respectively, which kicked off EU policy development regarding adaptation (Jensen et al., 2023, p.5). Alongside is drawn the resulting network, when connecting documents to the most similar previous document (with no minimum similarity threshold).

The matrix shows that the Danish 2008 strategy has high similarity to almost all other documents, including the 2009 and 2013 EU documents (which could indicate them drawing inspiration from the Danish strategy) and conforms well to the finding of Jensen et al. (2023) that at the time "Denmark saw itself more as a climate adaptation leader than a learner". The central role of the corresponding network node certainly supports this Danish self-image. Given that this dataset is just a small subset of all climate adaptation policies in Europe however, the dataset does not allow broad conclusions regarding whether Denmark indeed was a climate adaptation leader at the time. The centrality of the Danish 2008 node could just as well be a result of it simply being amongst the oldest nodes, like the spurious connections in the Figure 6.

The 2008 Danish strategy also shows high similarity to both older UK programmes, however, implying no significant divergence. The Danish 2012 plan has comparatively lower scores, albeit notably not to the UK's 2018 plan. This runs counter to the findings of Jensen et al. (2023), as we'd expect downwards-trending similarities scores over time if policies were diverging. The UK's most recent programme (not studied by

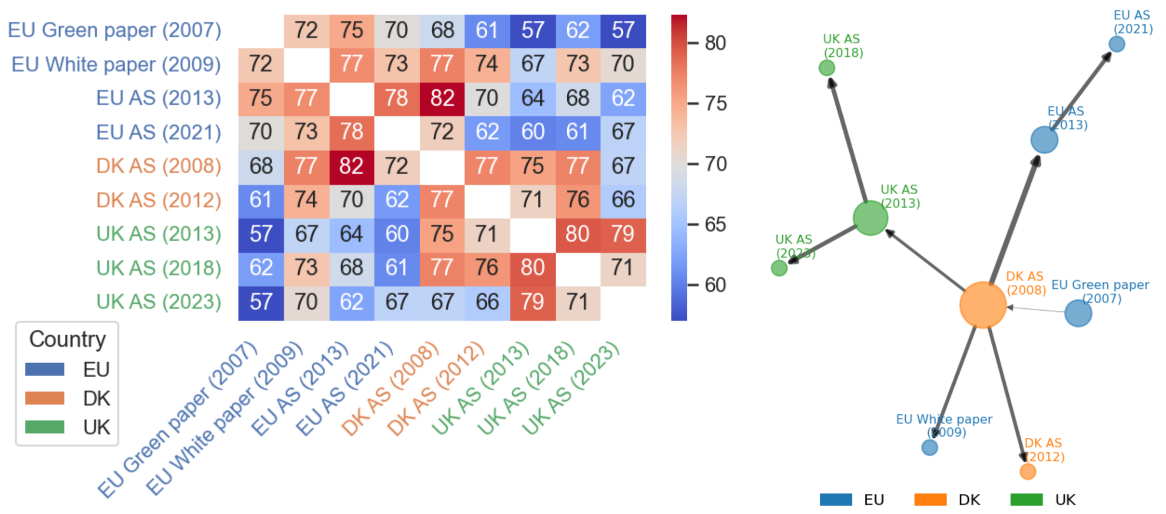


Figure 8: Similarity matrix for the Adaptation Strategies (AS) dataset and the corresponding network with connections drawn to most similar previous nodes (no similarity restriction).

Jensen et al. (2023)) does show divergence from the previous two programmes as it has lower similarity scores with all documents, except, crucially, the EU's 2021 strategy. In this light it is hard to interpret the falling scores as the UK diverging from the EU Adaptation Policy Tradition. That the EU 2021 strategy shows a similar pattern to the UK 2023 programme rather implies that both have evolved in a new direction from the rest of the strategies and plans. Notably, if we (like Jensen et al. (2023), who did not have the opportunity to include it) disregard the UK 2023 programme, the EU 2013 → 2021 evolution is reminiscent of divergence from the rest of the dataset documents, whereby one might be left with the impression of diverging policy traditions.

In conclusion, the LLM-based investigation of the Adaptation Strategies dataset cannot be said to replicate the findings of Jensen et al. (2023) entirely. While some of the broader patterns, especially based on document objectives and summaries, can seemingly be established, the finer detail is often lacking. Notably, the model-based policy instrument investigation has once again shown very mixed results. Still, it would be surprising if the data processing here performed comparably to human expert evaluation on a small dataset, especially given the lack of specific and directed prompting. The strengths of the methodology lie in its potential to treat larger datasets with something approaching the analytical skill of a trained expert. To evaluate the data process in such a setting, we turn to a dataset of larger-scale patterns.

4.2.2 Green Taxonomies

Green Taxonomies are a relatively recent policy innovation, which have become increasingly popular with the recent rise of green financial policies. Taxonomies define what are to be considered 'green', 'sustainable' or 'environmentally friendly' activities, and are thereby used to guide financial flows toward sustainable investment. Examples of applications include defining the scope of recipients for 'green' subsidies, or monitoring (un)sustainable international trade (Larsen, 2022). China kicked off the development in 2015, but since then 25 countries or organisations have joined the trend and developed their own taxonomies, including the EU. The Green Taxonomies dataset consists of 41 documents, each corresponding to a 'green' or 'sustainable' taxonomy of financial activities adopted or drafted by one of these 26 actors.[24] Of these, 1 was excluded from analysis as it failed to process.

Larsen (2022) studies some these taxonomies and notably finds that several of the taxonomies draw inspiration from the structure, principles and metrics of the EU Green Taxonomy. These include Bangladesh, Russia, South Africa, and China's updated 2021 taxonomy, where Larsen (2022) notes the influence the EU has had on the 'green' status of coal. He also notes that Mongolia as the sole nation seems mostly inspired by the original Chinese taxonomy. Others that have studied the issue have found that taxonomies under development at the time (now final and included in the dataset) seemed to either use the EU taxonomy as inspiration for drafting or as a benchmark for their own metrics and technical screening criteria for sustainable activities (Gondjian & Merle, 2021). In the following, we construct the similarity matrix for the dataset and investigate the resulting network structures, to see if they agree with these findings.

Similarity and Network Analysis. Figure 9 shows the similarity matrix for the Green Taxonomies dataset. The fairly obvious pattern is that two groups of documents are clear outliers and very different from both each other, and a larger 'main group' of documents. The pattern is reminiscent of the one made by the unrelated category of the Directives Test-set. This is a small surprise, given that we expected all these Taxonomies to have some common inspiration. Somewhat comforting is the fairly high baseline/average similarity compared to the Directives Test-set (Figure 4). This reflects the common topic among all documents in the dataset, and even the lowest-scoring pair similarities among these documents are significantly higher than for e.g. the unrelated group in the Directives Test-set. What is much more surprising is that the two groups of outlier documents are exactly the the EU and Chinese Taxonomies, respectively, and their related documents.

As the datasets size is clearly limiting the legibility of the matrix at this point, however, we instead draw a network map to investigate document relationships closer. Seeing as we (at least originally) had reason to believe these policies were related, we do not impose a similarity threshold and thereby create a single large network, displayed

[24] In some cases, a document was merged from several different sources, such as chapters or annexes. See the comments column in Appendix F.

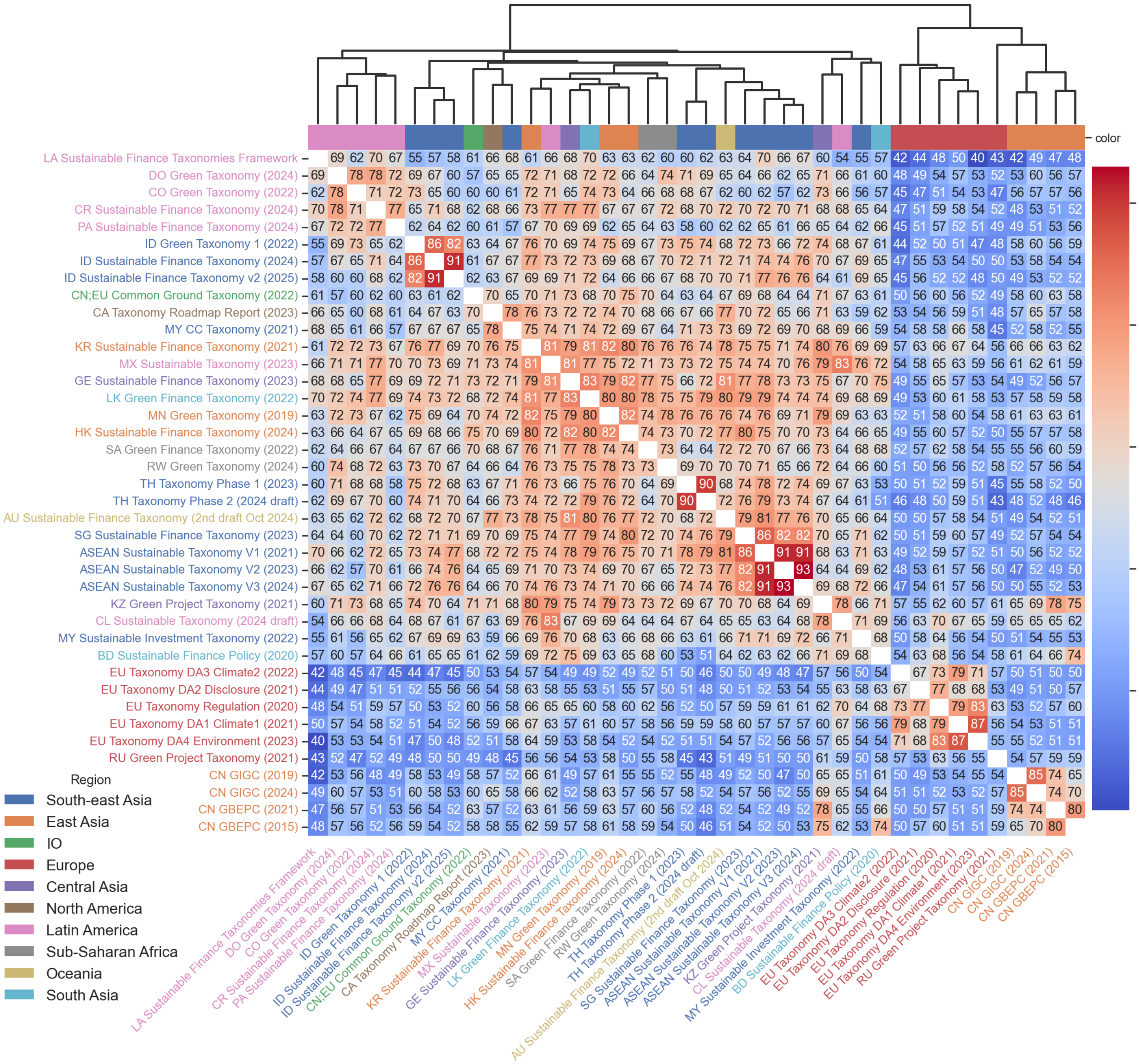


Figure 9: Green Taxonomies similarity matrix. The rows and columns have been sorted according to the clustering of the shown dendogram.

in Figure 10.[25]

The network displays the Mongolian 2019 taxonomy in a central role. This is quite surprising, given (Larsen, 2022)'s description of the Mongolian taxonomy as an outlier and the only taxonomy studied to be significantly inspired by China's original structure, rather than the EU's. Our results show just the opposite, that the EU and Chinese taxonomies are outliers, while the Mongolian taxonomy acts as the central

[25]For reference, the lowest pairwise similarity scores that lead to a connection in the network are CN (2015) → EU (2020), CN (2015) → MN (2019), EU (2020) → RU (2022), all with similarities just over 0.6. After that, the next connection doesn't break until the similarity threshold rises above 0.7

Figure 10: Network for the Green Taxonomies dataset, with no minimum similarity threshold. The words 'Taxonomy' and 'Finance' have been omitted from labels for brevity.

hub of the graph. While the Mongolian taxonomy technically *is* connected to the original Chinese taxonomy, this is a very weak connection that would be broken if any significant similarity threshold was imposed (as evident by the very thin arrow connecting them). Thereby, it doesn't support the fairly strong bond that Larsen suggests. As we discuss later, this surprising pattern may be related to the format of taxonomies and the documents presenting them.

Leaving aside the Mongolian node for now, the rest of the network contains several other interesting patterns. We see a large clump of South-east Asian nodes on the left of the graph, seemingly with point of departure in the ASEAN taxonomy, which seems very reasonable and in line with descriptions given by (Gondjian & Merle, 2021). Likewise, just north of the central Mongolian node, we see a few Latin American countries connected.

We may also note that the Common Framework of Sustainable Finance Taxonomies for Latin America and the Caribbean, a guidance document developed by UNEP as a reference for countries in the region drafting green taxonomies, seems inspired by the Sustainable Taxonomy developed by ASEAN. This is a reasonable result, suggesting that UNEP looked to ASEAN for inspiration in how to classify activities over a larger economic region. Indeed, inspecting the actual UNEP document reveals several explicit references to the ASEAN approach (UNEP, 2023, e.g. p. 22-23). Furthermore, the document also highlights Colombia as a source of inspiration for other LA countries, supporting the validity of the previously noted Latin American clump (UNEP, 2023).

Unlike other South-east Asian taxonomies, Malaysia's (2021) Climate Change taxonomy is not part of any larger clump, but this corresponds well to descriptions elsewhere that this is one of the sole taxonomies focused exclusively on climate change, whereby it *is* quite dissimilar to most other approaches and unlikely to have inspired many. (Gondjian & Merle, 2021). This finally brings us to some of the less-expected connections, such as Georgia → Australia/Costa Rica/Hong Kong, Korea → South Africa/Mexico, and of course Malaysia (2021) → Canada.

All in all, the network analysis of the Green Taxonomies yields another hit-and-miss performance from the diffusion-finding procedure presented here, with both encouraging connections and surprises - the largest of which is obviously the sidelining of the EU and Chinese Taxonomies. Following both this and the similar performance in the Adaptation Strategies dataset, as well as the relatively unconvincing display of instrument-extraction, we might have called it a day and concluded the procedure is simply not mature or trustworthy at this stage. Not to be discouraged by mediocrity and less-than-ideal results however, we move on to studying the FAOLEX dataset - if not for the significance of results, then to illustrate how a future, improved, diffusion-finding procedure *could* investigate the questions asked at the beginning of this thesis and inform further analysis.

4.3 FAOLEX

After cleaning and exclusions based on technical difficulties, the final FAOLEX dataset consists of 1779 documents from 1642 unique database entries.[26] Documents originate in 48 unique countries, and include documents from country subdivisions (e.g. regional laws), as well as documents listed with multiple authoring countries. For our purposes, the dataset contains five other relevant features, as it includes the 'type', 'primary subject(s)', 'domain(s)', and 'keywords' associated to each document, as well as document language. Each of these information fields list one or more labels given to the dataset from a limited set.[27] 'Primary subject' has 24 categories, 'domain' has 16, and 'keywords' over 400 (where documents were selected for inclusion in the dataset on the basis of including 'climate change' as one of these keywords). Subjects and domain labels have significant overlap, as 13 of the 16 options in 'domain' are also

[26] The 130 extra documents are mainly second-language or consolidated versions, but not exclusively.

[27] Fields with multiple labels aren't (always) sorted alphabetically, which implies some ordering, but the documentation available does not comment on this (FAO, n.d.).

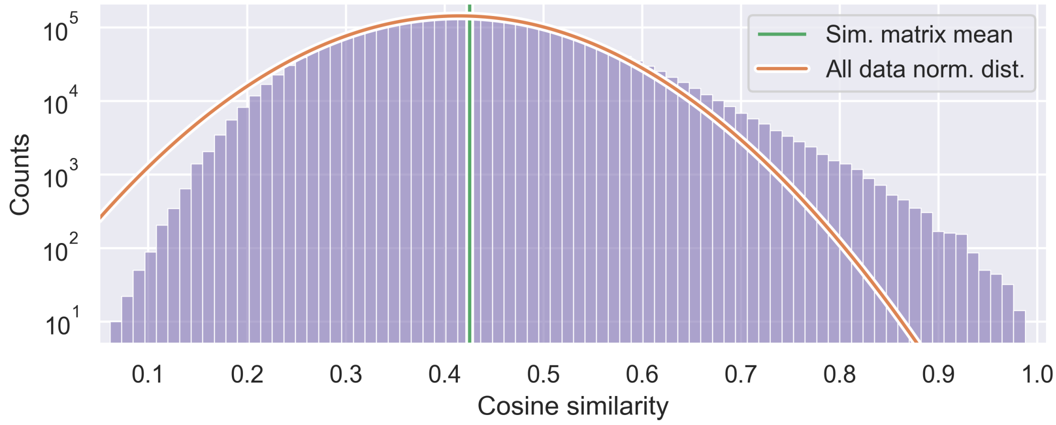


Figure 11: The histogram of pairwise similarity scores in the similarity matrix of the FAOLEX dataset, plotted alongside the best fitting Normal Distribution. Note that the y-axis has been scaled logarithmically to allow inspection of the tails.

found in ‘primary subject’. The domain labels can be seen in the legend of Figure 12. Each document lists on average 1.2 ‘primary subjects’, 2.9 ‘domains’ and 16.5 unique ‘keywords’. Type is either ’legislation’, ’regulation’, ’policy’, or in rare cases, combinations hereof. Given the size of the dataset, manual analysis of documents and corresponding model outputs is not feasible, nor is visualising using a similarity matrix. Instead we proceed directly to analysing the distribution of the similarity matrix, the large-scale diffusion patterns and the keyword distributions.

Similarity Analysis. Figure 11 shows the histogram of pairwise document similarity values for the FAOLEX dataset, once again alongside the best fitting normal distribution. The number of pairwise similarity scores grows quadratically with the size of the dataset, so the large dataset results in many (3.2 million) scores and a very smooth distribution. Furthermore, we see that the range of similarity scores is wider than in any of the previous datasets, indicating a larger variance among documents. While smooth however, the distribution is not normally distributed, but once again skewed towards higher similarity values, and away from lower values. Some of the very highest similarity values (approaching 1) can be explained by the pairwise comparison of duplicated documents (remaining below a similarity score of exactly one due to e.g. slight dissimilarities between different-language versions). However, given that these documents only number ~ 130, and should only show a particularly strong signal for one or two others (i.e. their other copies), they cannot explain the full tail, even if their exclusion does trim the very highest values. Rather, we once again resort to diffusion as the primary explanatory factor.

The relative smoothness of the curve means it shows no obvious similarity threshold. Instead, we choose a threshold based on how large a proportion of similarity scores above that threshold that are 'abnormal'. This proportion is calculated as the sum of the bin counts above that threshold, divided by the corresponding area under the normal curve.[28] If this ratio is high for a certain similarity threshold, it implies many more document pairs above that threshold than expected if scores were normally distributed. Assuming these overrepresented pairs have high similarity scores *because* of policy diffusion, a high ratio is an indicator of diffusion pressure in network resulting from that threshold.

Exactly which ratio to aim for depends on the scope of one's investigation. Too low of a threshold will pollute the sample with false-positive diffusion cases: connections drawn between documents that just 'happen to be' very similar to others. An example of this are the spurious connections discussed in relation to Figure 6. A higher ratio will imply a stronger indicator of diffusion, with fewer connections representing false positives. However, a high ratio also implies moving further right on the distribution curve, and thereby dramatically reducing the number of possible connections by virtue of the similarity threshold growing. This might lead excluding actual cases of diffusion due to a too strict similarity threshold.

In the end, the choice of threshold is a rather arbitrary one, depending on preference and tolerance for error. The previously displayed mixed performance of the present methodology might inspire a relatively high similarity threshold, to be relatively more certain that the results actually represent diffusion. In Figure 12, we've instead prioritised a useful visual representation and therefore networks shown have been drawn using a generous threshold of only sim > 0.62. This corresponds to a ratio of areas $A_{\mathrm{bin}}/A_{\mathrm{norm}} = 1.85$, indicating that above that threshold there are 1.85 document pairs with a sufficient similarity score to form a connection *for every one pair we'd expect in the normal case*. This is not a very restrictive threshold, with not even the majority of connections being 'unexpected'. However, networks resulting from higher thresholds very quickly become too fragmented to easily comment on the visuals of. Instead, more restricted networks are shown in Appendix E for threshold values of $0.65, 0.70, 0.75, 0.80$ and 0.85, representing area ratios of $2.5, 4.2, 9.5, 25$ and 65, respectively. These five networks will form the basis for the results that will be reviewed in the following sections, with plots shown for the $\mathit{Similarity} > 0.75$ case.

The network in Figure 12 is clearly already quite fragmented, and we'll not dwell on it too long, except to note that once again, we see splitting into sub-graphs of subject-related nodes. For example, near the top right corner is a group of documents dealing with *Forestry*, *Agriculture & Rural Dev.*, *Land & Soil* and *Livestock*. Just south-west of that, a cluster dominated by *Air and Atmosphere*, and near the bottom,

[28] We actually compute based on a Kernel Density Estimate (KDE) of the histogram distribution. We define the KDE and the Normal distribution on the same range of similarity values and compute the cumulative remaining distribution by summation.

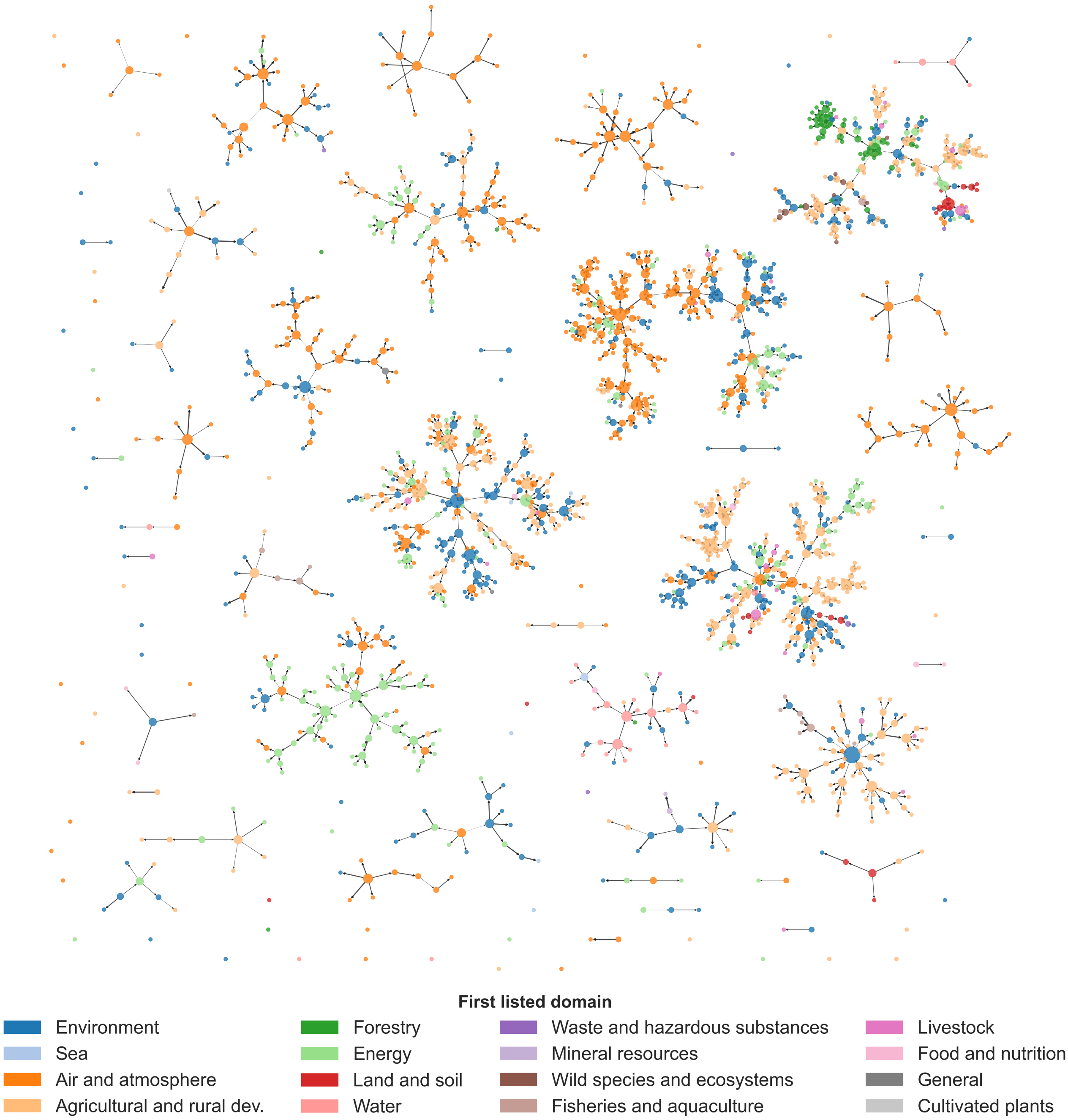


Figure 12: Network representation of the FAOLEX dataset. For each node, a connection was drawn to the most similar node from a previous year, if the document pairwise similarity is above 0.62. Nodes have been coloured by the first (of possibly several) domains listed in the corresponding FAO database entry.

clusters 'focused' on *Energy* and *Water* respectively. As the similarity threshold is increased, the weakest connections are 'cut' and sub-graphs gradually break into smaller pieces as central connections fall below the threshold - eventually leading to many isolated nodes or 2-node subgraphs at very high thresholds. For example, the *Forestry*, *Agriculture & Rural Dev.*, *Land & Soil* and *Livestock* sub-graph breaks down into smaller pieces more-or-less corresponding to each of those domains individually.

Table 10: Statistics for the *Similarity* > 0.75 Network. Node importance has been aggregated by mean and maximum for select countries in the FAOLEX database. The table is sorted by the mean values.

	Mean Katz centrality	Mean rank	Max Katz centrality	Max rank	Number of documents
Slovakia	2.34 (2.66)	1	10.19	3	11
Malta	1.92 (1.36)	3	6.16	18	24
Czechia	1.92 (2.03)	4	8.83	9	18
Germany	1.68 (1.71)	7	10.11	4	86
Italy	1.62 (1.39)	9	8.66	11	76
France	1.62 (1.28)	10	8.98	7	78
Finland	1.62 (1.36)	11	8.30	12	51
Denmark	1.61 (1.72)	13	10.32	2	45
Norway	1.55 (1.26)	15	7.00	16	36
European Union	1.48 (1.07)	20	7.40	14	176
UK	1.46 (0.96)	21	8.75	10	180
Sweden	1.26 (0.6)	37	3.31	34	27
Netherlands	1.22 (0.54)	39	3.14	35	19
Russian Federation	1.20 (0.36)	40	2.32	41	37

Network Statistics The observation that nodes connect to nodes with similar subject area is corroborated by network statistics, where the assortativity by first listed domain is generally in the range $0.4 - 0.6$, growing as the network becomes sparser at higher similarity thresholds. Likewise, the network shows strong, growing assortativity on document type (legislation, regulation, policy), in the range $0.7 - 0.85$.

Table 10 shows network statistics aggregated for a few select countries, based on both the average importance of nodes (mean) and the node with the highest importance (max). The table also shows the number of nodes for each country, and how the statistics rank among all other countries. From the table, it is apparent that e.g. Denmark is neither badly nor particularly well-placed, and does not seem to exert any particularly high influence on average. As evident by the large standard deviations and as we shall see in the following paragraph, the country differences here are *not* significant. Worth noting however is, that Denmark in this network has the second most important node, and the corresponding document is the 2008 National Adaptation Strategy studied earlier (the English language version, specifically). The ranking of nodes is volatile as the similarity threshold varies, but it is nonetheless encouraging that we have seemingly corroborated the results of section 4.2.1, and in a quite unexpected way at that! Note also the large variance in how well represented countries are in the database.

To wrap up, we move on to the distributions of keywords among influential nodes.

Keyword distributions. Here we repeat the procedure from section 4.1.3 for investigating keywords in the instrument fields of the model outputs. Outside of just the data fields from the model-extracted instruments however, we also analyse the FAOLEX database information fields on 'type', 'primary subject(s)', 'domain(s)', and 'keywords', as well as adopting country and document language. As with instrument keywords, we can study how the labels in each of these database features distributes among influential nodes, and if that distribution differs significantly from the dataset in general. This should be a more robust test than for instrument keywords, as the database information fields have been externally validated.

As in section 4.1.3, we find no significant differences for how the keywords found in model-extracted instruments are distributed. For the 'Subject' and 'Keywords' database features however, we do find significant results at the 5% level after 10000 permutations across all five tested networks..[29] Figure 13 shows theese distributions, and given the motivation of our study, we also show the in-all-cases-significant distribution of authoring countries.

A few things are worth noting from these distributions. Regarding countries, Germany, France, Italy, Malta, and to a lesser degree Denmark, are relatively more frequent as authors among influential documents. Conversely, Estonia, Russia, Serbia, Sweden and the Netherlands, are less frequent, while the EU and UK are relatively unchanged. This of course corresponds to the rankings in Table 10. However, we stress that these results are *not* significant.

Regarding subject, the 'Environment' tag (and perhaps 'Trade') stands out as significantly more frequent among influential nodes, but other major differences are hard to spot.

Finally, regarding keywords, the 'climate change' keyword is relatively more frequent across important nodes. This may seem confusing, given that all documents are tagged with the 'climate change' keyword as their basis for inclusion in the dataset. In this case, the only way the 'climate change' keyword can become *more* frequent is by other keywords being used *less*. In other words, the important nodes use fewer keywords, and are more focused on climate change. This is supported by the fact that it is not only 'climate change' which shows a stronger signal among the important nodes, but also air- and atmosphere related tags such as 'emissions', 'air quality/air pollution' and 'ozone layer', which might also be relevant for policies dealing with climate change mitigation. Notably, categories such as 'Innovation', 'Business/Industry/Corporations' and 'Circular Economy' show a reduced use among influential nodes.

Thus concludes the analysis. The following section discusses the results, as well as the methodology used to reach them.

[29] The keywords distribution divergence was in fact not *quite* significant at for the sim = 0.70 network, with $p = 0.0508$. For all other thresholds, $p_{\text{keywords}} < 0.025$, why it has been included.

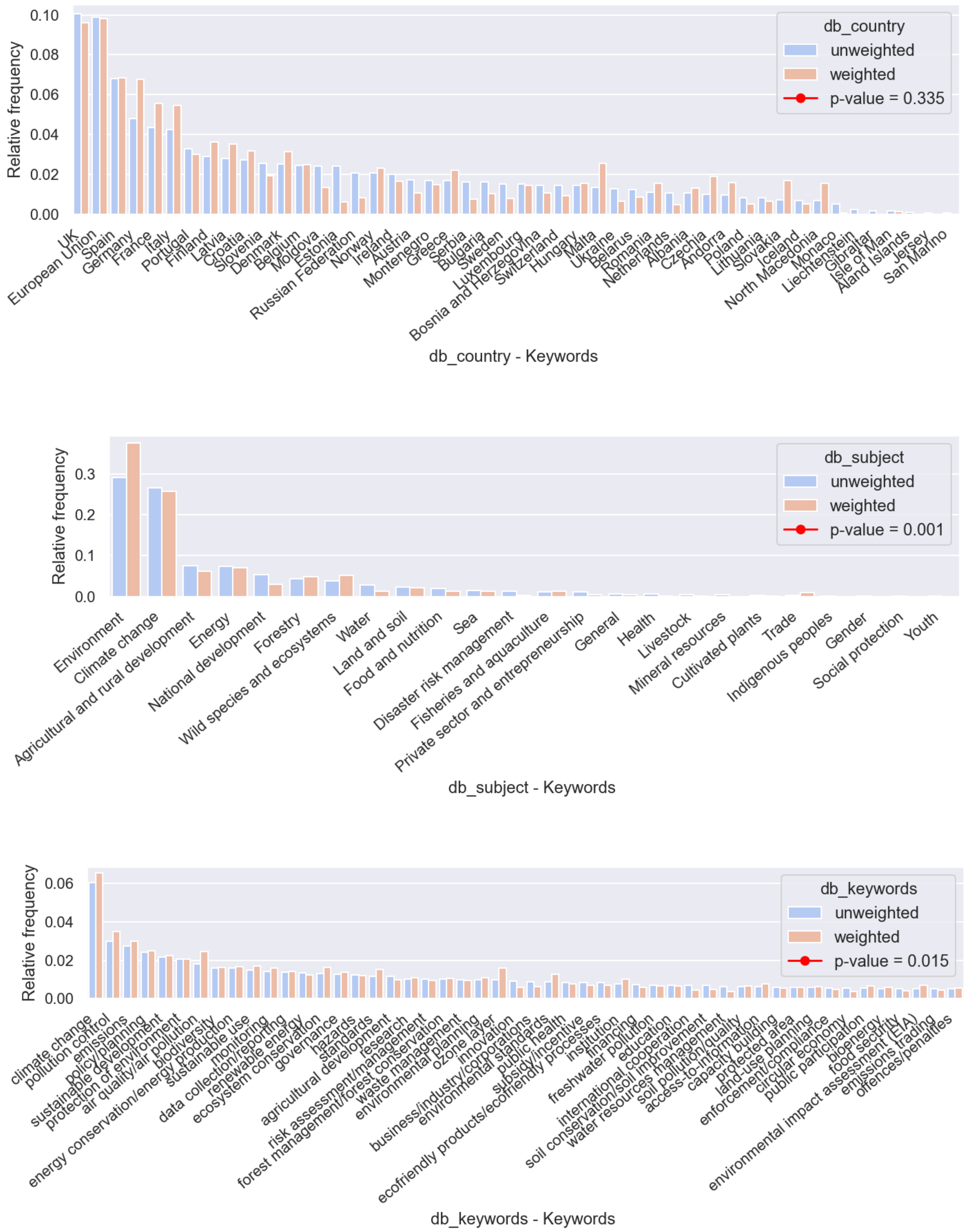


Figure 13: Relative frequency distributions for authoring countries (**top**), subject (**middle**) and keywords (**bottom**). The latter two database information fields have weighted distributions significantly different from the unweighted distribution at the 5% level. Distributions shown here were generated based on the network with similarity threshold *Similarity* > 0.75. The plots show the up to 50 most frequent labels in the unweighted distribution.

5 Discussion

The analysis has shown that the language model performs admirably when in recovering common thematics and substantive relations between documents. Throughout the analysis, the comparisons of embedded descriptions time and again recreated patterns, grouping and clusters found or expected in the original databases. Furthermore, there are strong indications that those found patterns, when combined with network analysis, *can* be used to trace diffusion patterns. With that said, there are several points worth discussing in relation to both the performance, design and future application of the methodology presented here.

Basic diffusion assumptions. Fundamentally, any conclusions we draw regarding policy diffusion rely on an *assumption* that document similarity is indicative of diffusion. This is a common assumption in the literature on quantitative investigations of policy diffusion, as we have noted in our background review, and one with intuitive appeal. It is however also an assumption that empirical evidence challenges or shows violations of. As an example, according to the findings of the 'policy mobilities' research field, policies evolve and mutate as they diffuse, implying that an increase in pairwise similarity could be small or entirely absent despite diffusion. A particular case is countries experimenting *un*successfully with policy, whereby they may act as inspiration for others on how *not* to approach a policy objective.

Still, even if policy diffusion is not *guaranteed* to cause similarity, we have attempted to corroborate that similarity *can* be a consequence of diffusion. We have done so by investigating the distributions of similarity scores among documents, and argued that the basic assumption can be strengthened by limiting the search for diffusion to the very highest pairwise similarity scores, which fall outside of a normal distribution. Here we further rely on 1) an expectation for similarity scores to be normally distributed and 2) skewing of the distribution being caused by diffusion.

That similarity scores are normally distributed is something that is at the very least roughly true for all the datasets presented here, but furthermore the natural consequence of random, unrelated variations around some baseline. It seems highly intuitive that policies should have some baseline similarity - being more alike than different, so to speak - regardless of subject area, and that differences in approach between unrelated policies would look like random variations around that baseline. Still, normality is an assumption that warrants confirmation. Ideally, we could confirm the normality by inspecting policy datasets of *unrelated* documents - though it is hard to imagine a large set of policy documents where no policy diffusion is present. Still, we could likely limit the influence of diffusion, e.g. by investigating the distribution of similarity scores only for document pairs from vastly different geographic origins, subject areas, and time periods.

The assumption that skewing can be attributed solely to diffusion seems more problematic. Fundamentally, skewing of a normal distribution implies some effect, acting on only a subset of data points, which systematically raises (or lowers) their pairwise scores. We can imagine several effects not representing policy diffusion that would amount to as much, most notably policy response in face of common challenges. Countries facing common challenges and independently reaching similar policy outcomes is something we cannot very well account for in the present investigation. Since challenges faced often form the motivation for action, we might imagine common challenges to show up primarily as an overlap in our extracted *motivations* for policies, and less so in the (summary of) substance. This is a testable hypothesis under the experimental design we have established here, albeit we have not pursued this route of enquiry. In particular, a dataset such as FAOLEX, which includes policies acting on the common issue of climate change, might provide opportunities for testing this. Even in these cases however, what is the result from independent *necessary* action against common challenges can be difficult to disentangle from *inspiration* to tackle a problem. Diffusion might not take the form of a direct inspiration from the legal or political formulations of others, but simply of inspiration to do *something* with common objectives. This is not in particular the sort of diffusion we set out to measure, but something that we likely capture to some degree nonetheless.

As touched upon during the analysis, even if the skewing is due to other factors than (just) diffusion, it certainly present an interesting phenomenon that deserves further examination. Important for the present investigation however, is that policy diffusion is (also) very likely to skew the distribution, adding to any skewing effect from other factors. The presence of other skewing effects simply means that the similarity threshold applied to be confident of policy diffusion as the difference maker must be increased. Still, the very highest pairwise similarities should be reserved for situations where all amplifying effects are present - including diffusion.

Inspiration of content and form. Another effect that likely influences document pairwise similarity is format and structure. We have attempted to account for the variance in type and format by prompting the language model to generate origin-neutral summaries of substance and content, without reference to the document type. As noted and evidenced by the various examples given throughout the analysis however, we have not been entirely successful.

The most significant impact of type and format references may have occurred in the Green Taxonomies dataset, though we have not shown examples of direct outputs from this dataset. Whereas literature instilled expectations that China and especially the EU taxonomy was an inspiration to many others, the similarity data did not reflect this, and instead placed the Mongolian taxonomy at the centre of a diffusion network. Notably however, the Mongolian taxonomy is in one aspect much more similar to other documents than the EU and China: in format and structure.

Whereas the EU taxonomy is a legal document with associated delegated acts, most other taxonomies are structured in a report-like format, more akin to various stock-take and strategy reports. These vary greatly in their lengths and detail, but share the same basic structure, the oldest example of which in the dataset is the Mongolian Taxonomy. It seems very likely that this defining feature is shining through to the descriptions, and increasing the resulting pairwise similarity scores between like-formatted documents. As we observed also for the Directives Test-set by example of the GDPR and *Ligestillingsloven*, legal document pairs have higher baseline similarity scores, and seemingly, so do pairs of policy strategies.

It is worth noting that while we have focused on their substance and contents, the format and structure of policies *is* something that can diffuse. Hereby, *if* we are capturing similarities in format between documents, we may be capturing diffusion and inspiration. In light of this, future applications and studies using our approach should perhaps more clearly formulate a definition and expectation as to which objects of diffusion to investigate - and adjust the model instructions accordingly.

We should also note that the taxonomy format innovation is likely to have originated, not in Mongolia, but in an earlier taxonomy by the Climate Bonds Initiative (CBI), a private organisation which since 2013 has been updating a report-structured approach, and which the organisation itself claims has also heavily inspired the EU Taxonomy (Climate Bonds Initiative, 2022). Unfortunately, the earliest CBI taxonomy we were able to locate was from late 2021 and thus could not be included in the analysis to test for early inspiration for e.g. Mongolia.

We can roughly simulate the effect a format-neutral treatment would have had on the Green Taxonomies dataset, by boosting pairwise scores for EU documents to an equivalent baseline as the report-formatted documents. Doing so *does* connect some documents more closely to the EU - especially some of those that were otherwise subject to seemingly spurious connections (such as South Africa) - but far from all, suggesting that other effects are likely also at play in this case.

In this context, if the format and structure of taxonomies did diffuse (whether from one source or the other), it hard to imagine that absolutely no taxonomy substance would diffuse alongside it. This could for example occur through policy consultants, who worked on drafting an early taxonomy, later moving on and applying similar methods elsewhere. Another hypothesis entirely for Mongolia's central role is of course, that the Mongolian taxonomy was a groundbreaking policy innovation widely copied, and that the investigations we have cited so far simply overlooked this, through one bias or the other. While our investigation here does not provide many answers regarding the diffusion of Green Taxonomies, it does provide us with questions to research and a starting point for a more detailed investigation of the topic.

We should also in this context note that, while Larsen (2022) and Gondjian and Merle (2021) do argue that inspiration is flowing from the EU taxonomy, they argue that it especially does so in the form of the underlying principles, such as the 'Do No (Significant) Harm'-principle. Underlying principles of action is not something specifically captured by the prompt we have used in this thesis, and as such, it is perhaps no surprise that our treatment does not well capture this object of diffusion. Once we look specifically for it though, the principle does show up in several of the instruments which the language model extracts. An improved set of prompt instructions could further elucidate such principles of action and their role in policy diffusion.

Instrument and prompt improvements. This all supports the greater point that the language models can and do capture meaningful information and policy diffusion in their descriptions - but that the instructions could be drastically improved. Improvements could take the form of new information fields to better capture the object of one's investigation, improving the 'neutrality' instructions, adding examples of the output format, and improving the instruments search.

Providing additional examples in the prompt would likely improve both instruments and the more general descriptions. For example, instructions could be added to start descriptions with e.g. "This policy aims to...", and thereby potentially avoiding formulations such as "This [Danish/EU/...] Directive/Regulation/Taxonomy aims to...". Not including more examples or specific instructions in the prompt was a conscious decision, partly to keep the prompt brief and readable when included in this report, and partly in light of bias-risks inherent. A related choice was not to tailor the prompt specifically to each datasets or to mirror the questions asked by the studies used for validation. This was motivated by wanting to reuse the prompt across all four datasets, whereby the first three could act as validation for the final analysis.

As seen throughout the analysis, model-extracted policy instruments have been inconsistent, and a disappointment in the light that instruments were intended as the main tool for probing document substance and content as part of the first research question. Primarily, problems arise due to the apparently many different ways of grouping and delineating instruments. To abate this issue, the prompt included an example of instruments grouping (specifically grouping various Extended Producer Responsibility (EPR) Schemes for plastics), and instructions to do so aggressively. Still, improved instructions regarding which instruments to focus on, include, and how to do so consistently are clearly needed.

One small, but likely crucial improvement, would be to limit which and how many words the models can select from when filling out the information fields of the response schema. The prompt applied here involved no restrictions on instrument fields, which lead to the model choosing freely and rarely using the same keyword twice. Were the model to instead select from a list (e.g. the most common keywords output by the unrestricted model), it would likely lead to consistency in outputs, including perhaps also in choice, grouping and titling of instruments.

A risk of designing a prompt with many or specific instructions is to bias the model towards results one expects or wants. As shown, models are eager to solve the task given, including interpreting formulations generously to suit the instructions given. The more specific restrictions one sets on the model in this regard, the higher the risk of hallucinations or biased responses. As an example, the aforementioned EPR-example may have biased the model towards finding primarily regulatory instruments. By providing an example of a regulatory instrument, but not a soft or economic one, we may have inadvertently caused the emphasis on regulatory instruments found across all datasets and models. Such a bias is likely amplified by the very formulations and word choices in the prompt, such as ‘instruments’ (rather than ‘measures’ or ‘interventions’) and that our prompt was initially built up around analysing legal texts specifically, rather than policy documents more broadly. This may have left its mark in the general wording of the prompt, which *might* bias the model towards a focus on regulatory aspects, and harder rules, rather than softer or economic incentives.

Output validation Since Large Language models are effectively black-boxes, we have few ways of probing *how* they arrive at a given output and what has biased them towards exactly those responses. We instead depend on instructing the model to output its reasoning, anchoring in textual quotes, and varying the prompt until arriving at acceptable outputs. Thereby, investigations such as this one often end up in trial-and-error prompt engineering, which carries several inherent risks.

The fundamental challenge is to determine when the prompt ‘works’ or is ‘good (enough)’. This process carries a large risk of confirmation bias, stamping a prompt as usable once it yields anticipated results. This risk is even greater where model outputs are but one step in a longer data pipeline, and thereby perhaps not inspected or validated as thoroughly as final results.

In general, our validation of model outputs has not been systematic or representative. Proper validation involves exhaustive or randomly sampled testing of whether model output conformed to expected results, by beforehand writing down expectations rather than confirming outputs ‘look reasonable’. Notably, we only came close to such a procedure for the Adaptation Strategies dataset. Since validity checks were only performed on English and Danish texts of small-to-medium length, the interpretations of much longer documents, or documents in other languages are in principle untested. For the latter, we rely on Language Model creators having tested and guaranteed model competence in other languages (Google, 2025a). Historically though, Language Model performance is worse in low-resource languages (Park et al., 2024; Wang et al., 2023).

We tried to reduce the risk of confirmation bias by using several test-sets, and thereby relying partly on external validation, rather than our own judgement of model descriptions. We included several information fields in the prompt schema intended for validation of outputs, but were not able to implement automatic checks of these within the project timeframe. Intended checks included confirming the extracted title, date and domain against those listed in the source database, but also using instrument anchors to check instruments against the actual document text - both checking whether the direct quotes can be found in the text and whether this texts semantically matches the generated description.

Likewise, time did not allow for implementing systematic hallucination checks. One technique for doing so is performing multiple, slightly varied treatments of the same text, and observing the resulting variation in outputs. Here the logic is that hallucinations, or outputs otherwise resulting from insufficient instructions and/or information regarding a task, will be inconsistent and thereby result in larger variance (Manakul et al., 2023). Different treatments could for instance be variations in language model, prompts, document pre-treatments, temperature setting or random seed. In particular, the outputs from different treatments can be compared using the resulting embeddings. Using this technique it is possible to validate outputs even where dataset size or language barriers prevent exhaustive manual validation. Documents that show large variations under different treatments can be excluded as outliers, or their resulting embeddings can be manipulated to isolate those dimensions that are stable regardless of treatment - e.g. by averaging embedding vectors. As we have stressed, instrument descriptions are not robust to perturbations of prompt and model choice, though as the appendices show, document-level descriptions mostly are. Still, it would have been fruitful to implement an automated assessment of variance, and this should be done if a process is begun to improve instrument extraction.

Data cleaning. As touched upon, language models are reluctant to point out errors, faulty data, or other situations where a human researcher would normally object to the task given. This stresses the importance of dataset curation and cleaning, to weed out documents with no or little policy relevance. Here, model-extracted information fields could also be applied towards cleaning out those documents which are duplicated, mislabelled, stored under the wrong entries, or otherwise mistreated.

In addition to mislabelling inherent in the original data, a bug during processing of PDF's led to faulty inclusion of documents without textual content. The metadata (title, author, type, domain, ...) for these documents were instead attached the content of another document, whereby documents with duplicated contents were included in the dataset under different and unrelated headings. We tried to remove these from the the final analysis through the document metadata, which included information on length of the processed PDF, why we excluded documents with token lengths under 300. This threshold was found to corresponded well to which documents were riddled with errors, but is not not guaranteed to have caught all sinners.

The FAOLEX database contained several other (nearly) identical PDF documents included under different entries. These we removed when stumbled upon, but not systematically, though such a procedure is possible by manually checking the documents, metadata and model outputs for document pairs with very high similarities. Furthermore, several entries contained, in addition to a primary document, different language versions, consolidated versions or the original policy in case the database entry corresponded to an update. We included these extra documents in the analysis to cast as wide a policy net as possible, which did potentially pollute the dataset with artificially high similarities for an extra 130 entries. Rerunning the analysis with only one file per entry does not significantly change the resulting similarity distribution or network rankings however, and as such these duplicated files are unlikely to have had much effect on the results presented here. We did however specifically avoid including results for networks with similarity restrictions above 0.9, as what few duplicates remained would likely dominate the distribution above this point.

Network rules and statistics. The network analysis presented here mirrors the treatment in previous diffusion studies and restricts networks to only one incoming connection per node, pointing forward in time and in particular to documents dated in later years (Garrett & Jansa, 2015). While such rules result in networks significantly more presentable than alternatives, they are not particularly well-founded theoretically. For instance, countries may draw inspiration from multiple sources or inspire each other through mutual discussions during policy formation, thereby receiving influence from legislation 'before it is adopted'. The literature on Green Taxonomies frequently refers to cases where taxonomies have influenced one another through public consultation drafts, or where documents explicitly draw on design and experience from multiple other taxonomies. The codebase created for this project is flexible enough to accommodate more complex rules and assumptions, but these complicate an already extensive data treatment and its presentability, why we abandoned this line of enquiry.

The choice of how many connections to allow relates to the measure of node importance. Here, we use Katz centrality since it captures the intuitive feeling that if a node has influenced an important node, it is also important. This was in large part motivated by considerations of pioneership and studies like Linsenmeier et al. (2023), that illustrate the large effect early adopters can have on e.g. carbon emissions. What Katz centrality accommodates less well, is that younger nodes have less opportunity to exert influence than older nodes, since the forward-in-time restriction limits them to only few possible connections. This may seem intuitive, but on the other hand, we could wish for a measure of importance that drew attention to younger nodes who have quickly made a large impact. As we shall see in a moment, this may have implications for the interpretations of the keyword distributions. The inclusion of out-degree and descendants as fraction of younger nodes were intended as illustrations of this Katz centrality weakness, but were not widely applied as they were volatile for the very youngest nodes, and overall correlated well with Katz centrality.

How strongly Katz centrality favours older nodes depends on the choice of α parameter. Higher α implies that more influence is attributed to older nodes due to distant descendants. This skews importance towards older nodes, compared to lower α, though the effect is smaller for higher similarity threshold networks, where larger sub-graphs are fragmented into smaller pieces, and distant descendants thereby limited. The relatively high choice of α (0.65) used here was partly motivated by attributing importance to influencing nodes even if spurious connections interceded between a node and what 'should' be its direct descendant. For example, in Figure 6, Irish RE-regulations pop up between Danish RE nodes and the EU RE-directive. If instead we'd allow multiple connections per node, such that the directive could connect to additional sufficiently similar implementations, the 'need' for a significant influence flow through descendants would disappear, and α could be reduced. In any case, the importance of the parameter means the choice of α warrants a more thorough investigation - both theoretically and empirically - of which value is suitable.

The measures studied here can very easily be adapted to study *incoming* (rather than outgoing) influence, and thereby which countries are most vulnerable to or reliant on external policy influence. As a consequence of the network rules used however, the number of ingoing connections to each node is exactly one and nodes are in that sense equally vulnerable, why we did not highlight this investigation.

A likely more fruitful method for investigating both diffusion 'vulnerability' and country statistics more generally, would be to apply network inference algorithms to the networks of policies. Such algorithms infer an underlying network structure of e.g. country-country diffusion, by considering how documents from different countries relate to each other. This would be a much more illuminating quantification than simply averaging state nodes, as it would highlight the pairwise relationships between states - including any particularities in those relationships. This would allow one to study what works for promoting diffusion between certain country pairs, and the various roles a country can take in different relationships, e.g. whether some countries act as amplifiers for policy or policy filters (Desmarais et al., 2015). Rather than assessing importance at the level of each document then, it could be assessed at the higher-level country diffusion network. Here, other network importance measures might be more suitable, given that the resulting graph would not be acyclic.

Keywords and distributions. Network inference would also provide a sounder method for studying diffusion-relevant patterns of policy characteristics, which can account for different characteristics having varying importance in different contexts. For example, one country's policy tradition might be more inclined towards hard regulation than another, which favours voluntary agreements, or certain instruments might be conducive to diffusion in one policy area but counterproductive in another. Such patterns would be hard to observe using the aggregation methods applied thus far. Even so, the demonstrated methods and results deserve some comments.

The relative rarity of keywords such as 'innovation' and 'circular economy' among the influential FAOLEX nodes is interesting, especially considering the recent rise of e.g. mission-based research and innovation policy (Mazzucato et al., 2018). An explanation hereof may be Katz centrality favouring of older nodes with many descendants, especially given the relatively large α applied. Together, these imply a significant temporal lag before nodes representing policy innovations build enough descendants to show importance. Furthermore, the methodology applied won't result in any keywords going 'out of date'. Rather, nodes keep their importance for all future time. Given that diffusion research has shown trends to wax and wane, it might however be fruitful to apply some temporal filter to node importance (Mallinson, 2021). For instance, nodes importance could be set to decay over time, or a Gaussian filter applied to only investigate diffusion within a certain time period. These are all relatively uncomplicated to implement, but were not performed.

It is tempting to conclude from the FAOLEX 'Keyword' distribution (Fig. 13) that a narrower policy focus (i.e. fewer applicable keywords) helps policies diffuse. This is, however, most likely an artefact of which policies from the full database are included in the dataset - and which are left out. Documents only tangentially related to climate change but strongly related to other domains, would likely have several associated keywords. Such documents would likely not have high similarity to many documents in this dataset, and as a result, are unlikely to be considered 'important'. If the dataset were expanded to include domains they relate to more strongly however, that could very well change. Unfortunately, we could not confirm this, as computational constraints limited the number of policies we could include in the dataset. In this context it is also worth reminding that a significant difference between keyword distributions does not clarify causality. Such a difference could be due to certain keywords resulting in more diffusion, but alternatively that drafting a policy which is to be promoted heavily internationally usually comes with certain verbiage. We should note here that we have not, in this analysis, restricted importance or diffusion to only count international links, though it could in principle be done.

Applications. In summary the methodology demonstrated shows promise as a tool for investigating diffusion patterns, and refinements would likely lead to better performance. Still, the methods are not yet mature enough to stand on their own. What this methodology can offer is an initial processing of data that might reveal interesting patterns suitably studied in greater detail. An example is the apparent influence of the Mongolian taxonomy. Whether it is real and how it came about, we cannot answer, but the questions illustrate well how the methodology could be used for case selection and as a check for biases that might be present in other studies of diffusion.

Before embarking on any such investigation though, it seems likely that that point of departure could relatively easily be improved by using Language Models to further clean the datasets and implementing automated validation checks to increase trust in the results. Furthermore, simple prompt improvements and the use of more advanced Language Models will almost certainly improve the substantial probing of policies. Technological development is rapid in this field and already since this work began, several such models have been released. Larger models furthermore enable more detailed citations and textual anchors, thereby improved validation, as well as more detailed instrument descriptions. As demonstrated in Appendix D, individual instruments can be compared much like policy documents, and this further provides a route of inquiry for investigating the flow of inspiration between countries and regions - assuming a standardised and reproducible method of producing instruments can be achieved.

Any conclusions regarding country pioneership should likely also be held off until such improvements are implemented and a fuller picture established, e.g. by implementing network inference algorithms. As it stands, no strong evidence points towards Denmark clearly being in a distinguished pioneer position in general, but it could be that Denmark acts in more subtle ways, inspiring e.g. a subset of the states studied here more heavily, or having a leading role in specific sub-fields. In this context it is worth mentioning the that the FAOLEX network happened to turn up the Danish 2008 National Adaptation Strategy as an important node, in line with Jensen et al. (2023)'s observation that Denmark considered itself a policy pioneer in this area.

While the performance demonstrated is not yet up to par with human qualitative analysts, a more fair comparison is rather to previous text-as-data approaches within policy diffusion. Such studies share most of the assumptions and many of the difficulties presented here, but furthermore have been limited in scope to single-language analysis and unable to comment on document content. In this context, the use of language models to process and compare texts present a significant methodological leap forward. While further refinements are likely to improve performance, already the methodology overcomes the severe limitations of previous approaches and thus represents a significant contribution to the methodological portfolio of policy diffusion scholars.

6 Conclusion

This thesis has explored the use of Large Language Models for analysis of environmental policy and its diffusion patterns. Using generative language models to process, summarise and extract information from several small-to-large datasets, we have demonstrated an experimental design that is able to establish reliable large-scale patterns of policy document similarities. Distributions of document pairwise similarity scores indicate that scores are skewed towards higher similarity, more so than random variation can explain. While several explanations can be hypothesised for such skewing, we have shown that it can at least partially be accounted for by direct or shared inspiration among high-scoring document pairs. When combined with network analysis, we have demonstrated how the similarity distribution can be used to restrict a network of policies to influence-exerting candidates, to study their properties and content.

The methodology presented should not be considered a final formula for Language Model-based investigations of policy, but rather a rough sketch of how policy diffusion and the question of 'Pioneership' *could* be investigated. Detailed probing of document contents remains to be improved sufficiently to create consistently reproducible results, and attempts to reproduce the findings of previous diffusion investigations have shown that the current approach does not match the abilities of human analysts. Notably however, improvements seem achievable. In any case, benchmarking against human analysts is likely both premature and unfair. Rather, the methodology should be evaluated in context of previous text-as-data approaches in policy studies, where the demonstrated procedure presents a significant methodological leap forward, which enables analysts to broaden the geographic and lingual scope of previous investigations, and to dive deeper into the reasons for document similarity.

For the moment, the presented hit-and-miss results regarding causal links between policies do not translate into reliable conclusions regarding who or what act in frontrunner or pioneering roles on the European environmental policy scene. Likewise, the results presented here are unlikely to be of much use informing future policy such that the regional and global impact of environmental policies is maximised. Nonetheless, the glimmers of success displayed do inspire some confidence that informative results can eventually be reached. A long road of research, trial and error is ahead before both technology and its application are mature enough to stand alone as convincing evidence of inspiration flows and their causes - if that time will ever come. Until then, the methods displayed here seem to hold significant promise as both a starting point for further mixed-methods research and a check of results achieved through other means.

References

Abel, D., & Mertens, A. (2023). United in disagreement: Analyzing policy networks in EU policy making. *Review of Policy Research*, *41*(1), 59–82. https://doi.org/10.1111/ropr.12546

Allen-Zhu, Z., & Li, Y. (2024). Physics of Language Models: Part 3.3, Knowledge Capacity Scaling Laws [arXiv:2404.05405 [cs]]. https://doi.org/10.48550/arXiv.2404.05405

An, S., Ma, Z., Lin, Z., Zheng, N., & Lou, J.-G. (2024). Make Your LLM Fully Utilize the Context [arXiv:2404.16811 [cs]]. https://doi.org/10.48550/arXiv.2404.16811

Artifex Software. (2025). PyMuPDF4LLM. Retrieved May 26, 2025, from https://pymupdf.readthedocs.io/en/latest/pymupdf4llm/

Baldwin, E., Carley, S., & Nicholson-Crotty, S. (2019). Why do countries emulate each others' policies? A global study of renewable energy policy diffusion. *World Development*, *120*, 29–45. https://doi.org/10.1016/j.worlddev.2019.03.012

Benara, V., Singh, C., Morris, J. X., Antonello, R., Stoica, I., Huth, A. G., & Gao, J. (2024). Crafting Interpretable Embeddings by Asking LLMs Questions [arXiv:2405.16714 [cs]]. https://doi.org/10.48550/arXiv.2405.16714

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623. https://doi.org/10.1145/3442188.3445922

Berry, F. S., & Berry, W. D. (1990). State Lottery Adoptions as Policy Innovations: An Event History Analysis. *American Political Science Review*, *84*(2), 395–415. https://doi.org/10.2307/1963526

Berry, F. S., & Berry, W. D. (2018). Innovation and Diffusion Models in Policy Research. In C. M. Weible & P. A. Sabatier (Eds.), *Theories of the Policy Process* (4th ed.). Routledge.

Berti, L., Giorgi, F., & Kasneci, G. (2025). Emergent Abilities in Large Language Models: A Survey [arXiv:2503.05788 [cs]]. https://doi.org/10.48550/arXiv.2503.05788

Blatter, J., Portmann, L., & Rausis, F. (2022). Theorizing policy diffusion: From a patchy set of mechanisms to a paradigmatic typology. *Journal of European Public Policy*, *29*(6), 805–825. https://doi.org/10.1080/13501763.2021.1892801

Boehmke, F. J., Brockway, M., Desmarais, B. A., Harden, J. J., LaCombe, S., Linder, F., & Wallach, H. (2020). SPID: A New Database for Inferring Public Policy Innovativeness and Diffusion Networks. *Policy Studies Journal*, *48*(2), 517–545. https://doi.org/10.1111/psj.12357

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., … Amodei, D. (2020). Language models are few-shot learners. *Proceedings of the 34th International Conference on Neural Information Processing Systems*, 1877–1901.

Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., Lee, P., Lee, Y. T., Li, Y., Lundberg, S., Nori, H., Palangi, H., Ribeiro, M. T., & Zhang, Y. (2023, April). Sparks of Artificial General Intelligence: Early experiments with GPT-4 [rXiv:2303.12712 [cs]]. https://doi.org/10.48550/arXiv.2303.12712

Burstein, P. (1991). Policy Domains: Organization, Culture, and Policy Outcomes. *Annual Review of Sociology*, *17*, 327–350. Retrieved May 30, 2025, from https://www.jstor.org/stable/2083346

Calderon, N., & Reichart, R. (2025). On Behalf of the Stakeholders: Trends in NLP Model Interpretability in the Era of LLMs [Xiv:2407.19200 [cs]]. https://doi.org/10.48550/arXiv.2407.19200

Cao, X. (2010). Networks as Channels of Policy Diffusion: Explaining Worldwide Changes in Capital Taxation, 1998–2006. *International Studies Quarterly*, *54*(3), 823–854. https://doi.org/10.1111/j.1468-2478.2010.00611.x

Chalmers, A., Klingler-Vidra, R., & Malou van den Broek, O. (2024). From Diffusion to Diffuse-ability: A Text-as-Data Approach to Explaining the Global Diffusion of Corporate Sustainability Policy. *International Studies Quarterly*, *68*(1), sqad106. https://doi.org/10.1093/isq/sqad106

Chew, R., Bollenbacher, J., Wenger, M., Speer, J., & Kim, A. (2023). LLM-Assisted Content Analysis: Using Large Language Models to Support Deductive Coding [arXiv:2306.14924 [cs]]. https://doi.org/10.48550/arXiv.2306.14924

Chung, H. W., Hou, L., Longpre, S., Zoph, B., Tay, Y., Fedus, W., Li, Y., Wang, X., Dehghani, M., Brahma, S., Webson, A., Gu, S. S., Dai, Z., Suzgun, M., Chen, X., Chowdhery, A., Castro-Ros, A., Pellat, M., Robinson, K., … Wei, J. (2022). Scaling Instruction-Finetuned Language Models [arXiv:2210.11416 [cs]]. https://doi.org/10.48550/arXiv.2210.11416

Climate Bonds Initiative. (2022). Climate Bonds Taxonomy. Retrieved June 1, 2025, from https://www.climatebonds.net/standard/taxonomy

Colvin, S., Jolibois, E., Ramezani, H., Garcia Badaracco, A., Dorsey, T., Montague, D., Matveenko, S., Trylesinski, M., Runkle, S., Hewitt, D., Hall, A., & Plot, V. (2025). Pydantic. Retrieved May 26, 2025, from https://github.com/pydantic/pydantic

Cook. (1984). *The Black Company*. Tor Fantasy.

Creswell, J. W., & Poth, C. N. (2016). *Qualitative Inquiry and Research Design: Choosing Among Five Approaches*. SAGE Publications.

Dansk Industri. (n.d.). Danmark som globalt, grønt foregangsland. Retrieved May 26, 2025, from https://www.danskindustri.dk/politik-og-analyser/dis-politiske-udspil/gront-foregangsland/

Desmarais, B. A., Harden, J. J., & Boehmke, F. J. (2015). Persistent Policy Pathways: Inferring Diffusion Networks in the American States. *American Political Science Review*, *109*(2), 392–406. https://doi.org/10.1017/S0003055415000040

Dodge, J., Sap, M., Marasović, A., Agnew, W., Ilharco, G., Groeneveld, D., Mitchell, M., & Gardner, M. (2021). Documenting Large Webtext Corpora: A Case Study on the Colossal Clean Crawled Corpus [arXiv:2104.08758 [cs]]. https://doi.org/10.48550/arXiv.2104.08758

Doshi-Velez, F., & Kim, B. (2017). Towards A Rigorous Science of Interpretable Machine Learning [arXiv:1702.08608 [stat]]. https://doi.org/10.48550/arXiv.1702.08608

Dreyer, E., Hansen, T., Holmberg, K., Olsen, T., & Stripple, J. (2024). *Towards a Global Plastics Treaty: Tracing the UN Negotiations*. Lund University.

Efron, B., & Tibshirani, R. J. (1994). 15: Permutation tests. In *An Introduction to the Bootstrap* (pp. 202–214). Chapman; Hall/CRC. https://doi.org/10.1201/9780429246593

Elhage, N., Hume, T., Olsson, C., Schiefer, N., Henighan, T., Kravec, S., Hatfield-Dodds, Z., Lasenby, R., Drain, D., Chen, C., Grosse, R., McCandlish, S., Kaplan, J., Amodei, D., Wattenberg, M., & Olah, C. (2022). Toy Models of Superposition. *Transformer Circuits Thread*. Retrieved May 26, 2025, from https://transformer-circuits.pub/2022/toy_model/

FAO. (n.d.). Faolex Metadata. Retrieved May 26, 2025, from https://www.fao.org/faolex/opendata/metadata/en

FAO. (2025, April). Open Data. Retrieved April 14, 2025, from https://www.fao.org/faolex/opendata

Fu, Y. (2022). A Closer Look at Large Language Models Emergent Abilities. Retrieved May 26, 2025, from https://yaofu.notion.site/A-Closer-Look-at-Large-Language-Models-Emergent-Abilities-493876b55df5479d80686f68a1abd72f

Gansner, E. R., & North, S. C. (2000). An open graph visualization system and its applications to software engineering. *Software: Practice and Experience*, *30*(11), 1203–1233. https://doi.org/10.1002/1097-024X(200009)30:11<1203::AID-SPE338>3.0.CO;2-N

Garrett, K. N., & Jansa, J. M. (2015). Interest Group Influence in Policy Diffusion Networks. *State Politics & Policy Quarterly*, *15*(3), 387–417. https://doi.org/10.1177/1532440015592776

Gilardi, F., & Füglister, K. (2008). Empirical Modeling of Policy Diffusion in Federal States: The Dyadic Approach. *Swiss Political Science Review*, *14*(3), 413–450. https://doi.org/10.1002/j.1662-6370.2008.tb00108.x

Gilardi, F., & Wasserfallen, F. (2019). The politics of policy diffusion. *European Journal of Political Research*, *58*(4), 1245–1256. https://doi.org/10.1111/1475-6765.12326

Glick, H. R., & Hays, S. P. (1991). Innovation and Reinvention in State Policymaking: Theory and the Evolution of Living Will Laws. *The Journal of Politics*, *53*(3), 835–850. https://doi.org/10.2307/2131581

Gondjian, G., & Merle, C. (2021). Sustainable Taxonomy development worldwide: A standard-setting race between competing jurisdictions. Retrieved May 26, 2025, from https://gsh.cib.natixis.com/our-center-of-expertise/articles/sustainable-taxonomy-development-worldwide-a-standard-setting-race-between-competing-jurisdictions

Google. (2025a). Google models | Language support. Retrieved May 30, 2025, from https://cloud.google.com/vertex-ai/generative-ai/docs/models

Google. (2025b). Structured output | Gemini API. Retrieved May 26, 2025, from https://ai.google.dev/gemini-api/docs/structured-output

Google. (2025c, May). Googleapis/python-genai. Retrieved May 26, 2025, from https://github.com/googleapis/python-genai

Graham, E. R., Shipan, C. R., & Volden, C. (2013). The Diffusion of Policy Diffusion Research in Political Science. *British Journal of Political Science*, *43*(3), 673–701. https://doi.org/10.1017/S0007123412000415

Hagberg, A. A., Schult, D. A., & Renieris, M. (2024, September). Pygraphviz/pygraphviz. Retrieved May 26, 2025, from https://github.com/pygraphviz/pygraphviz

Hagberg, A. A., Schult, D. A., & Swart, P. J. (2008). Exploring Network Structure, Dynamics, and Function using NetworkX. In G. Varoquaux, T. Vaught, & J. Millman (Eds.), *Proceedings of the 7th Python in Science Conference* (pp. 11–15).

Hoffmann, J., Borgeaud, S., Mensch, A., Buchatskaya, E., Cai, T., Rutherford, E., Casas, D. d. L., Hendricks, L. A., Welbl, J., Clark, A., Hennigan, T., Noland, E., Millican, K., Driessche, G. v. d., Damoc, B., Guy, A., Osindero, S., Simonyan, K., Elsen, E., … Sifre, L. (2022). Training Compute-Optimal Large Language Models [arXiv:2203.15556 [cs]]. https://doi.org/10.48550/arXiv.2203.15556

Holm, A. (2024). Danmark har på magisk vis skabt sig et rygte som grønt foregangsland. Det er vist ikke en beskyttet titel. *Altinget*. Retrieved May 26, 2025, from https://www.altinget.dk/artikel/danmark-har-paa-magisk-vis-skabt-sig-et-rygte-som-groent-foregangsland-nu-er-tiden-inde-til-at-vi-faktisk-bliver-det

Hu, P., Liu, S., Gao, C., Huang, X., Han, X., Feng, J., Deng, C., & Huang, S. (2025). Large Language Models Are Cross-Lingual Knowledge-Free Reasoners [arXiv:2406.16655 [cs]]. https://doi.org/10.48550/arXiv.2406.16655

Hunter, J. D. (2007). Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering*, *9*(3), 90–95. https://doi.org/10.1109/MCSE.2007.55

Islas, L. (2023). Gemini, Google's new Artificial Intelligence explained in 5 key facts. *Revista Merca2.0*. Retrieved May 26, 2025, from https://www.merca20.com/gemini-googles-new-artificial-intelligence-explained-in-5-key-facts/

Jänicke, M. (2005). Trend-setters in environmental policy: The character and role of pioneer countries. *European Environment*, *15*(2), 129–142. https://doi.org/10.1002/eet.375

Jensen, A., Nielsen, H. Ø., & Russel, D. (2023). Diffusion of climate policy integration in adaptation strategies: Translating the EU mandate into UK and Danish national contexts. *Regional Environmental Change*, *23*(4), 130. https://doi.org/10.1007/s10113-023-02110-6

Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y., Chen, D., Dai, W., Chan, H. S., Madotto, A., & Fung, P. (2023). Survey of Hallucination in Natural Language Generation [arXiv:2202.03629 [cs]]. *ACM Computing Surveys*, *55*(12), 1–38. https://doi.org/10.1145/3571730

Jiao, W., Wang, W., Huang, J.-t., Wang, X., Shi, S., & Tu, Z. (2023). Is ChatGPT A Good Translator? Yes With GPT-4 As The Engine [arXiv:2301.08745 [cs]]. https://doi.org/10.48550/arXiv.2301.08745

Jørgensen, D. (2022). Dan Jørgensen: Hele verden kalder Danmark for et grønt foregangsland. Fordi vi er det. *Information*. Retrieved May 26, 2025, from https://www.information.dk/debat/2022/10/dan-joergensen-hele-verden-kalder-danmark-groent-foregangsland-fordi

Kammerer, M., & Namhata, C. (2018). What drives the adoption of climate change mitigation policy? A dynamic network approach to policy diffusion. *Policy Sciences*, *51*(4), 477–513. https://doi.org/10.1007/s11077-018-9332-6

Kaplan, J., McCandlish, S., Henighan, T., Brown, T. B., Chess, B., Child, R., Gray, S., Radford, A., Wu, J., & Amodei, D. (2020). Scaling Laws for Neural Language Models [arXiv:2001.08361 [cs]]. https://doi.org/10.48550/arXiv.2001.08361

Kavukcuoglu, K. (2025). Gemini 2.5: Our most intelligent AI model. Retrieved May 26, 2025, from https://blog.google/technology/google-deepmind/gemini-model-thinking-updates-march-2025/

Klima-, Energi- og Forsyningsministeriet. (2021). Bekendtgørelse af lov om klima. Retrieved May 26, 2025, from https://www.retsinformation.dk/eli/lta/2021/2580

Klimarådet. (2023). *Danmarks global klimaindsats* (tech. rep.). https://klimaraadet.dk/da/analyse/danmarks-globale-klimaindsats

Kojima, T., Gu, S. S., Reid, M., Matsuo, Y., & Iwasawa, Y. (2023). Large Language Models are Zero-Shot Reasoners [arXiv:2205.11916 [cs]]. https://doi.org/10.48550/arXiv.2205.11916

Kryściński, W., McCann, B., Xiong, C., & Socher, R. (2019). Evaluating the Factual Consistency of Abstractive Text Summarization [arXiv:1910.12840 [cs]]. https://doi.org/10.48550/arXiv.1910.12840

Larsen, M. L. (2022). Driving Global Convergence in Green Financial Policies: China as Policy Pioneer and the EU as Standard Setter. *Global Policy*, *13*(3), 358–370. https://doi.org/10.1111/1758-5899.13105

Lin, C., & Ahrens, K. (2005). How Many Meanings Does a Word Have? Meaning Estimation in Chinese and English. In J. W. Minett & W. S.-Y. Wang (Eds.), *Language Acquisition, Change and Emergence: Essays in Evolutionary Linguistics* (pp. 437–464).

Lin, G. (2025). Using cross-encoders to measure the similarity of short texts in political science. *American Journal of Political Science*, *n/a*(n/a). https://doi.org/10.1111/ajps.12956

Lin, J. (1991). Divergence measures based on the Shannon entropy. *IEEE Transactions on Information Theory*, *37*(1), 145–151. https://doi.org/10.1109/18.61115

Linder, F., Desmarais, B., Burgess, M., & Giraudy, E. (2018). Text as Policy: Measuring Policy Similarity through Bill Text Reuse. *Policy Studies Journal*, *48*(2), 546–574. https://doi.org/10.1111/psj.12257

Linsenmeier, M., Mohommad, A., & Schwerhoff, G. (2023). Global benefits of the international diffusion of carbon pricing policies. *Nature Climate Change*, *13*(7), 679–684. https://doi.org/10.1038/s41558-023-01710-8

Liu, H., Yan, W., Zaharia, M., & Abbeel, P. (2025). World Model on Million-Length Video And Language With Blockwise RingAttention [arXiv:2402.08268 [cs]]. https://doi.org/10.48550/arXiv.2402.08268

Liu, N. F., Lin, K., Hewitt, J., Paranjape, A., Bevilacqua, M., Petroni, F., & Liang, P. (2023). Lost in the Middle: How Language Models Use Long Contexts [arXiv:2307.03172 [cs]]. https://doi.org/10.48550/arXiv.2307.03172

Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H., & Neubig, G. (2021). Pre-train, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing [arXiv:2107.13586 [cs]]. https://doi.org/10.48550/arXiv.2107.13586

Lu, S., Bigoulaeva, I., Sachdeva, R., Tayyar Madabushi, H., & Gurevych, I. (2024). Are Emergent Abilities in Large Language Models just In-Context Learning? In L.-W. Ku, A. Martins, & V. Srikumar (Eds.), *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 5098–5139). Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.acl-long.279

Maggiori, E. (2023). ChatGPT's vocabulary: The words that ChatGPT knows and how they were chosen – Emmanuel Maggiori. Retrieved May 26, 2025, from https://emaggiori.com/chatgpt-vocabulary/

Mallinson, D. J. (2021). Who Are Your Neighbors? The Role of Ideology and Decline of Geographic Proximity in the Diffusion of Policy Innovations. *Policy Studies Journal*, *49*(1), 67–88. https://doi.org/10.1111/psj.12351

Manakul, P., Liusie, A., & Gales, M. J. F. (2023). SelfCheckGPT: Zero-Resource Black-Box Hallucination Detection for Generative Large Language Models [arXiv:2303.08896 [cs]]. https://doi.org/10.48550/arXiv.2303.08896

Maynez, J., Narayan, S., Bohnet, B., & McDonald, R. (2020). On Faithfulness and Factuality in Abstractive Summarization [arXiv:2005.00661 [cs]]. https://doi.org/10.48550/arXiv.2005.00661

Mazzucato, M., Directorate-General for Research and Innovation, & European Commission. (2018). *Mission-oriented research & innovation in the European Union: A problem solving approach to fuel innovation led growth*. Publications Office of the European Union. Retrieved May 30, 2025, from https://data.europa.eu/doi/10.2777/360325

McCann, E., & Ward, K. (2013). A multi-disciplinary approach to policy transfer research: Geographies, assemblages, mobilities and mutations [Publisher: RSA Website _eprint: https://doi.org/10.1080/01442872.2012.748563]. *Policy Studies*, *34*(1), 2–18. https://doi.org/10.1080/01442872.2012.748563

Meta. (2025). The Llama 4 herd: The beginning of a new era of natively multimodal AI innovation. Retrieved May 26, 2025, from https://ai.meta.com/blog/llama-4-multimodal-intelligence/

Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2*, *2*, 3111–3119.

Min, S., Lyu, X., Holtzman, A., Artetxe, M., Lewis, M., Hajishirzi, H., & Zettlemoyer, L. (2022). Rethinking the Role of Demonstrations: What Makes In-Context Learning Work? [arXiv:2202.12837 [cs]]. https://doi.org/10.48550/arXiv.2202.12837

Natixis. (2023). *The New Geography of Taxonomies* (tech. rep.). Natixis. Retrieved May 26, 2025, from https://gsh.cib.natixis.com/our-center-of-expertise/articles/the-new-geography-of-taxonomies

Newman, M. E. J. (2010). Measures and metrics. In *Networks: An Introduction* (pp. 168–234). Oxford University Press. Retrieved May 26, 2025, from https://doi.org/10.1093/acprof:oso/9780199206650.003.0007

Oliveira, O. P. d., Romano, G. C., Volden, C., & Karch, A. (2023). Policy Diffusion and Innovation. In C. M. Weible (Ed.), *Theories Of The Policy Process* (5th ed.). Routledge.

OpenAI. (2025a). Introducing GPT-4.1 in the API. Retrieved May 26, 2025, from https://openai.com/index/gpt-4-1/

OpenAI. (2025b). *OpenAI o3 and o4-mini System Card* (tech. rep.). Retrieved May 26, 2025, from https://openai.com/index/o3-o4-mini-system-card/

OpenAI. (2025c). Openai/openai-python. Retrieved May 26, 2025, from https://github.com/openai/openai-python

Opitz, J., Möller, L., Michail, A., & Clematide, S. (2025). Interpretable Text Embeddings and Text Similarity Explanation: A Primer [arXiv:2502.14862 [cs]]. https://doi.org/10.48550/arXiv.2502.14862

Park, G., Hwang, S., & Lee, H. (2024). Low-Resource Cross-Lingual Summarization through Few-Shot Learning with Large Language Models [arXiv:2406.04630 [cs]]. https://doi.org/10.48550/arXiv.2406.04630

Patel, D., & Wong. (2023, July). GPT-4 Architecture, Infrastructure, Training Dataset, Costs, Vision, MoE. Retrieved May 26, 2025, from https://semianalysis.com/2023/07/10/gpt-4-architecture-infrastructure/

Patton, M. Q. (2014). *Qualitative Research & Evaluation Methods: Integrating Theory and Practice*. SAGE Publications.

Peck, J., & Theodore, N. (2010). Mobilizing policy: Models, methods, and mutations. *Geoforum*, *41*(2), 169–174. https://doi.org/10.1016/j.geoforum.2010.01.002

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, *12*(85), 2825–2830. Retrieved May 26, 2025, from http://jmlr.org/papers/v12/pedregosa11a.html

Peters, M. E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., & Zettlemoyer, L. (2018). Deep Contextualized Word Representations. In M. Walker, H. Ji, & A. Stent (Eds.), *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)* (pp. 2227–2237). Association for Computational Linguistics. https://doi.org/10.18653/v1/N18-1202

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). *Language models are unsupervised multitask learners* (tech. rep.). OpenAI. Retrieved May 26, 2025, from https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why Should I Trust You?": Explaining the Predictions of Any Classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1135–1144. https://doi.org/10.1145/2939672.2939778

Ritzau. (2009). Time: Danmark er grønt foregangsland. *DR*. Retrieved May 26, 2025, from https://www.dr.dk/nyheder/udland/time-danmark-er-groent-foregangsland

Rothernborg, M. (2001). Gult lys til grønne investeringer. *Politiken*. Retrieved May 26, 2025, from https://politiken.dk/danmark/oekonomi/art4947890/Gult-lys-til-gr%C3%B8nne-investeringer

Saldana, J. (2021). *The Coding Manual for Qualitative Researchers*. SAGE.

Sanderson, G. (2024). How might LLMs store facts. Retrieved May 26, 2025, from https://www.youtube.com/watch?v=9-Jl0dxWQs8

Sardana, N., Portes, J., Doubov, S., & Frankle, J. (2025). Beyond Chinchilla-Optimal: Accounting for Inference in Language Model Scaling Laws [arXiv:2401.00448 [cs] version: 3]. https://doi.org/10.48550/arXiv.2401.00448

Schaeffer, R., Miranda, B., & Koyejo, S. (2023). Are Emergent Abilities of Large Language Models a Mirage? [arXiv:2304.15004 [cs]]. https://doi.org/10.48550/arXiv.2304.15004

Schoenefeld, J. J., Schulze, K., & Bruch, N. (2022). The diffusion of climate change adaptation policy. *WIREs Climate Change*, *13*(3), e775. https://doi.org/10.1002/wcc.775

Sennrich, R., Haddow, B., & Birch, A. (2016). Neural Machine Translation of Rare Words with Subword Units. In K. Erk & N. A. Smith (Eds.), *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 1715–1725). Association for Computational Linguistics. https://doi.org/10.18653/v1/P16-1162

Shipan, C. R., & Volden, C. (2008). The Mechanisms of Policy Diffusion. *American Journal of Political Science*, *52*(4), 840–857. Retrieved May 26, 2025, from https://www.jstor.org/stable/25193853

Shuster, K., Poff, S., Chen, M., Kiela, D., & Weston, J. (2021). Retrieval Augmentation Reduces Hallucination in Conversation. Retrieved May 26, 2025, from https://arxiv.org/abs/2104.07567v1

Si, C., Gan, Z., Yang, Z., Wang, S., Wang, J., Boyd-Graber, J., & Wang, L. (2023). Prompting GPT-3 To Be Reliable [arXiv:2210.09150 [cs]]. https://doi.org/10.48550/arXiv.2210.09150

Software, A. (2025, May). Pymupdf/PyMuPDF [original-date: 2012-10-06T18:54:25Z]. Retrieved May 26, 2025, from https://github.com/pymupdf/PyMuPDF

Srivastava, A., Rastogi, A., Rao, A., Shoeb, A. A. M., Abid, A., Fisch, A., Brown, A. R., Santoro, A., Gupta, A., Garriga-Alonso, A., Kluska, A., Lewkowycz, A., Agarwal, A., Power, A., Ray, A., Warstadt, A., Kocurek, A. W., Safaya, A., Tazarv, A., … Wu, Z. (2023). Beyond the Imitation Game: Quantifying and extrapolating the capabilities of language models [arXiv:2206.04615 [cs]]. https://doi.org/10.48550/arXiv.2206.04615

Starke, P. (2013). Qualitative Methods for the Study of Policy Diffusion: Challenges and Available Solutions. *Policy Studies Journal*, *41*(4), 561–582. https://doi.org/10.1111/psj.12032

Steinebach, Y., Fernández-i-Marín, X., & Aschenbrenner, C. (2021). Who puts a price on carbon, why and how? A global empirical analysis of carbon pricing policies. *Climate Policy*, *21*(3), 277–289. https://doi.org/10.1080/14693062.2020.1824890

Temenos, C., & McCann, E. (2013). Geographies of Policy Mobilities. *Geography Compass*, *7*(5), 344–357. https://doi.org/10.1111/gec3.12063

Templeton, A., Conerly, T., Marcus, J., Lindsey, J., Bricken, T., Chen, B., Pearce, A., Citro, C., Ameisen, E., Jones, A., Cunningham, H., Turner, N. L., McDougall, C., MacDiarmid, M., Freeman, C. D., Sumers, T. R., Rees, E., Batson, J., Jermyn, A., … Henighan, T. (2024). Scaling Monosemanticity: Extracting Interpretable Features from Claude 3 Sonnet. *Transformer Circuits Thread*. https://transformer-circuits.pub/2024/scaling-monosemanticity/index.html

Turpin, M., Michael, J., Perez, E., & Bowman, S. R. (2023). Language Models Don't Always Say What They Think: Unfaithful Explanations in Chain-of-Thought Prompting [arXiv:2305.04388 [cs]]. https://doi.org/10.48550/arXiv.2305.04388

Udenrigsministeriet. (n.d.). Climate and energy. Retrieved May 26, 2025, from https://vietnam.um.dk/en/climate-energy

UNEP. (2023). *Common Framework of Sustainable Finance Taxonomies for Latin America and the Caribbean*. Retrieved May 26, 2025, from https://wedocs.unep.org/xmlui/handle/20.500.11822/42967

UNSD. (n.d.). Standard country or area codes for statistical use (M49). Retrieved May 26, 2025, from https://unstats.un.org/unsd/methodology/m49/overview/

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 6000–6010.

Villalobos, P., Ho, A., Sevilla, J., Besiroglu, T., Heim, L., & Hobbhahn, M. (2024). Will we run out of data? Limits of LLM scaling based on human-generated data [arXiv:2211.04325 [cs] version: 2]. https://doi.org/10.48550/arXiv.2211.04325

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., ... van Mulbregt, P. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python [Publisher: Nature Publishing Group]. *Nature Methods*, *17*(3), 261–272. https://doi.org/10.1038/s41592-019-0686-2

Volden, C. (2006). States as Policy Laboratories: Emulating Success in the Children's Health Insurance Program. *American Journal of Political Science*, *50*(2), 294–312. https://doi.org/10.1111/j.1540-5907.2006.00185.x

Volden, C., Ting, M. M., & Carpenter, D. P. (2008). A Formal Model of Learning and Policy Diffusion. *American Political Science Review*, *102*(3), 319–332. https://doi.org/10.1017/S0003055408080271

Walker, J. L. (1969). The Diffusion of Innovations among the American States. *American Political Science Review*, *63*(3), 880–899. https://doi.org/10.2307/1954434

Wang, J., Liang, Y., Meng, F., Zou, B., Li, Z., Qu, J., & Zhou, J. (2023). Zero-Shot Cross-Lingual Summarization via Large Language Models [arXiv:2302.14229 [cs]]. https://doi.org/10.48550/arXiv.2302.14229

Waskom, M. L. (2021). Seaborn: Statistical data visualization. *Journal of Open Source Software*, *6*(60), 3021. https://doi.org/10.21105/joss.03021

Wasserman, S., & Faust, K. (1994). Centrality and Prestige. In *Social Network Analysis: Methods and Applications* (pp. 169–219). Cambridge University Press. Retrieved May 26, 2025, from https://www.cambridge.org/core/books/social-network-analysis/centrality-and-prestige/D845CD4DD5B5780E3860DF4DD36D5509

Webson, A., & Pavlick, E. (2022). Do Prompt-Based Models Really Understand the Meaning of their Prompts? [arXiv:2109.01247 [cs]]. https://doi.org/10.48550/arXiv.2109.01247

Wei, J., Tay, Y., Bommasani, R., Raffel, C., Zoph, B., Borgeaud, S., Yogatama, D., Bosma, M., Zhou, D., Metzler, D., Chi, E. H., Hashimoto, T., Vinyals, O., Liang, P., Dean, J., & Fedus, W. (2022). Emergent Abilities of Large Language Models [arXiv:2206.07682 [cs]]. https://doi.org/10.48550/arXiv.2206.07682

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Ichter, B., Xia, F., Chi, E., Le, Q., & Zhou, D. (2023). Chain-of-Thought Prompting Elicits Reasoning in Large Language Models [arXiv:2201.11903 [cs]]. https://doi.org/10.48550/arXiv.2201.11903

Westersø, R. S. (2023). Mette Frederiksen adresserer klimakritik: - Jeg har en helt anden version - TV 2. *TV2*. Retrieved May 26, 2025, from https://nyheder.tv2.dk/politik/2023-08-13-mette-frederiksen-adresserer-klimakritik-jeg-har-en-helt-anden-version

White, J., Fu, Q., Hays, S., Sandborn, M., Olea, C., Gilbert, H., Elnashar, A., Spencer-Smith, J., & Schmidt, D. C. (2023). A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT [arXiv:2302.11382 [cs]]. https://doi.org/10.48550/arXiv.2302.11382

Yoder, N. (2025, February). Beyond Orthogonality: How Language Models Pack Billions of Concepts into 12,000 Dimensions. Retrieved May 26, 2025, from https://nickyoder.com/johnson-lindenstrauss/

Yu, W., Iter, D., Wang, S., Xu, Y., Ju, M., Sanyal, S., Zhu, C., Zeng, M., & Jiang, M. (2023). Generate rather than Retrieve: Large Language Models are Strong Context Generators [arXiv:2209.10063 [cs]]. https://doi.org/10.48550/arXiv.2209.10063

AI Declaration

Declaration of using generative AI tools (for students)

☒ I/we have used generative AI as an aid/tool (please tick)
☐ I/we have NOT used generative AI as an aid/tool (please tick)

List which GAI tools you have used and include the link to the platform (if possible):

- Copilot with enterprise data protection (UCPH license), https://copilot.microsoft.com
- Google Gemini, https://aistudio.google.com/
- ChatGPT, https://chatgpt.com/
- OpenAI Playground https://platform.openai.com/playground

Describe how generative AI has been used in the exam paper:

1. *Purpose (what did you use the tool for?)*
2. *Work phase (when in the process did you use GAI?)*
3. *What did you do with the output? (including any editing of or continued work on the output)*

1. Conceptual sparring; Literature search; Methodological suggestions; Coding assistance; Word clarifications; Prompt Generation
2. Conceptualisation, Data treatment (python coding assistance, methodological suggestions, prompt correction and improvement), Writing and Editing (Literature search for sources and supporting literature, word clarifications, Latex coding assistance)
3. Verified suggestions, adapted and incorporated suggestions if they were applicable. Rejected, reprompted or continued conversation if initial output not useful. Iterative and step-wise improvements, as well as coding assistance for bug fixing, optimisation and prompting.

Appendices

Appendix A gives longer, more technical introductions to the academic topics relevant for this thesis.
Appendix B lists the prompt used in the main body analysis, as well as results from an earlier prompt version to highlight both robustness and fragility of the prompt. The following Appendix (C) contain results for the EU Directive's testset when varying the Large Language Model used to process documents.
Appendix D demonstrates how instruments can be compared using similarity matrixes.
Appendix E contains snapshots of the FAOLEX network under different network-drawing rules.
Appendix F shows the three smaller datasets included in this study in tabular format.

A Extended Background

The following subsections provide soft and accessible introductions to the academic topics relevant to this thesis, aimed at those readers unfamiliar with the subject(s). It begins with a review of Policy Diffusion studies, with emphasis on the methodological challenges encountered in this field. Then follows an introduction to Text Embeddings. Furthermore we describe how Large Language Models can be used to pre-process documents and prepare them for embedding. The final subsection introduces network analytics.

A.1 Policy Diffusion and related fields of research

The study of if, how and why policies spread between different administrative and geographical areas has a long history under many different headings and fields of academic research (Graham et al., 2013). In political science alone, research on the topic has been published under the headings of (policy-) 'diffusion', 'innovation', 'convergence', 'transfer', 'circulation', 'mobilities', 'harmonization', 'contagion', 'integration', 'process tracing', and others (Graham et al., 2013; Oliveira et al., 2023).Throughout these disciplines, scholars vary in how exactly they define diffusion (or related concepts), including notably on whether diffusion occurs over time to subsequent policies in other jurisdictions, or whether the definitions allow 'policy choices to be interdependent, simultaneous or anticipatory' (Graham et al., 2013, p. 675). Scholars have largely agreed on four mechanisms through which policies diffuse: *Learning*, *Competition*, *Coercion* and *Emulation*, although the exact categories, including their number and mutual delineations, is not entirely undisputed, and alternative typologies and categorisations exist (Berry & Berry, 2018).[30] *Learning* is the process where policy-makers base their policy choices on experiences from other

[30]For alternatives, see e.g. Blatter et al. (2022) and Gilardi and Wasserfallen (2019)

jurisdictions and governments, allowing policies to diffuse if they are deemed 'successful'. *Competition* is when governments base policy choices on the objective of achieving competitive advantages over others. *Coercion* is when governments adopt certain policies as a result of explicit outside pressures, e.g. in the form of force or incentives. Finally, *Emulation* occurs when similar perceptions of 'proper conduct' across polities leads to adoption of similar policies. Note that these diffusion mechanisms may overlap conflict to push different policies which can delay the policy process (Berry & Berry, 2018).It is hopefully apparent that policy diffusion does not necessarily lead to improved outcomes, as various perverse motives and incentives can force the diffusion of sub-optimal policies, and that the success of policy in one location does not guarantee its success elsewhere (Shipan & Volden, 2008).

Quantitative diffusion studies. The birth of diffusion studies is often attributed to Walker (1969), who introduced 'external factors' as a relevant consideration when analysing policy innovations in American states. External factors complement 'internal factors' in affecting and explaining policy choices. Since then, research under the different diffusion-related traditions have employed different methods. Notably, the 'policy diffusion' subfield has largely been focused on large-N quantitative studies. This followed Berry and Berry (1990) seminal study, which introduced 'Event History Analysis' as a quantitative approach to uniting studies of internal and external factors.[31] Event History Analysis (EHA) in its simplest form uses statistics to model the likelihood of adopting a certain (type of) policy and its dependency on variables representing both the internal and external characteristics of the jurisdiction in question. More advanced 'dyadic' approaches furthermore incorporate and model the influence of pairwise interactions between jurisdictions (Gilardi & Füglister, 2008; Volden, 2006).

Using these techniques, researchers have demonstrated positive correlations between policy adoption likelihoods and indicators of external influence such as: adoption in neighbouring jurisdictions (Berry & Berry, 2018; Linsenmeier et al., 2023), ideological similarities to adopters (Berry & Berry, 2018; Schoenefeld et al., 2022), trading partners and similar trade patterns to other adopters (Steinebach et al., 2021), diplomatic interactions with other adopters (Kammerer & Namhata, 2018), donors of foreign aid (Baldwin et al., 2019) and many more (Berry & Berry, 2018; Graham et al., 2013). Several of these relationships are however disputed, or show evolutions in relative importance over time. For instance, the importance of geographic proximity has seemingly diminished over time (Mallinson, 2021). This may be attributable to sometimes questionable choices of indicators for inter-jurisdictional contact

[31] Event History Analysis is also known by other names in other fields, including survival analysis, hereunder the popular (proportional) hazard models which one often finds references to in Policy Diffusion studies. We'll use to term to cover all of these related approaches.

relationships (Blatter et al., 2022). In any case, researchers have used the estimated diffusion rates to model the resulting impacts of policy diffusion, e.g. estimating and showing significant global reductions in carbon emissions resulting from the diffusion of carbon pricing policies (Linsenmeier et al., 2023).
Despite their popularity, EHAs (and quantitative diffusion studies more generally) have several methodological limitations, shortcomings and issues. For one, EHA studies rarely engage with the actual text or substance of policies, including their mutual similarities and differences. Instead, policies within a certain policy/issue area are most often included only as ones or zeros, indicating whether or not a certain jurisdiction has taken *some* policy action within the given area. This approach also often limits EHAs to studying a single policy area at once, and thereby not capturing broader diffusion trends or relationships (Oliveira et al., 2023, p.234). Researchers will often conclude that positive correlations between adoption likelihood and independent variables representing diffusion implies increased similarities between adopted policies in correlated jurisdictions, but as mentioned often without any actual engagement with the text or content of policy documents.[32] Fundamentally however, the largest methodological issue with such quantitative investigations remain the lack of causal explainability. Causation is not to be conflated with correlation, and correlations between 'external factors' in adopting jurisdictions does not serve as evidence that diffusion of policies have occurred. Governments might adopt policies within an issue area at similar times, simply in response to facing similar challenges (Volden et al., 2008). In such cases, neighbouring, co-trading or ideologically aligned governments *may* exchange policy experiences, but the simple observation of mutual policy action in principle does not suffice to evidence that such an exchange occurred.

Quanlitative diffusion studies have been viewed as key to overcoming the challenges of quantitative analysis (Starke, 2013). For instance, the 'policy mobilities' sub-field has applied qualitative methods to follow policies through adoption in multiple jurisdictions and study how they evolve along the way, as well as who molds them (Peck & Theodore, 2010; Temenos & McCann, 2013). Notably, this research tradition finds that policies are re-elaborated and re-invented as they spread, rather than circulating unchanged (Glick & Hays, 1991; McCann & Ward, 2013; Oliveira et al., 2023). Key techniques for qualitative studies include cross-case analysis, as well as process tracing. Cross-case analysis involves in-depth analysis of several systematically selected policies, often including deductive coding of policy contents. Process tracing is especially fruitful in illuminating causal relationships between policy adoptions (Starke, 2013).
Most often however, direct and unquestionable 'smoking-gun' evidence of diffusion is

[32] See e.g. Kammerer and Namhata (2018)

difficult to find and access in either research tradition, and studies instead rely on building convincing narratives of diffusion, and the pathways through which policies have spread in a given case (Starke, 2013). While qualitative methods potentially patch many of the holes of quantitative diffusion studies, they also suffer from the challenges faced by qualitative methods more generally. Namely, that human investigators are far from identical and thereby may arrive at different conclusions when presented with the same more-or-less ambiguous evidence, due to different knowledges and implicit biases (Creswell & Poth, 2016; Patton, 2014). Furthermore, humans get tired, make mistakes, or are otherwise not entirely consistent, and rarely possess sufficient language skills and technical knowledge to single-handedly analyse diverse policies from a heterogenous-language region such as Europe. All in all, qualitative processing of policy documents usually require several passes over any given textual source and rounds of consistency checks between multiple more-or-less independent researchers to arrive at accurate and reproducibly results (Saldana, 2021). This is a time- and resource-consuming process, and thus usually limits applications of these methods to small datasets.

Network-based diffusion studies have, more recently, entered the scene. Networks have been used to model diffusion networks based on ever-growing policy adoption datasets (Boehmke et al., 2020). Desmarais et al. (2015) use a network inference algorithm on an EHA-dataset spanning 187 policies to infer the underlying policy diffusion network connecting American states. Garrett and Jansa (2015) use a text-as-data approach to match documents based on similarity and explore diffusion networks. Several more recent studies elaborate on the text-as-data approach in a diffusion context (Abel & Mertens, 2023; Chalmers et al., 2024; Linder et al., 2018). Notably, text-as-data approaches often use document similarity as a proxy for diffusion, assuming that document similarity is indicative of diffusion, and thereby that similar documents are a result of causal influence by one document on the other. Similarly to adoption patterns in EHAs however, such similarities might be random or spurious. As the sub-field of Policy Mobilities has shown, policies often mutate as they diffuse, and as such the validity of the "similarity = diffusion" assumption can be questioned. Diffusion studies using text-as-data approaches view them as a potential method for probing the substantive contents of policies otherwise only achieved through qualitative studies, while avoiding the prohibitively expensive manual labour associated with reproducible qualitative investigation. Thereby these approaches potentially offer a way to marry the strengths of qualitative and quantitative approaches. The following sections will expand upon the text-as-data approach, its weaknesses, and how they might be overcome using recent advances in the computerised processing of text through *Natural Language Processing*.

A.2 Natural Language Processing

The basic barrier to performing computer-based processing of text is the fact that computers operate on *numbers*, while text is a symbolic representation of *language*. Human language embodies complexity extending far beyond simple strings of characters, but even translating just a single of the many written human language to numbers present several fundamental challenges. For instance, words are ambiguous and carry different meanings depending on the surrounding context, with nuances often conveyed by subtle word choices or shared understanding unstated in the text. Metaphors, idiomatic expressions and cultural differences further complicates interpretation. Moreover, most languages are constantly evolving, with new or updated meanings emerging regularly and diffusing between geographic contexts, which has to be accounted for when interpreting text.

One simple but common way to represent a text document as numbers is to simply count the number of occurrences of unique words. This is often referred to as a 'bag-of-words' approach, and allows for easily interpretable comparisons between texts. For example, texts might be considered similar if their word-distributions are similar. Such approaches have been used to study policy diffusion, but have several fundamental challenge (Abel & Mertens, 2023; Chalmers et al., 2024; Garrett & Jansa, 2015). For instance, representations for the sentences `"dog bites man"` and `"man bites dog"` would be identical. More generally, bag-of-words methods fail to account for context, paraphrasing, word ordering and document length. Essentially, bags-of-words do not carry much information about the substance or meaning present in the 'bagged' text, and do not identify or favour words carrying meaning or substance. Moreover, such methods cannot be applied to compare documents in different languages. Other, more advanced methods, such as 'bag-of-concepts', 'n-grams', and text-reuse approaches, similarly suffer from many, if not always all of these challenges (G. Lin, 2025).

Modern approaches for using computers to uncover the complex relationships between words in written language - the field of Natural Language Processing - all involve statistical models achieved through application of machine learning. While the term 'machine learning' often brings to mind robots, artificial intelligence and apocalyptic scenarios, Machine learning is an umbrella term for statistical methods that estimate parameters of a function through iterative prediction-and-updates, by minimizing the error when compared to known ('training') data. In essence, most modern methods are similar to estimating the slope and intercept of a linear function given a set of data points. Like in that exercise, one requires a certain amount of data points to provide a reliable estimate for each parameter but once estimated, the function, or 'model', can make predictions that generalise beyond the initial set of data points. *Language* models are extensions of these concepts, except the data points they 'learn' from are numerical representations of language. Given the complexity of language, effective models of language are also extremely complex. They are typically decidedly *non*-linear functions, but nonetheless operate on similar parameter-estimating principles, only involving an very large amount of parameters and thereby enormous amounts of training data. Fortunately, written human language is relatively common and the internet provides an abundant source of publicly available training data. Recent years have seen colossal progress in using machine learning models for language processing ('language models'), which has spurred much of the 'AI boom' and surrounding hype. The following sub-sections detail two such developments: first, Text Embeddings as a representation of text, and secondly, large language models as a processing tool for documents.

A.3 Embeddings

Embeddings allow for machine treatment and comparisons of words and text by creating numerical representations of text in the form of vectors. These vectors live in often high-dimensional vector spaces, where each dimension might be thought of as representing some trait, and the value of a given word or text's representation for some dimension will thereby correspond to how well the text embodies that particular trait (Elhage et al., 2022; Templeton et al., 2024). Effectively, one can think of the dimensions as a list of yes-or-no questions that might be answered regarding the contents of some text.[33] To illustrate how this works, consider the following example, where we represent text strings in a six-dimensional vector space, where each dimension corresponds to some trait that might be useful in describing fantasy villains. The dimensions, shown in Eq. 1, encode information about whether the text string corresponds to the name or description of a 'Dark Lord'; whether the subject of said string stores their soul in jewellery; is vanquished by a child; has a nose; resides on 'Earth'; and has intact limbs.
We can then 'embed' text strings in this space. For instance, in the case of the text `"Sauron"`, we might answer affirmatively to the first and second question, somewhat wavering regarding the third (depending on whether one considers Isildur or Frodo as the vanquisher of Sauron, and furthermore Hobbits as 'children' or 'child-like') and so on. Eq. 1 shows how three strings of text; `"Sauron"`, `"Voldemort"` and `"Darth Vader"`, each corresponding to the name of a fantasy villain, might be represented as vectors in this space. Here, a value of 1 indicates maximum presence of a trait, -1 indicates the complete absence of a trait, and values in between represent

[33] This is an embedding technique called 'question-answering (QA) embeddings' (Benara et al., 2024). It is worth stressing that this is *not* how the embedding models applied in this thesis are constructed, but it is a useful example.

partial presences, with 0 specifically indicating neutrality, inapplicability or unclarity surrounding a trait.

$$\begin{pmatrix} \text{Dark Lord} \\ \text{Jewellery-soul} \\ \text{Child-vanquished} \\ \text{Has nose} \\ \text{Lives on Earth} \\ \text{Hand cut off} \end{pmatrix} \quad \text{Sauron:} \begin{pmatrix} 1 \\ 1 \\ 0.2 \\ 0 \\ 0.5 \\ 1 \end{pmatrix} \quad \begin{matrix}\text{Volde-} \\ \text{mort}\end{matrix} : \begin{pmatrix} 1 \\ 0.43 \\ 1 \\ -1 \\ 1 \\ -1 \end{pmatrix} \quad \begin{matrix}\text{Darth} \\ \text{Vader}\end{matrix} : \begin{pmatrix} 0.75 \\ -1 \\ 1 \\ 1 \\ -1 \\ 1 \end{pmatrix} \tag{1}$$

In this example, all three vectors show a strong positive signal in the "Dark Lord" category (albeit Darth Vader slightly less than the other two, due to his redemption arc). Sauron has his soul stored entirely in The One Ring, and thereby gets a '1' in the second dimension, wheres Voldemort is left with $3/7 \approx 0.43$ as only three of his seven Horcruxes were pieces of jewellery. Darth Vaders soul is not at all stored in jewellery, so he gets a '-1'. Sauron's physical form is mostly described as a dark shadow or a large eye, so the presence of a nose is hard to evaluate and hence receives a '0' - Voldemort and Darth Vader are more straightforward in this regard, being on opposite ends of the spectrum.

Embedding Similarity. Aside from providing a compact and machine-friendly way to represent words or text, vector representations allow for several very useful operations. Especially pertinent to this thesis, vectors can be compared for similarity, and embeddings thereby provide an avenue for assessing how similar pairs of words or text snippets are to each other. Many different measures of similarity exist, but in this thesis we will be applying the most commonly used measure 'Cosine Similarity':[34]

$$\text{Cosine similarity}(\mathbf{A}, \mathbf{B}) = \cos\theta_{\mathbf{A},\mathbf{B}} = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|} \tag{2}$$

Cosine Similarity measures the angle θ between vectors $\mathbf{A}$ and $\mathbf{B}$ (specifically, the cosine of the angle) and thereby provides information on whether the vectors are aligned ($\cos\theta = 1$), orthogonal ($\cos\theta = 0$), opposite-pointing ($\cos\theta = -1$) or somewhere in between ($\cos\theta \in [-1, 1]$). We can compute this similarity for pairs of

[34] Henceforth, the term 'similarity' will be used interchangeably with 'Cosine similarity'

embedding vectors:

$$
\begin{aligned}
&\text{Cosine similarity}(\text{Sauron}, \text{Voldemort}) &&= 0.27\\
&\text{Cosine similarity}(\text{Sauron}, \text{Darth Vader}) &&= 0.11\\
&\text{Cosine similarity}(\text{Voldemort}, \text{Darth Vader}) &&= -0.31
\end{aligned}
$$

Here it is apparent that the representations of `"Sauron"` and `"Voldemort"` are most similar amongst the three villains and that `"Darth Vader"` is not very similar to the other two villains. Notably, cosine similarity does not provide information on *how* the strings of text differ, or in which dimensions, only that they are more or less aligned.

Toy-model limits. In principle, we might embed any string of text in this toy example vector space. For instance, embedding the string `"He, who must not be named"` (Hwmnbn), would yield a vector representation identical to that of `"Voldemort"`, while embedding `"Anakin Skywalker"` might yield a representation similar to, but with distinct differences from, that of `"Darth Vader"` (e.g. the "Dark Lord" dimension might be negative for Anakin). This also includes embedding strings somewhat or entirely unrelated to fantasy villains, such as `"Dog"`, `"Scooby-Doo"`:

$$
\text{Hwmnbn:}\begin{pmatrix}1\\0.43\\1\\-1\\1\\-1\end{pmatrix}\quad
\text{Anakin:}\begin{pmatrix}-0.75\\-1\\1\\1\\-1\\1\end{pmatrix}\quad
\text{Dog:}\begin{pmatrix}-1\\-1\\-1\\1\\1\\0\end{pmatrix}\quad
\begin{matrix}\text{Scooby}\\\text{-Doo}\end{matrix}:\begin{pmatrix}-1\\-1\\-1\\1\\1\\0\end{pmatrix}
$$

Clearly, the 6 dimensions representing villain-y traits do not do a very good job of encoding the information in all of these strings. For instance, the representations of `"Dog"` and `"Scooby-Doo"` are identical and do not capture the cartoonish, mystery-solving, stoner-culture nuances which separates Scooby-Doo from any other dog or indeed the concept of a dog. To be sure, a faithful Anakin representation would also differ from Vader's in more than just 'evilness', such as presence of hair or a robot suit. To capture such distinctions, we might include additional dimensions encoding information on whether the subject of the text string is something furry, drawn in a cartoon, voiced by James Earl Jones, etc.Notably however, we would have to introduce new dimensions to our vector space to accomplish this. This complexity only increases as the size, length and variation amongst text pieces we wish to embed increases. A good general embedding scheme might be able to distinguish between the different possible meanings of a polyseme and/or homonym like 'bank' (e.g. either a financial institution, a river bank, a verb, etc.) depending on the surrounding context. Meaningfully capturing information in strings such as `"Directive 2019/904 of the European Parliament and of the Council of 5 June 2019 on the reduction of the impact of certain plastic products on the environment"` requires entirely different dimensions and traits to what we have considered so far. Embedding the *contents* of the directive would require handling of even more information still. What should be apparent following our toy example is that **a)** the representation of a given words depends on the information (and opinions) available to whomever is assessing the value of vector entrances (i.e. the interpretation of those answering the yes-or-no questions), but should be consistent and reproducible for any given information-state - in other words, deterministic. **b)** the dimensions of our vector space are fairly arbitrary (we might as well have described our villains using by the colour of their clothes, whether their name contains a 'V' or the release-dates of work depicting them'). In particular, we might through an iterative process *update* and improve upon the 'questions' we ask, such that the vector representation best captures the similarities and differences between text in our data set.

Modern text embedding is done through statistical models that are trained to capture the context and meaning of different strings of text. Differences in training data and model design means the information available to different models varies, but they are each internally consistent and deterministic. These statistical models are high-dimensional (encoding information in vector spaces of several thousand dimensions), and dimensions rarely correspond cleanly to a single traits, which limits the direct interpretability of dimensions that we have observed above (Elhage et al., 2022; Opitz et al., 2025; Templeton et al., 2024). Still, the principles of the process are similar to that described above (Mikolov et al., 2013; Peters et al., 2018).[35] Currently, very capable general embedding models are commonly available, which allows direct comparison of text similarity despite differences in language, subject and context.[36]

[35]See, though, nn. 33.

[36]The reader might object, that even thousands of dimensions aren't sufficient to effectively capture, in individual dimensions, every separate meaning present in written human language. The meanings of individual words alone likely numbers at the very least in the millions (C. Lin & Ahrens, 2005). Embedding models likely circumvent this issue by embedding more features than there are dimensions. This involves not encoding information or traits cleanly in individual dimensions orthogonal to each other, but rather doing some very slight degree of 'mixing'. While the number of mutually orthogonal vectors in a space scales as the number of the dimensions, the number of *almost* mutually orthogonal vectors (e.g. vectors with mutual angles within 1 degree of orthogonal) scales *exponentially* with space dimensions. In other words, allowing slight 'mixing' means that the amount of meaning that can be crammed into a vector space increases very quickly as its dimensionality grows. Such mixing would

The direct applicability of embedding models for text comparison, especially amongst legal and policy documents, is however still limited by especially two challenges: First, embedding models are limited in the size of inputs, what is typically termed the 'context window' of the model. In short, embedding models have an upper limit on the character length of a given input string.[37] For the largest and most advances models, this limit corresponds to around 30000 characters (≈ 6000 words ≈ 12.5 Danish standard pages), which is far from enough to process longer policy and legal documents (Maggiori, 2023). A way to circumvent this challenge is to embed sub-divisions of a given document, e.g. individual chapters or paragraphs. However, using such an approach, the model cannot consider the full context of a given document when encoding the meaning of text. This may be especially problematic in legal documents, given the abundance of cross references to previous articles, paragraphs, etc.
The second major challenge for embedding models in the task at hand concerns the inclusion of irrelevant context or information in embeddings. For our purposes, we are interested in comparing the *substance* of policies and legal acts - rather than the specific language and form with which policies are implemented. However, legal and policy documents often contain numerous standardised formulations, boilerplate language, and country-specific peculiarities which will influence the embedding of a given text. For instance, Danish laws begin with the text

> "We, Frederik the Tenth, by the Grace of God King of Denmark, do hereby declare: The Parliament has passed and We, by Our consent, ratified the following law: (...)".

Inclusion of such country specifics in the text to embed would result in reduced similarity between a Danish law and a substantially identical law from another country

ordinarily introduce interference between meanings, and thereby quite some noise into the model. However, as long as any given input is quite meaning-sparse (i.e. any piece of text (especially one of limited length) only ever contains a small subset of all possible meanings in language), this noise is limited and can be filtered. For more on this, see e.g. Elhage et al. (2022), Sanderson (2024), and Yoder (2025).

[37]More generally, language models process text in the form of 'tokens', which are smaller units of text, acting as the vocabulary of the model. Any input is converted into a series of tokens (a process called 'tokenization'). Tokens might be individual words (implying a large 'vocabulary' of tokens - one for each word in any given language - but few tokens necessary to represent a sentence - one for each word in the sentence) or characters (implying a smaller token vocabulary - one unique token for each character in the alphabet of a given language - but more tokens required to build a sentence). Modern models encode using a 'sub-word tokenisation' (specifically, byte pair encodings; Maggiori, 2023; Sennrich et al., 2016), where tokens are common sub-elements of words (e.g. prefixes, suffixes, common word parts), intended to strike a balance between the number of unique tokens and the amount of tokens necessary to represent a sentence (Radford et al., 2019). For instance, the suffix 'ing' could be considered a token and can follow several different verbs in their base form (such as 'eat', 'jump' and 'sleep' - also all tokens) to conjugate these into progressive present form. On average, a token corresponds to approximately 4 characters in English, and the embedding models relevant to this thesis have context windows of up to 8191 tokens. 'Tokenization' is not to be confused with embedding, which encodes information and meaning into vectors - tokenisation simply represents text in a different vocabulary.

(and likely more so if that country is not a constitutional monarchy), while increasing similarity between substantially unrelated Danish laws, based entirely on their common origin. These parts of the text will serve to hide and dilute the substance, and will skew any comparison based directly on that text. This should not be considered a flaw in the embedding model, which is doing its job of embedding similar texts to similar vectors, but rather a result of including unwanted context or information in the text string to embed. In this sense the second challenge might be formulated as our inability to make embedding models consider the objectives of our specific study and the associated relevant context.[38]
Fortunately, a second major recent advance in Natural Language Processing provides us with tools to overcome both of these challenges. The following section will detail how Large Language Models can be used for unsupervised directed extraction of information from lengthy documents - information, which can in turn be embedded and compared.

A.4 Large Language Models (LLMs)

Large Language Models (LLMs) is the commonly used term for modern natural language processing machine learning models. In the present context we shall use the term specifically to refer to autoregressive text-processing models with *generative* capabilities: models that process text and generate text in response.[39] These models function by iteratively generating a probability distribution for the next word in a sequence, sampling that distribution, updating the sequence with the sampled word,[40] and repeating this process until a certain stop trigger is reached (Brown et al., 2020; Vaswani et al., 2017).[41] In this sense, LLMs function not much differently to the spell-check or autofill suggestions on any mobile phone, which stop short of generating text by only producing suggestions based on the most likely words of the output distribution. Where LLMs differ from other language models is in their complexity and abilities.[42]

[38]One can, of course, train oneself out of these issues, creating an embedding model that takes large input and learns to ignore such unwanted context, or somehow isolating the dimensions related to those meanings and disregarding them when performing similarity comparison. Given that we do not have the resources to train such a model, we opt for another route of inquiry.

[39]This use of the term excludes the previously discussed embedding models, which are also natural language processing machine learning models, but process text and generate high-dimensional vectors. We'll use the term 'Embedding Models' to cover these models.

[40]More accurately, LLMs process and predict *tokens*, see nn. 37. For most introductory purposes however, thinking of tokens as individual words makes the process significantly more intuitive. Typically, a stop trigger such as `"<end of text>"` is encoded as a specific token, and generation terminates when this token is appended to the sequence (Maggiori, 2023).

[41]For those interested in exactly how this process works, Grant Sanderson provides an excellent introduction in his YouTube series on Deep Learning.

[42]LLMs are often confused with and mistakenly referred to by the names of products applying them, such as chatbots (e.g. 'ChatGPT' or 'Gemini'). Such products are specific applications of

The 'largeness' of language models is not a well-defined measure, but is typically described by their parameter count. The modern LLMs used in this thesis are proprietary, and their model architectures are not public knowledge, but the number of parameters is in the trillions.[43] The number of parameters however, seems key to the abilities of modern LLMs. Notably, models seemingly exhibit emergent abilities when the number of parameters scale past a certain point (Fu, 2022; Wei et al., 2022).[44] These abilities include the ability for 'common sense' reasoning, but also more complex, multi-step reasoning and problem solving (Kojima et al., 2023; Wei et al., 2023).[45] Furthermore, model parameters store information,[46] and larger models are thereby able to reason with inherent knowledge, without resorting to external information retrieval or training on a specific knowledge base (Yu et al., 2023).
In combination, reasoning and knowledge allow models to perform language processing tasks which they were not explicitly trained on, due to a general understanding of language and context. In particular, statistical model predictions are generally only effective when faced with input data similarly distributed as the training base.[47] LLM abilities however generalise and are robust when faced with inputs that differ significantly from the distribution of its training base (Si et al., 2023). This implies a great potential to use language models as general text processing tools, without specifically training or fine-tuning a model for the task at hand. For instance, even LLMs that are now considered dated due to the rapid pace of development have shown excellent performance in e.g. sentiment analysis, information extraction, deductive coding, text summarisation, translation, and much more (Chew et al., 2023; Chung et al., 2022; Jiao et al., 2023). This opens for the application of machine-learning methods for textual analysis, where historically the intensive work of constructing and labelling a representative test-set for supervised machine learning has precluded application (Chew et al., 2023).
Language models can in summary perform competitively to humans on a variety of

the more general technology of LLMs, most often with hidden instructions and fine-tuned for e.g. chatbot interactions, to produce a specific type of response, both in format, content, style, language and tone. For that reason, responses from a chatbot to e.g. the prompts and documents presented in this thesis will often differ noticeably from the responses presented here, obtained by prompting the same underlying language model.

[43]GPT4 not public but rumored 1.8T (Patel & Wong, 2023); Llama 4 Behemoth is 2T (Meta, 2025); Gemini unknown but earlier versions rumored 1.6T (Islas, 2023).

[44]Note that there is some debate and scepticism around emergent properties, with some authors arguing that the appearance of 'emergence' regarding certain abilities are instead a result of improper metrics for evaluation or other factors - see Berti et al. (2025), Lu et al. (2024), and Schaeffer et al. (2023). Whether emergent or not, the abilities are present in modern LLMs.

[45]Model reasoning performance is typically evaluated on math and logic problems as these have easily verifiable solutions. However, indications are that emergent abilities extend far beyond these domains - see (Bubeck et al., 2023; Srivastava et al., 2023; Wei et al., 2022).

[46]Up to 2 bits of knowledge stored per parameter, with modern models thereby easily capable of storing e.g. all information in English Wikipedia and textbooks (Allen-Zhu & Li, 2024).

[47]In other words, training data has to be representative.

complex language-based tasks, based only on brief instructions in the form of a 'prompt'. Various techniques exist for prompting models effectively and increasing performance (P. Liu et al., 2021), and the prompt is most often the single-biggest, but also most easily improved, influence on model performance for a given task (Kojima et al., 2023; Wei et al., 2023; White et al., 2023). While training models on specific examples can improve results, it is often sufficient to supply just a few, or even no examples, in the prompt. Providing examples of desired outputs generally improves model performance, especially when faced with ambiguity or unclear tasks, where models can latch onto example formats and receive cues on e.g. labels and classifications (Brown et al., 2020; Min et al., 2022; Radford et al., 2019; Webson & Pavlick, 2022).

Context window length can be a limiting factor for using embedding models to process policy documents. LLMs, however, typically have longer context windows. The latest generation of models (as of writing) allow context windows of millions of characters, whereby all but the very longest texts can be processed without issues (Kavukcuoglu, 2025; Meta, 2025; OpenAI, 2025a). Historically, models have however shown trouble processing information from longer texts, and research suggests models preference information at the beginning and end of longer inputs when generating responses (An et al., 2024; H. Liu et al., 2025; N. F. Liu et al., 2023). It is not yet clear if this problem persists in the newest generation of models, which are increasingly optimised for longer contexts and needle-in-a-haystack information retrieval (Kavukcuoglu, 2025; Meta, 2025; OpenAI, 2025a).

Other potential issues with LLMs include the fact, that models are generally eager to follow instructions - sometimes to the point of over-eagerness where they confidently output wrong, untrue or fabricated information, rather than disobeying instructions or indicating that a task cannot be solved or a question cannot be answered given the information present.[48] One example is in the case of a classification task where no categories fit a given sample, but the model confidently places the sample in an ill-fitting category nonetheless; another is including information in summaries not present in the source (Kryściński et al., 2019; Maynez et al., 2020). This tendency to 'hallucinate' is a major issue for models that is seemingly not improved upon in more recent releases (OpenAI, 2025b).[49]

[48] For a review, see Ji et al. (2023)

[49] Hallucinations can be mitigated by grounding LLMs in external knowledge banks, but considering the motivation for applying LLMs in the first place, construction of such knowledge banks can require significant resources and manual labour, and might not be feasible for novel, multilingual datasets (Shuster et al., 2021). A rather more managable approach is detecting non-factual statements across multiple, slightly varied, processings of the same text (Manakul et al., 2023).

Related to this, the black-box nature of deep-learning methods, including LLMs, fundamentally limit interpretability of model outputs (Doshi-Velez & Kim, 2017; Templeton et al., 2024). Specifically, it is difficult to probe how the model arrives at any given output. The large number of parameters and their complex interrelationships makes it very difficult to determine what each part of the model is effectively doing, but investigation is moreover precluded for many models due to their proprietary nature. Instead, outputs can be probed by prompting the model to also return its reasoning and step-by-step thinking, as well as textual anchors that support its interpretation of text (Wei et al., 2023). Such approaches are at best simulations of actual behaviour, and at worst post-rationalisations of potentially hallucinated outputs (Turpin et al., 2023). This fundamentally means, that scepticism is required with respect to model outputs, and that outputs are ideally manually validated, either exhaustively or for larger datasets by sampling (Calderon & Reichart, 2025; Ribeiro et al., 2016).

Randomness and Temperature are central concepts in LLMs. As mentioned, language models fundamentally output probability distributions over words, which are sampled and updated iteratively to produce the final output. As such, LLMs are inherently stochastic models, which limits the reproducibility of results. However, most LLM developers provide options to control randomness through model configuration. The inherent randomness in language models is mostly determined by the 'Temperature' setting, which determines the flatness of the probability distribution from which the word a model outputs is sampled. Internally, the probability distributions are generated by the 'softmax' function:

$$\sigma(\mathbf{z})_i = \frac{e^{z_i/T}}{\sum_j e^{z_j/T}}, \tag{3}$$

which converts a vector $\mathbf{z}$ of weights (one for each word in the model vocabulary) to a vector of probabilities σ, which is sampled to find the next word in a sequence. Increasing the temperature (T), results in a flatter distribution, where low-weighted words are more likely to be sampled than at lower temperatures. In the extreme case of $T \to \infty$, this will result in a completely random string of words.[50] Conversely, reducing the temperature makes the largest values of $\mathbf{z}$ dominate the distribution, and thereby preferences the most likely words. Specifically, setting $T = 0$ implies that the most likely word is always selected, which provides a way to increase reproducibility.[51] Temperature, however, does not control *all* stochastic processes in the model, and as

[50] Model providers typically don't allow this, limiting temperature settings to the range $[0; 2]$

[51] This is a simplified explanation, as more generally models operate on tokens (see nn. 37 and nn. 41), and softmax layers are present throughout the model architecture. For details, Sanderson (2024) or Vaswani et al. (2017).

such $T = 0$ does not guarantee consistent output. Some providers provide the option to control the random seed of the underlying algorithms, which further increases reproducibility, but still variations can occur between runs.

Training data. Large parameter counts require large amounts of training data, typically in the range of one to several hundred words per parameter (Hoffmann et al., 2022; Kaplan et al., 2020; Sardana et al., 2025). For modern models, this implies several trillion words, and the growth rates of LLMs have led to estimates that training data requirements will soon be surpassing publicly available text content (Villalobos et al., 2024). This means that models are trained on, and more-or-less capable in, several hundred languages. Moreover, as a notable consequence of the exhaustive use of publicly available text, any modern model will almost certainly already have seen any given public policy document we might show it, and thereby have some 'baked-in' knowledge of related policies.[52] As with any other statistical model, language models reflect the biases of their training data. Given that models are mostly trained on web data, this includes an English language bias, which is likely reflected in the models language abilities (Dodge et al., 2021; Hu et al., 2025). Biases also include potentially racist, sexist, ideological, etc. positions that occur in data (Bender et al., 2021). Models may inadvertantly (and subtly) restate such learned positions as fact, or frame interpretations according to such biases.
Following all this, the reader should hopefully have be convinced that LLMs, at least in principle, provide a tool which can assist in overcoming the barriers to embedding-based comparison of documents, namely length and unwanted context considerations. LLMs can be used to write condensed summaries of long documents, and be specifically instructed to only include the context which pertains to the substance of a policy or regulation. Moreover, LLMs can be used to extract information from such documents on characteristics of the policies they implement, such as instruments implemented, targetted actors, regulatory type, etc. This amounts to deductive coding, where models potentially can extract features of documents at speeds and levels of cross-document and cross-language consistency unmatched by human coders (Chew et al., 2023). These characteristics can in turn be used to find patterns of common elements in similar policies, potentially enabling conclusions about policy characteristics that might promote policy diffusion, or more/less influential actors in international policy development. The next section will detail how, given the ability to compare texts for similarity, networks can be used to elucidate and explore the similarity relationships between documents.

[52]Models typically have a knowledge cut-off date corresponding to the latest data harvest in its training data. The cut-off date for this depends on model, but the models used in this thesis are trained on information up to June 2024.

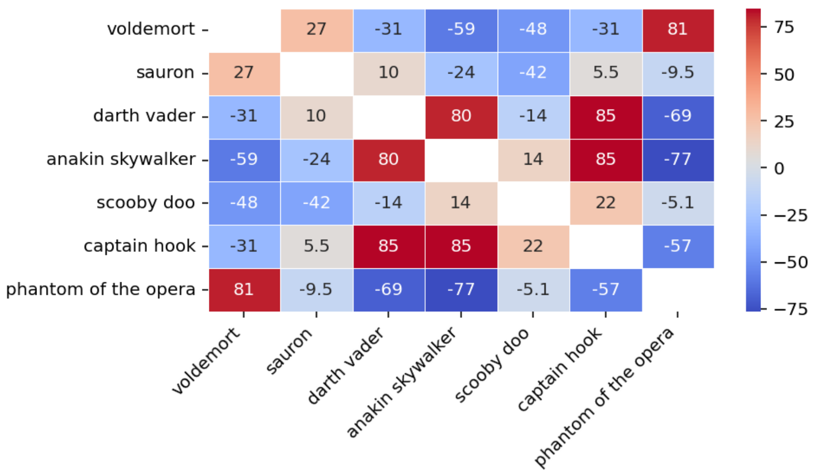


Figure 14: An example of a similarity matrix, using the toy embedding model example vectors. Diagonal (self-similarity) values have been hidden.

A.5 Network Analysis

Similarity, whatever exact measure one uses to calculate it, is a pairwise relationships between units of analysis (in this case textual documents). A common way to visualise similarities is through a similarity matrix, where rows and columns each correspond to every item in the dataset, and the pairwise similarities between vectors are entries in the corresponding row/column combinations. An example is given in Figure 14, which shows the cosine similarity matrix for the toy embeddings studied earlier (and a few additions).

As the number of sample pairs in a dataset grows quadratically with dataset size, visual representations through similarity matrices quickly become overwhelming and impractical except to read off all but the most superficial patterns in data (see e.g. Figure 9). Instead, one useful approach is to construct a network graph based on a similarity matrix. Here, the units of analysis are represented as nodes, and connections (also termed 'edges' or 'ties') are created between based on similarity. Connections are typically 'weighted' according to the pairwise similarity score. Networks can come in directed forms, where ties furthermore have a direction, which can correspond to e.g. physical, financial, information or influence flows.

If connections are drawn based on all similarity scores, the graph will be fully connected, albeit with varying weight of connections. In this case a large graph would be no easier to visualise than an equivalent similarity matrix. However, it is often fruitful to limit connections to only those edges which satisfy some combination of conditions and rules, such as exceeding a minimum similarity score, being amongst the n-highest scores for a given node, or satisfying some other outside conditions not directly related to similarity. These conditions will often represent the assumptions and expectations we have for the system of study, as well as limiting edges to specific aspects we wish to investigate (Cao, 2010). For instance, in a directed network studying policy diffusion, where edges represent influence between documents, we might establish causal rules such as only allowing a node to receive connections (and thereby influence) from nodes representing documents older than itself. Furthermore, we might only expect direct influence to be received from the most similar documents and/or for there to be some minimum threshold of similarity for us to consider it likely that diffusion has occurred.

Network Statistics. Restricting network connections simply amounts to filtering or masking out certain parts of the similarity matrix, and while a sparse network is easier to visualise, it is not the primary motivation for these operations. Rather, the field of network analytics has developed several measures of node influence, importance and centrality which become very useful once the network becomes more sparse. Notably, networks where connections only point forward in time do not have cycles, and thereby correspond to an important subset of networks called 'Directed Acyclic Graphs' (DAGs) for which the concepts of 'descendants' becomes meaningful. Several measures of node importance, useful in characterising those nodes which most likely exert or receive policy influence, have been devised for such networks (Newman, 2010; Wasserman & Faust, 1994).

- **Out-degree centrality** represents one of the simplest measures of node importance, simply counting the number of edges connecting outwards from a given node. In other words, this measures the number of direct descendants or 'children' of a node. Out-degree is commonly divided through by the total number of possible connections, to instead arrive at out-degree as the *fraction* of possible connections that are achieved.

  One of the disadvantages of using edge out-degree as a measure of node importance, is that it doesn't consider the importance of a node's descendants - i.e. if a node with a single outgoing connection to a dead-end node has the same importance as a node with a single outgoing connection to a high out-degree node.

- **Number of descendants**: Counting the number of descendants is another very simple measure of node importance, as having many descendants implies that one's influence has propagated to many others. As opposed to out-degree this measure does not differentiate in importance between nodes that are far-removed from their descendants and those that are closer. As an example, the node at the root of a 5-step chain of descendants is deemed as important as the node with 5 direct outgoing connections.

Clearly, neither of these simple measures fully match intuition of what node importance implies. We therefore turn to a more complicated measure, which effectively mixes considerations of immediate connections to nodes with considerations of descendant importance.

- **Katz centrality** is a more involved calculation commonly used in analysis of Directed Acyclic Graphs, that measures node importance based on summing the influence flowing through all other paths from any other node, but diminishing contributions along longer paths (Newman, 2010). When summing, the influence flowing through a connection is also commonly weighted by connection strength (e.g. document similarity). Katz centrality can be considered a generalisation of degree centrality, at it computes the number of nodes that connect to a given node directly or through other nodes, but attenuates the contributions from further away nodes by a factor $\alpha^k \in]0, 1[$, where k is the number of steps separating nodes. Nodes are furthermore given a baseline level of influence determined by the parameter β.

  For Directed Acyclic Graphs, these two parameters can be chosen freely, and represent an arbitrary but important methodological choice. Especially the parameter α is important, as it determines how many steps removed a node can be from its descendants, while still gaining significant 'credit' for their influence. As an example, for $\alpha = 0.25$ it takes 2 step for the aggregated weight given to a descendant to drop below 0.1 ($0.25^2 = 0.0625$); for $\alpha = 0.5$ it takes 3 steps ($0.5^3 = 0.0625$), and for $\alpha = 0.75$ it takes 9 ($0.75^8 \approx 0.1$). The value of β is less important, as we are rarely concerned with the actual value of the centrality measure, but more so the relative values between nodes as a measure of relative importance.

These measures represent node-level statistics - but can be aggregated based on node attributes (e.g. document metadata) to arrive at e.g. country-level measures of influence, whereby conclusions can be drawn as to which countries are more or less successful in influencing other countries' policy.
Furthermore, we can use the measures of importance to weight documents when studying policy characteristics such as the distribution of keywords, instrument types, etc. Hereby we can determine whether the distribution of policy characteristics differs between the most influential policies and the general population, and thereby whether some characteristics are more or less common among policies that diffuse.
Finally, we can use sparse networks to make some general observations regarding the structure of policy diffusions. Notably, we can measure the 'assortativity' of the network. Assortativity describes the tendency of nodes to connect to other nodes with similar attributes, and ranges in values from -1 (disassortative; nodes connect to nodes

with dissimilar attributes) to 1 (perfectly assortative) (Newman, 2010). As such, assortativity is measured with respect to some attribute, such as policy domain, country of origin, or specific policy aspects.

All in all, networks provide a useful tool for probing the contents and substance of diffusing policies, both individually and in the aggregate. With all this background in mind, we now move on to clarifying more precisely the methods we will use to construct and study Policy Diffusion networks.

B Prompt variations

This appendix contains the language model prompt used throughout this thesis, as well as example outputs from previous prompt versions.

B.1 Main prompt

The main prompt was the result of an iterative process of refinement and testing. The first few prompt versions were written manually, but were since improved using both Google's Gemini prompt generator and OpenAIs prompt design tools to include markdown formatting and more specific instructions. From then, the prompts were iteratively updated by hand to correct errors, improve performance or add functionality. Larger changes to the prompt or schema structures necessitated changes to the rest of the project codebase and as such were generally performed in batches, accompanied by internal consistency using the Google/OpenAI tools. As an example, the imprint left by these tools is noticeable in the final checks section of the prompt. This is generally a prompting technique reserved for reasoning models, that generate a hidden thought stream which can be corrected before outputting a response. The main model used in this thesis is not a reasoning model, but several of those tested were, including Gemini 2.5 which served a backup-role for erroneous documents, why this part of the prompt was kept.

Listing 1: The primary prompt (v3.2) used for model calls in the main body analysis.

```
Your primary task is to analyze legal texts and policy documents to summarize the policy's objectives and strategy in a **neutral, comparable, and concise manner** and identify **core policy instruments**. Your output **MUST** strictly adhere to the `PolicyAnalysis` JSON schema provided.

#CORE REQUIREMENTS & PHILOSOPHY:

1.  **CONCISENESS (Balanced):** Output must be concise, but fields like `anchor_quote` should prioritize readability for location verification over extreme brevity. Respect word limits.
2.  **INSTRUMENT FOCUS:** Identify the fundamental underlying **policy instruments** (e.g., Ban, Tax, Reporting Rule, Standard).
3.  **AGGRESSIVE & CONSISTENT GROUPING (MANDATORY):**
    *   You **MUST** group all instances or variations of the *same core policy instrument* under a single `PolicyInstrument` entry. Base grouping on the **fundamental mechanism** (e.g., imposing a binding rule, creating a financial incentive/disincentive, requiring information disclosure, promoting voluntary action).
    *   Do *not* create separate entries for applications to different products/substances if the *mechanism* is identical (e.g., one 'Substance Ban' entry covering multiple substances banned via the same article type).
```

```
    *   Do *not* group fundamentally *different types* of instruments (e.g., do
        not group a 'Tax' with a 'Reporting Obligation'). The goal is one entry per
        unique *type* of instrument used.
4.  **READABLE ANCHORING:** Use citations and `anchor_quotes` (~10-20 words
including the start of the relevant sentence/passage) to allow clear location
verification. They are anchors, not summaries.
5.  **DIRECT VALIDATION:** Justify `instrument_type` using a short, direct
`type_evidence_phrase` quoted from the text.
6.  **STRICT NEUTRALITY:** Summaries and descriptions **MUST NOT** contain
specific names or numbers of laws, directives, regulations, policies,
institutions, or countries of origin. Use generic terms ONLY (e.g., "the
legislation", "competent authority", "Member State", "economic operator").

#STEPS:

1.  **Identify and Translate Title and Date:** Extract and store the original
title (`document_title`) and date (`document_date`) of the document and translate
the title to english (`english_title`).
2.  **Identify Objectives:** Extract core policy objectives
(`policy_objectives`).
3.  **Summarize Policy:** Create a neutral summary (~500 words) of the overall
strategy, main instruments and target sectors/actors/industries/products/etc.
**Strictly avoid specific legislative/policy/institutional names/numbers**
(`policy_summary`).
4.  **Identify Policy Domain:** Identify the primary substantative issue area
which the policy is organised around (`policy_domain`) , one of the following:
'Agricultural and rural dev.', 'Air and atmosphere', 'Cultivated plants',
'Energy', 'Environment', 'Fisheries and aquaculture', 'Food and nutrition',
'Forestry', 'General', 'Land and soil', 'Livestock', 'Mineral resources', 'Sea',
'Waste and hazardous substances', 'Water', 'Wild species and ecosystems'.
5.  **Identify & Group Instruments:**
    *   Scan text for distinct **core policy instruments**. Apply grouping rules
        consistently (see Core Requirement #3).
    *   For each identified core instrument type:
        *   Create **ONE** `PolicyInstrument` object.
        *   **`instrument_title`**: Assign a concise, generic name (e.g., 'Market
            Restriction', 'Substance Ban', 'EPR Scheme', 'Reporting Obligation',
            'Financial Levy', 'Voluntary Guideline').
        *   **`instrument_type`**: Classify 'Regulatory', 'Economic', 'Soft', or
            `null`.
        *   **`type_evidence_phrase`**: Extract shortest key phrase (max 10
            words, exact quote) justifying the type (e.g., 'shall prohibit', 'fee
            shall be paid', 'promote awareness'). Null if type is null.
        *   **`summary_description`**: Write a **brief** (max 75 words) neutral
            description of the instrument's application *in this text*. If grouped,
            list key targets/areas. **Strictly avoid specific
            legislative/institutional names/numbers.**
```

```
        *   **`anchors`**: Find ALL relevant locations. For each unique `citation` string (e.g., '§ 1', 'Art. 5(a)'), include AT MOST ONE `SourceAnchor` in the list for this specific `PolicyInstrument`.** Do NOT repeat the same citation string multiple times within a single instrument's `anchors` list.
            *   For each **unique relevant citation**:
                *   Create **one** `SourceAnchor`.
                *   **`citation`**: Provide the unique citation string.
                *   **`anchor_quote`**: Extract a **concise quote anchor** (~10-20 words) including the **start of the relevant sentence/passage** after the citation. Use ellipses `(...)` **within the quote** if needed to skip less relevant initial words while preserving context and ensuring it allows clear location finding.
        *   **`target_actors`:** If clearly identifiable, list concise keywords for primary groups affected (e.g., `["Producers", "Importers"]`). Keep list brief.
        *   **`intervention_objects`:** If clearly identifiable, list concise keywords for primary objects of intervention, i.e. products and entities targeted by the policy (e.g., 'Wind Turbines', 'Plastic Waste', 'Migratory Birds', 'Data Infrastructure'). Keep list brief and group elements to achieve this.
        *   **`implementation_mechanisms`:** If clearly identifiable, list concise keywords for methods used (e.g., `["Permitting", "Reporting"]`). Keep list brief.
6.  **Format Output:** Assemble results into a single JSON object matching `PolicyAnalysis` schema. Ensure **strict adherence**.

#Examples
##EXAMPLE 1 `PolicyInstrument` (Illustrating Readability, Grouping, Neutrality, Optional Fields):**

```json
    {
      "instrument_title": "Extended Producer Responsibility Scheme",
      "instrument_type": "Regulatory",
      "summary_description": "Mandates financing by producers/importers for end-of-life management (take-back, handling, some cleanup) across various specified product streams, including packaging, electronics, and batteries.", // Generic description, no specific directive names
      "anchors": [
        {
          "citation": "Chapter 2 c § 9 h",
          // Slightly longer, more readable anchor including start of sentence
          "anchor_quote": "Producenter og importører af fiskeredskaber, der indeholder plast, skal (...) foranstalte tilbagetagning og særskilt håndtering (...)"
        },
```
```

```
        {
          "citation": "Chapter 2 c § 9 j",
          "anchor_quote": "Producenter og importører skal for egen regning foranstalte tilbagetagning samt særskilt håndtering af affald (...)"
        }
        // ... potentially more anchors ...
      ],
      "type_evidence_phrase": "skal (...) foranstalte", // Direct quote for obligation
      "target_actors": ["Producers", "Importers"], // Optional: Concise list
      "intervention_objects": ["Packaging", "Electronics", "Batteries"], // Optional: Concise list
      "implementation_mechanisms": ["Take-back Systems", "Separate Handling", "Financing Obligation"] // Optional: Concise list
    }

##EXAMPLE 2 SourceAnchor (Illustrating Readable Anchor Quote):

Original Passage Start: "Article 5(1): The competent authority shall, by 31 December 2025, establish and maintain a public database containing information reported under Article 7 related to...".

citation: "Article 5(1)",
anchor_quote: "The competent authority shall (...) establish and maintain a public database containing information reported under Article 7 (...)" (~18 words).

#FINAL CHECKS:

*Is the output valid JSON matching PolicyAnalysis schema?
*Are policy_summary, summary_description free of specific law/directive names/numbers?
*Are anchor_quotes readable anchors (~10-20 words, start of sentence/passage)?
*Is type_evidence_phrase a short, direct quote?
*Is grouping applied consistently based on the core instrument mechanism?
*Are optional fields (target_actors, intervention_objects, implementation_mechanisms) used concisely with keywords only, if present?
*Are word/sentence limits respected?
*Is each `citation` string unique within any single `PolicyInstrument`'s `anchors` list (NO duplicate citations per instrument)?
```

Listing 2: The python code for the Pydantic Schema passed alongside the prompt, which includes descriptions inspired by Burstein (1991) and Dreyer et al. (2024)

```python
from pydantic import BaseModel, Field
from typing import Literal, List, Optional

class SourceAnchor(BaseModel):
    """Provides a readable anchor point back to the source text."""
    citation: str = Field(..., description="Citation in the policy text (e.g., '§3', 'Art. 5(1)(a)').")

    anchor_quote: str = Field(..., description="**Concise quote anchor** (~10-20 words) including the **beginning of the relevant sentence/passage** following the citation marker. Use '(...)' for omissions and to skip less relevant initial words while preserving context. MUST uniquely identify passage with citation for location.")

class PolicyInstrument(BaseModel):
    """Represents a single core policy instrument, potentially grouping multiple applications."""
    instrument_title: str = Field(..., description="Concise, generic name of the core policy instrument type (e.g., 'Market Restriction', 'Extended Producer Responsibility Scheme', 'Reporting Obligation', 'Financial Incentive', 'Information Provision').")

    summary_description: str = Field(..., description="**Brief** (max 75 words) neutral description of how this instrument is applied within this text. If grouped, mention main areas/targets. **DO NOT include specific law/directive names or numbers.** Use generic terms." )

    instrument_type: Optional[Literal["Regulatory", "Economic", "Soft"]] = Field(None, description="Classification: 'Regulatory (establishing laws and regulations), 'Economic' (leveraging monetary incentives and disincentives), 'Soft' (based on voluntary, non-binding compliance), or null if unclassifiable.") # Inspired by definition in Dreyer et al (2024)

    type_evidence_phrase: str = Field(None, description="The **specific shortest key phrase/term** (max 10 words, exact quote) from the source text that MOST clearly indicates the assigned `instrument_type` (e.g., 'shall prohibit', 'fee shall be paid', 'promote awareness'), or null if type is null.")

    target_actors: Optional[List[str]] = Field(None, description="List of **concise keywords** for primary groups targetted by the instrument(e.g., 'Producers', 'Consumers', 'Operators'). Optional.")

```

```
    intervention_objects: Optional[List[str]] = Field(None, description="List of **concise keywords** for primary products, entities affected by the instrument or objects of intervention (e.g., 'Wind Turbines', 'Plastic Waste', 'Migratory Birds', 'Data Infrastructure'). Optional.")

    implementation_mechanisms: Optional[List[str]] = Field(None,
        description="List of **concise keywords** for methods used (e.g., 'Permitting', 'Reporting', 'Audits', 'Funding'). Optional.")

    anchors: List[SourceAnchor] = Field(..., description="List of source anchors showing ALL locations where this core instrument is defined or applied.")

class PolicyAnalysis(BaseModel):
    document_title: str = Field(..., description = "Exact original-language title of the document.")

    english_title: str = Field(..., description = "A translation to english of the original language title.")

    document_date: str = Field(..., description = "The date of the document in format DD/MM/YYYY. If the day or month cannot be determined, specify '01' (e.g. 'Mar 2020' -> 01/03/2020). If the year cannot be determined, return 'UNKNOWN'.")

    policy_objectives: str = Field(..., description="A brief, origin-neutral describtion of the core policy objectives. **DO NOT include specific names or numbers for legal texts or policies.**")

    policy_summary: str = Field(..., description="Origin neutral summary of overall strategy, main instruments and target sectors/actors/industries/products/etc. **DO NOT include specific names or numbers for legal texts or policies.**")

    policy_domain: Literal[
        'Agricultural and rural development', 'Air and atmosphere', 'Cultivated plants', 'Energy', 'Environment', 'Fisheries and aquaculture', 'Food and nutrition', 'Forestry', 'General', 'Land and soil', 'Livestock', 'Mineral resources', 'Sea', 'Waste and hazardous substances', 'Water',
        'Wild species and ecosystems'] = Field(..., description = 'The primary substantative issue area which the policy is organised around.') # Inspired by formulation of Burnstein (1991)

    instruments: List[PolicyInstrument] = Field(...,  description="List of the core policy instruments identified, grouping related applications.")
```

B.2 Previous prompt examples

Prompt version 3.2 is almost identical to prompt version 3.1, and thus provides a good example of what happens when we perform small perturbations in the prompt. The two differ mainly by prompt 3.2 introducing some metadata and classification fields, i.e. the fields extracting the document title, date, English title, and classify according to FAO domain. They also swap around some ordering between fields and have slight wording differences, but nothing substantial. Overall the structuring is the same, and specifically the instrument fields and instructions regarding them do not differ between the two. Here we list the plots generated using prompt 3.1, to be compared to those of Section 4.1. We only describe them very briefly in their captions, assuming the reader has knowledge of the main analysis. We also show the instruments found in the Nature Protection Act and the summaries and descriptions for the same documents as in Table 1.

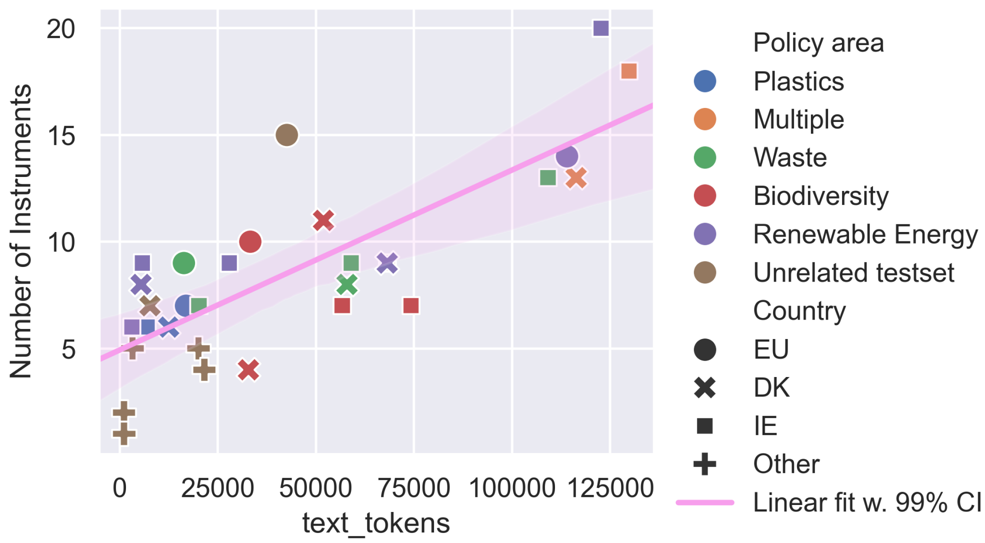


Figure 15: Prompt 3.1 Instruments vs Token length scatter - compare Fig. 3. We see the same general increase in instruments with documents length. The Nature Protection Act is not an outlier for this prompt version, but notably the amount of instruments found in several other documents increases and the slope of the curve is slightly higher.

Table 11: Prompt 3.1 Model-generated objectives and summaries. Descriptions are very similar to the Prompt 3.2 ones, except for the bottom two documents.

Policy	Objectives	Summary
EU Single-use plastics directive	The directive aims to reduce the impact of certain plastic products on the environment and human health, particularly in the aquatic environment, and to promote a transition to a circular economy.	This directive addresses the environmental impact of single-use plastics, oxo-degradable plastics, and fishing gear containing plastic. It employs a combination of measures, including consumption reduction targets, market restrictions, product design requirements, marking requirements, and extended producer responsibility schemes. The directive targets producers, consumers, and Member States, focusing on reducing plastic waste and promoting sustainable alternatives. It also sets separate collection targets for single-use plastic beverage bottles and mandates awareness-raising measures to encourage responsible consumer behavior. The directive includes provisions for monitoring, reporting, and evaluation to ensure its effective implementation and to inform future policy adjustments.
DK Miljøbeskyttelsesloven	The law aims to protect nature and the environment, promote sustainable development, prevent and combat pollution, establish hygiene standards, limit resource use, promote cleaner technology, and encourage recycling and waste management.	This legislation outlines a comprehensive framework for environmental protection. It covers a wide range of activities, including pollution control from various sources, waste management, and the use of resources. The core strategy involves setting environmental quality standards, implementing permitting and regulatory schemes for polluting activities, promoting cleaner technologies and waste reduction, establishing extended producer responsibility for certain products, and ensuring public access to environmental information. The legislation targets a broad range of actors, including businesses, producers, importers, public authorities, and individual citizens, and it addresses various environmental media, including air, water, soil, and the subsurface. It also establishes monitoring and enforcement mechanisms to ensure compliance and provides for penalties for violations.
IE Waste Management Act	The act aims to prevent, manage, and control waste, give effect to European Community acts regarding waste management, and promote waste recovery and reduction.	The legislation provides a framework for waste management, emphasizing waste reduction, recovery, and environmentally sound disposal. It establishes waste management planning processes, including hazardous waste management. The legislation outlines duties for waste holders, collection services, and facilities, and introduces permitting and licensing requirements. It enables measures to prevent waste production, promote waste recovery, and ensure public authority compliance. The legislation also covers monitoring, inspection, and enforcement provisions, including penalties for offences. Furthermore, it addresses specific waste streams like end-of-life vehicles and electrical and electronic equipment.
DK Habitatbekendtgørelsen	The policy objectives are to designate and administer international nature conservation areas, protect certain species, and ensure the preservation and restoration of favorable conservation status for species and habitats within these areas.	This legislation designates international nature conservation areas, including bird protection areas, habitat areas and Ramsar areas, and establishes rules for their administration. It sets binding regulations for authorities regarding planning and administration that affect these areas. The legislation outlines specific considerations for authorities, including assessing the impact of projects on Natura 2000 areas, promoting the protection of Ramsar areas, and adhering to prohibitions against damaging breeding and resting areas for animal species and plant species. It includes provisions for granting permissions, dispensations, and approvals, as well as for potential deviations from protective measures under specific circumstances, such as overriding public interests. The legislation also addresses the general protection of certain species, including assessments of project impacts on species and habitats, and allows for deviations under specific conditions, such as protecting public health or preventing damage to property.
IE Renewable Energy Regulations 2020	The regulations aim to implement articles of a European Union directive regarding the promotion of energy from renewable sources. The primary objective is to establish a framework for support schemes that encourage the integration of renewable energy into the energy market.	These regulations establish a framework for support schemes designed to promote the use of energy from renewable sources. The regulations outline the application of support schemes, emphasizing the integration of electricity from renewable sources into the market. These schemes should be responsive to market conditions, avoid distortions, consider system integration costs and grid stability. Support is granted through open, transparent, competitive, and non-discriminatory processes, though exemptions may be provided for small-scale installations or demonstration projects. The regulations allow for competitive tendering processes to be limited to specific technologies under certain conditions. The regulations also address the stability of financial support for renewable energy projects, ensuring that the terms and conditions are not revised in a way that undermines the economic viability of supported projects. A long-term schedule for the allocation of support is to be published and updated annually, and regular assessments of the effectiveness of support schemes are to be carried out.
Fiction chapter - BlackCompanyVol1Chap4	The primary objectives are to gather intelligence, eliminate key enemy figures, and disrupt enemy operations to maintain control and stability in the region.	The policy revolves around military actions and intelligence gathering to undermine the enemy's capabilities. Key strategies include identifying and eliminating high-value targets, disrupting enemy plans, and gathering information to anticipate and counter threats. The policy employs a combination of direct military intervention, covert operations, and intelligence networks to achieve its objectives. The main actors involved are military units, intelligence operatives, and political leaders. The policy targets enemy forces, leaders, and infrastructure, aiming to weaken their ability to wage war and maintain control. The policy also involves managing internal dissent and maintaining morale within the ranks through propaganda and control of information.
News - Sports	The policy objective is to win the green jersey in the cycling race.	The cycling team strategically employed repeated attacks and strong climbing to secure the green jersey in the race. A cyclist launched multiple attacks, creating breakaways and securing points, ultimately winning the jersey despite challenges in the final stage.

Table 12: Prompt 3.1 Instruments found in the Danish Nature Protection Act. Note the much more reasonably described area restrictions and inclusion of protected area designation as an instrument, as well as Nature National Parks.

	Instrument Title	**Description**	**Type**	**Target actors**	**Intervention Objects**	**Implementation Mechanisms**
1	Area Restriction	Imposes restrictions on activities such as construction, planting, and terrain modification within defined buffer zones around various natural features, including coastlines, lakes, watercourses, forests, and historical monuments.	Regulatory	'Landowners', 'Developers'	'Coastal areas', 'Lakes', 'Watercourses', 'Forests', 'Historical monuments']	'Zoning', 'Permitting'
2	Protected Area Designation	Allows for the designation of specific land and water areas as protected zones, with the aim of preserving natural habitats, biodiversity, and landscape values. Includes provisions for managing activities within these areas.	Regulatory	'Landowners', 'Local authorities'	'Lakes', 'Watercourses', 'Coastal areas', 'Klitfredede areas'	'Zoning', 'Management plans'
3	Activity Notification Obligation	Requires notification to the local authority before undertaking certain activities within international nature conservation areas, allowing for assessment of potential impacts on conservation objectives.	Regulatory	'Landowners', 'Operators'	'International nature conservation areas'	'Notification', 'Impact assessment']
4	Species Protection	Prohibits the intentional disturbance of animal species listed in the law, protects their breeding and resting areas, and regulates the exploitation of wild flora and fauna.	Regulatory	'General public', 'Landowners'	'Protected animal species', 'Plant species'	'Prohibitions', 'Regulations'
5	Land Management Agreement	Enables local authorities to enter into agreements with landowners or users to manage land within international nature conservation areas to achieve Natura 2000 plan objectives.	Regulatory	'Landowners', 'Local authorities'	'International nature conservation areas'	'Agreements', 'Land management'
6	Land Management Order	Allows local authorities to mandate specific land management practices on properties within or outside international nature conservation areas to achieve Natura 2000 plan objectives, if agreements cannot be reached.	Regulatory	'Landowners', 'Local authorities'	'International nature conservation areas'	'Orders', 'Land management'
7	Financial Support	Provides grants and loans to municipalities, associations, and private landowners for land acquisition, nature conservation, habitat restoration, and improving public access to nature.	Economic	'Municipalities', 'Associations', 'Landowners']	'Natural areas', 'Public access'	'Grants', 'Loans'
8	Sand Drift Management	Enables the Ministry to implement measures to control sand drift in protected areas and to order landowners to take necessary measures on their property.	Regulatory	'Landowners', 'Municipalities'	'Klitfredede arealer'	'Orders', 'Direct intervention'
9	Public Access Restriction	Allows for restricting public access to certain areas to protect vulnerable natural environments or ensure safety during specific activities like hunting or forestry work.	Regulatory	'General public'	'Natural areas'	'Signage', 'Enforcement'
10	Advertising Restriction	Prohibits the placement of advertising and propaganda materials in open landscapes, with certain exceptions for business advertisements and informational signs.	Regulatory	'Businesses', 'Advertisers'	'Open landscapes'	'Enforcement', 'Permitting'
11	Nature National Park Establishment	Allows for the establishment of nature national parks on state-owned land, managed primarily for nature and biodiversity conservation, free from forestry and agricultural production.	Regulatory	'State landowners', 'Park managers'	'State-owned land'	'Permitting', 'Management plans'

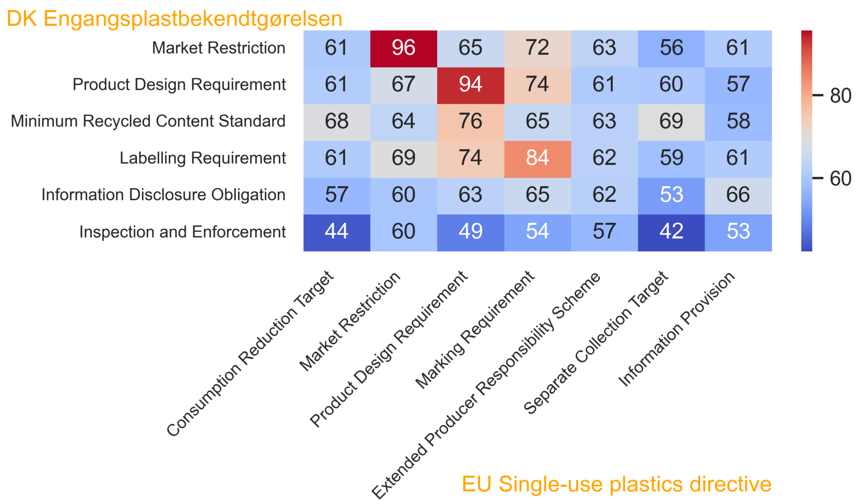


Figure 16: Prompt 3.1 Instrument-comparison matrix between the European Single-Use Plastics Directive and the Danish *Engangsplastbekendtgørelse*. Compared to fig 19, an extra instrument is found in the Danish text (Minimum recycled content standard). The best matching pairs have higher scores, indicating better matching descriptions of these (almost identical in fact). Outside of that the matrices are very similar.

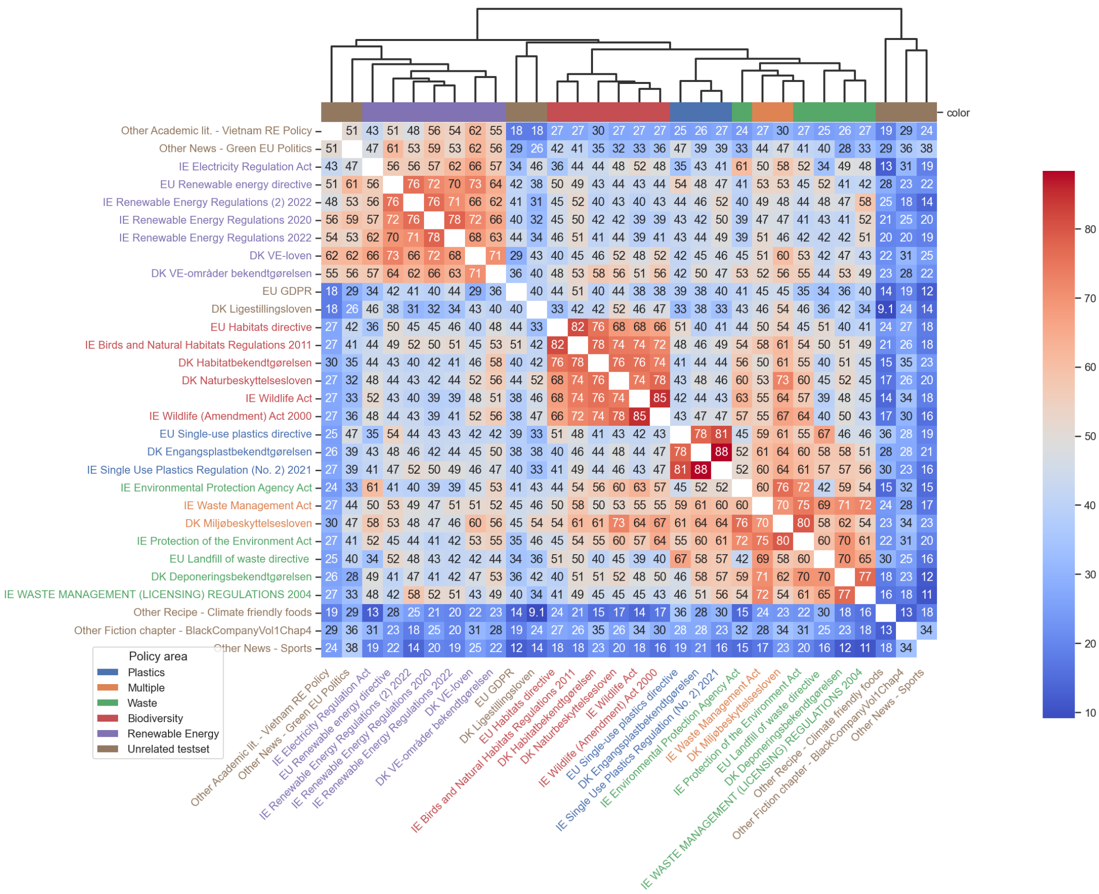


Figure 17: Prompt 3.1 Similarity matrix of document embeddings for the Directives Test-set. Note that the ordering of labels has changed compared to the main analysis. Small scale variations compared to Fig 4, but large scale structure and clusters remains the same. The RE cluster shows somewhat higher similarities in Prompt 3.1.

C Large Language Model variation

This appendix tables comparing variations in outputs of different language models to the same prompt. It compares Google's Gemini 2.0 Flash, Gemini 2.5 Flash and OpenAI's GPT 4.1 Mini and o4 Mini. Gemini 2.5 and o4 Mini are both 'reasoning' model, that produce a thought stream of their response beforehand, and are capable of self checking this response before their final answer. These models are more 'intelligent', but show similar abilities in the summarisation task here. Table 14 and ?? shows example outputs from the four models for 2 texts in the Directives testset. Table 15 shows the instruments found for the Danish Nature Protection Act, for each model. Figure 18 shows Principal Component Analysis (PCA) of the Embedding vectors for the 30 texts, for each model choice. PCA is a technique for visualising high-dimensional vectors in 2D, by projecting them onto the plane that best explains the variance of the dataset. Note the very similar structures and placements of points (the vector dimensions are arbitrary up to a sign, so can be flipped across both axis - in this light all four figures are roughly identical). This implies the embedding vectors are very similar across all four treatments - and thereby implies robustness towards model choice for both embeddings and the analysis of them described in Section 4.1. Thereby, aside from these examples, the clusters and networks resulting from the similarity matrix are more-or-less identical for all four models, and all four show the same skewing in the histogram of similarity values.

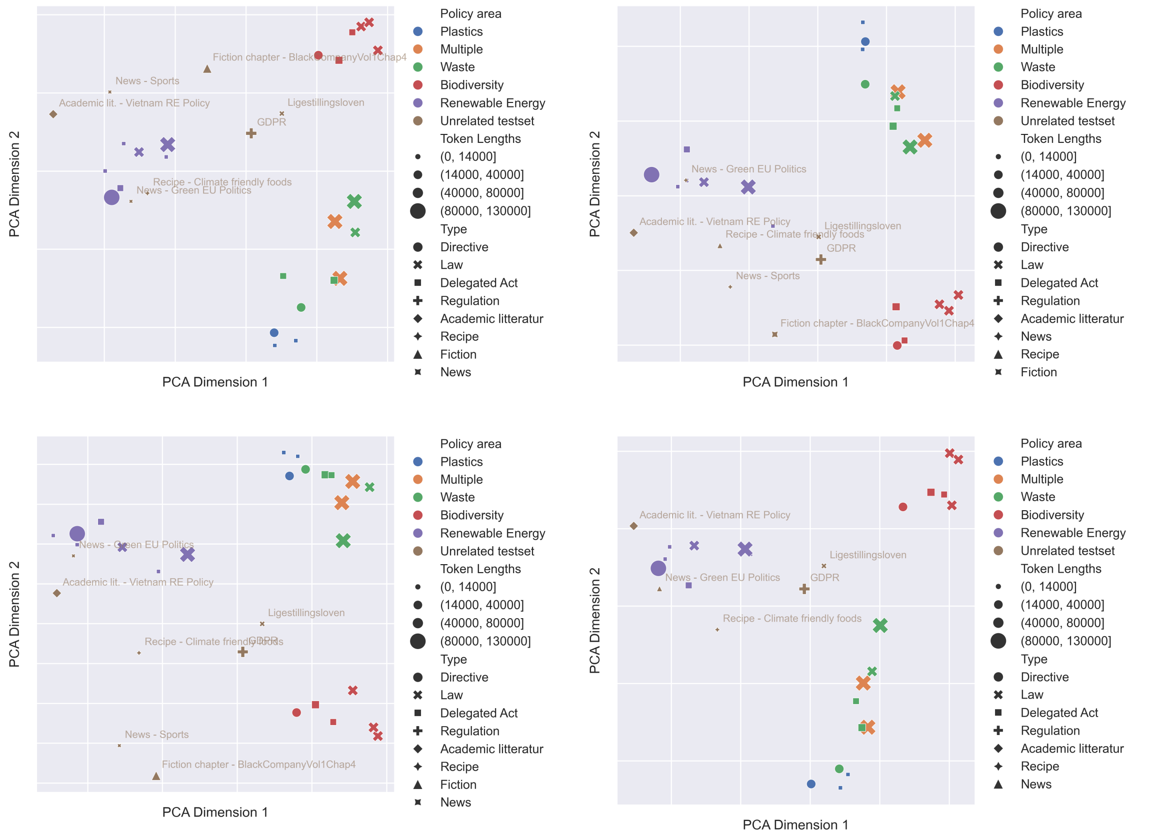


Figure 18: Comparison of Principal Component Analysis for the four different model choices. The structure for all four models are very similar and show the same large scale patterns.

Table 13: Model comparison of objectives and summary for the Vegetarian recipe

Model	Objectives	Summary
2.0 Flash	The recipe aims to provide a vegetarian/vegan alternative to a traditional Greek dish, while also considering the carbon footprint of the ingredients.	The document provides a recipe for a vegetarian version of a traditional Greek dish. It details the ingredients and steps for preparing the dish, offering variations for vegetables and herbs. The recipe also includes an estimated carbon footprint analysis, comparing the impact of different ingredients, such as cheese and vegetables, on the overall emissions of the dish. The analysis aims to inform consumers about the environmental impact of their food choices.
2.5 Flash	This document is a recipe and does not contain policy objectives.	This document provides a recipe for vegetarian giouvetsi, including ingredients, preparation steps, and an estimated carbon footprint for the dish. It compares the carbon footprint to other recipes on the same blog and highlights the relative impact of different ingredients like cheese and vegetables.
4.1 Mini	To provide a recipe for a vegetarian version of a traditional oven-baked dish, emphasizing the use of vegetables and plant-based ingredients to reduce carbon footprint and promote sustainable cooking.	The document presents a vegetarian adaptation of a traditional oven dish originally made with meat, substituting it with a variety of vegetables and rice noodles. The recipe includes detailed preparation steps, ingredient lists, and cooking instructions designed to maintain the dish's flavor and texture while reducing environmental impact. It highlights the carbon footprint of the dish, noting that the cheese topping contributes significantly to emissions, whereas vegetables contribute less relative to their weight. The recipe encourages flexibility in vegetable and herb choices, promoting sustainable and climate-friendly cooking practices.
o4 Mini	Present a vegetarian version of a traditional dish and estimate its carbon footprint.	This document provides a detailed recipe for preparing vegetarian giouvetsi, including ingredients, step-by-step cooking instructions, suggestions for variation, and an estimated carbon footprint per two portions. It also compares the impact of key ingredients—highlighting the substantial share from cheese versus vegetables—and ranks the recipe among others on the site based on its estimated emissions.

Table 14: Model comparison of objectives and summary for the Danish Nature Protection Act.

Model	Objectives	Summary
Gemini 2.0 Flash	The law aims to protect the country's nature and environment, enabling sustainable societal development while respecting human living conditions and preserving animal and plant life. It focuses on protecting nature, improving habitats, and enhancing public access to nature.	The legislation outlines measures for nature protection, including general protective provisions for natural sites such as lakes, watercourses, heaths and coastal areas. It restricts activities like spraying, fertilization and modification of these areas. Specific protections are in place for dune areas, coastal zones, lakes, rivers and forests, limiting construction and other changes. International nature conservation areas are subject to special regulations and management plans. The legislation also addresses public access to nature, regulating activities like traffic, camping and waste disposal. It includes provisions for protecting plant and animal species, nature management, and financial support for nature conservation efforts. The law establishes administrative structures and processes for conservation, including designating protected areas and handling violations.
Gemini 2.5 Flash	To protect the country's nature and environment to ensure sustainable societal development, preserving wildlife, plants, their habitats, and landscape values. It also aims to improve, restore, or create areas important for wildlife, plants, and landscape interests, and to provide public access to nature and improve opportunities for outdoor recreation.	The legislation establishes a framework for protecting diverse natural areas, including lakes, streams, heaths, bogs, coastal meadows, swamps, wet meadows, dry grasslands, dune areas, coastal zones, areas near lakes and streams, forests, and ancient monuments, by prohibiting alterations to their state or restricting specific activities within defined protection lines. It also includes provisions for the protection of wild animal and plant species, prohibiting disturbance and damage to breeding/resting sites, and regulating the introduction of non-native species. The framework supports the implementation of international conservation obligations, including measures within designated international nature protection areas, through management plans, agreements, and mandatory actions. Financial instruments such as loans and subsidies are available for nature conservation, restoration, and public access initiatives. The legislation also covers land acquisition through pre-emption rights and expropriation for conservation purposes, sand drift control, nature monitoring, and procedures for handling environmental damage. Administration and enforcement involve local and central authorities, with provisions for appeals and penalties for non-compliance. Specific rules govern public access to beaches, forests, uncultivated areas, roads, and paths, balancing public recreation with protection needs and landowner interests. The establishment and management of nature parks on state-owned land are also regulated.
GPT 4.1 Mini	To protect and preserve natural habitats, wild flora and fauna, and cultural and landscape values; to restore and improve areas important for wildlife and cultural interests; to ensure public access to nature and promote outdoor recreation; to combat sand drift, increase forest areas, and restore wetlands and river valleys to improve water environments.	The legislation establishes a comprehensive framework for nature protection, focusing on conserving natural habitats, wild species, and cultural landscapes. It prohibits alterations to protected natural areas such as lakes, streams, heaths, bogs, and coastal zones, with exceptions for maintenance and specific authorized activities. It regulates public access to natural areas, including forests, beaches, and dunes, balancing recreation with conservation. The law mandates protection and management of species, including prohibitions on disturbance and provisions for conservation plans. It sets up procedures for designating protected areas, including international nature protection zones, and mechanisms for agreements or compulsory measures to achieve conservation goals. The law empowers authorities to impose restrictions, grant dispensations, and enforce compliance through monitoring and penalties. It includes provisions for acquisition, compensation, and expropriation to facilitate conservation projects. The legislation also governs advertising in rural areas, waste disposal, and navigation restrictions to protect natural environments. Administrative structures, including specialized committees and appeals processes, are established to oversee implementation. The law incorporates digital communication and public involvement in decision-making. Overall, it employs regulatory instruments such as bans, prohibitions, permitting, compulsory measures, and enforcement mechanisms, targeting landowners, public authorities, businesses, and the general public to safeguard biodiversity, landscape, and public enjoyment of nature.
o4 Mini	Protect and conserve wild flora and fauna, their habitats and landscape, restore and create areas of ecological importance, ensure sustainable development, public access to nature, combat sand drift, expand forests and wetlands to improve the water environment.	This law establishes a comprehensive framework for nature protection through direct prohibitions on altering designated habitats (lakes, watercourses, heaths, bogs, dunes, shorelines, forests, heritage and church zones), mandates public access rights, and empowers authorities to grant permits or dispensations under strict conditions. It obliges advance notification of activities in international nature sites, enables municipalities to impose conservation orders where voluntary agreement fails, and requires immediate reporting and remediation of environmental damage. The minister may define habitat types, set measurement, registration and exemption rules, and issue delegated regulations for implementation and public involvement. Financial instruments include compensation to landowners for restrictions and grants or loans for land acquisition, habitat restoration and management by public bodies, NGOs and private owners. Enforcement is secured through fines, criminal sanctions and confiscation. The law also regulates exploitation of wild species via authorisation schemes, provides for land management orders to control sand drift and habitat maintenance, and prescribes public notification and consultation procedures for conservation proposals and decisions.

Table 15: Titles of model-extracted instruments for the Danish Nature Protection Act

Model	Gemini 2.0 Flash	Gemini 2.5 Flash	GPT 4.1 Mini	o4 Mini
1	Area Restriction	Land Use Alteration Restrictions	Prohibitions on Alterations to Protected Natural Areas	Habitat Modification Prohibition
2	Activity Restriction	Activity Restrictions	Prohibitions on Use of Chemicals and Land Management Practices in Protected Areas	Ministerial Delegated Rulemaking
3	Area Restriction_2	Public Access Rules	Restrictions on Development and Land Use in Protected Coastal and Dune Areas	Activity Notification Requirement
4	Land Use Prohibition	Species and Habitat Protection	Regulation of Public Access to Natural Areas	Natura 2000 Conservation Orders
5	Area Restriction_3	Conservation Planning and Implementation	Protection and Management of Wild Species	Financial Support and Incentives
6	Building Restriction	Nature Conservation Board Decisions (Fredning)	Designation and Management of Protected Areas and International Nature Protection Zones	Fines and Criminal Sanctions
7	Building Restriction_2	Financial Support for Nature Management	Compulsory Nature Management and Sand Drift Control Measures	Public Access Regulation
8	Area Restriction_4	Land Acquisition (Preemption and Expropriation)	Acquisition, Compensation, and Expropriation for Conservation Purposes	Mandatory Land Management Orders
9	Building Restriction_3	Sand Drift Control	Regulation of Advertising in Rural Areas	Dispensation and Exemption Permits
10	Activity Notification	Nature Monitoring	Monitoring, Enforcement, and Penalties	Public Notification and Consultation
11	Land Use Agreement	Environmental Damage Remediation		Environmental Damage Reporting and Remediation
12	Land Use Mandate	Fee Collection		Wild Species Exploitation Authorisation
13	Public Facility Planning Mandate	Establishment and Management of Nature Parks		
14	Advertising Restriction			
15	Waste Disposal Prohibition			
16	Species Disturbance Prohibition			
17	Habitat Damage Prohibition			
18	Species Management Plan			
19	Species Exploitation Regulation			
20	Species Introduction Prohibition			
21	Reed Cutting Restriction			

D Instrument Similarities

If diffusion occurs, we'd expect this to also affect the *instruments* implemented by a given policy. Embedded descriptions of each model-extracted instrument allow comparisons between the instruments found in one document to the instruments found in another. In particular, we can produce a cosine-similarity matrix for each document pair, and thereby investigate overlapping use of instruments between high-similarity document pairs. An example is given in Figure 19, where we compare the instruments of one of the more similar document pairs in the dataset, the European Single-Use Plastics Directive and the Danish *Engangsplastbekendtgørelse*.

The matrix shows relatively high baseline similarity, likely stemming at least partly from the common subject area, but also the common schema and field labels of the embedded text. It is also apparent that there is significant 1:1 overlap between several of the instruments - as expected, seeing as the Danish order transposes most of the restrictions from the directive. Table 4 shows side-by-side comparison of the descriptions embedded for the two most similar instrument pairs. Note that not every instrument corresponds exactly 1:1, which could both be a result of the Danish order implementing more than just the directive, some parts of the directive being implemented elsewhere in Danish law (for instance, the Danish Environmental Protection Act (*Miljøbeskyttelsesloven*) implements some Extended Producer Responsibility Schemes), and the previously investigated variance in how the model delineates instruments. In the extreme case, the latter could result in no apparent overlap between instruments for a document pair that each implement the same regulation, if simply the model mixes, groups or splits similar rules into different instruments in different documents. Thereby, the fruitfulness of this line of inquiry is likely limited at this stage.

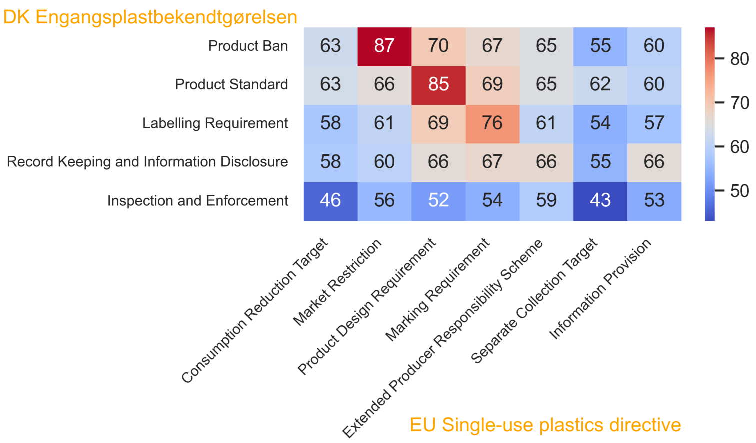


Figure 19: Instrument-comparison matrix between the European Single-Use Plastics Directive and the Danish *Engangsplastbekendtgørelse*

E Network rules

This appendix shows the structure of the FAOLEX network under different network similarity thresholds.

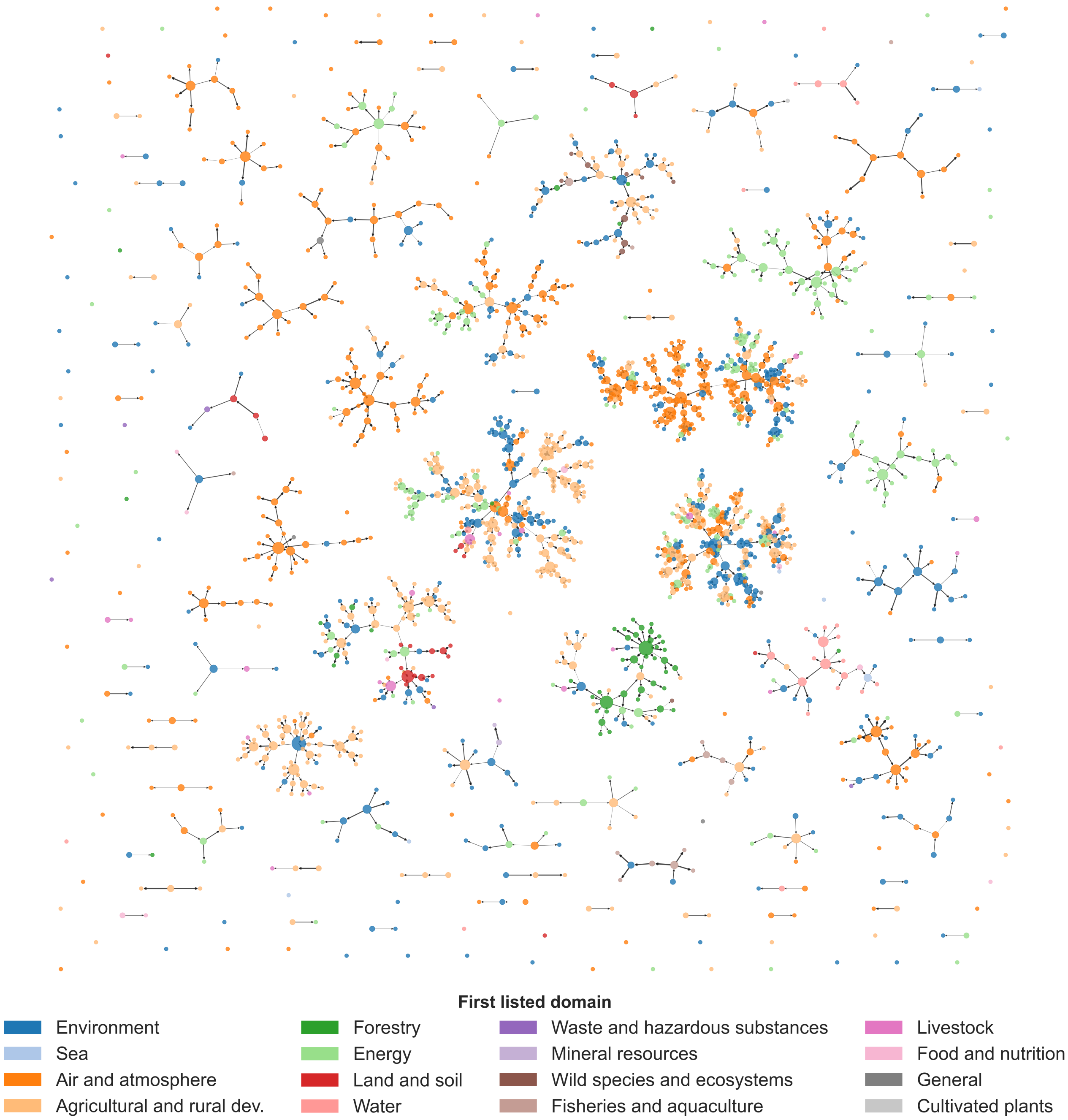


Figure 20: FAOLEX network for similarity threshold > 0.65

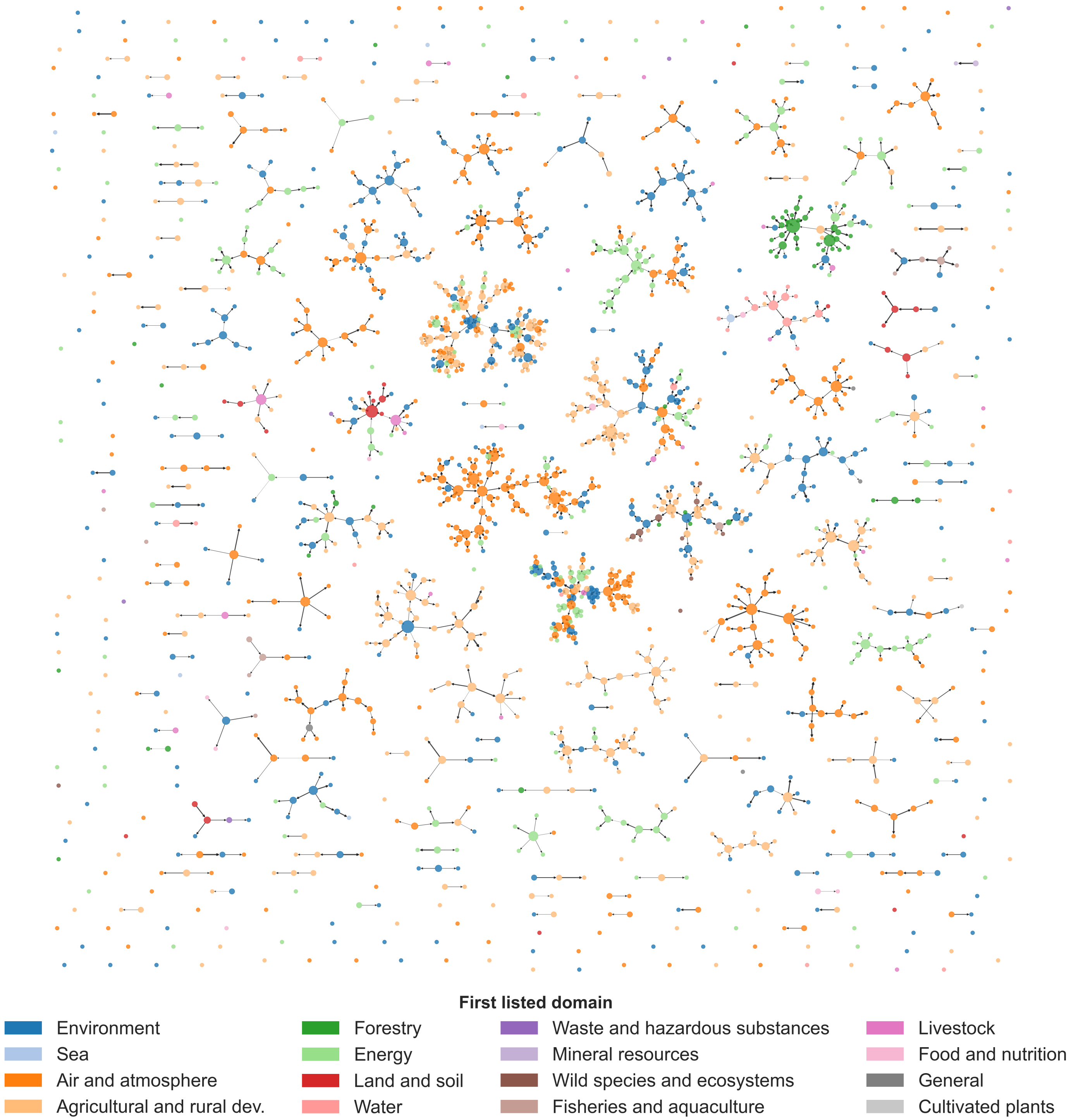


Figure 21: FAOLEX network for similarity threshold > 0.70

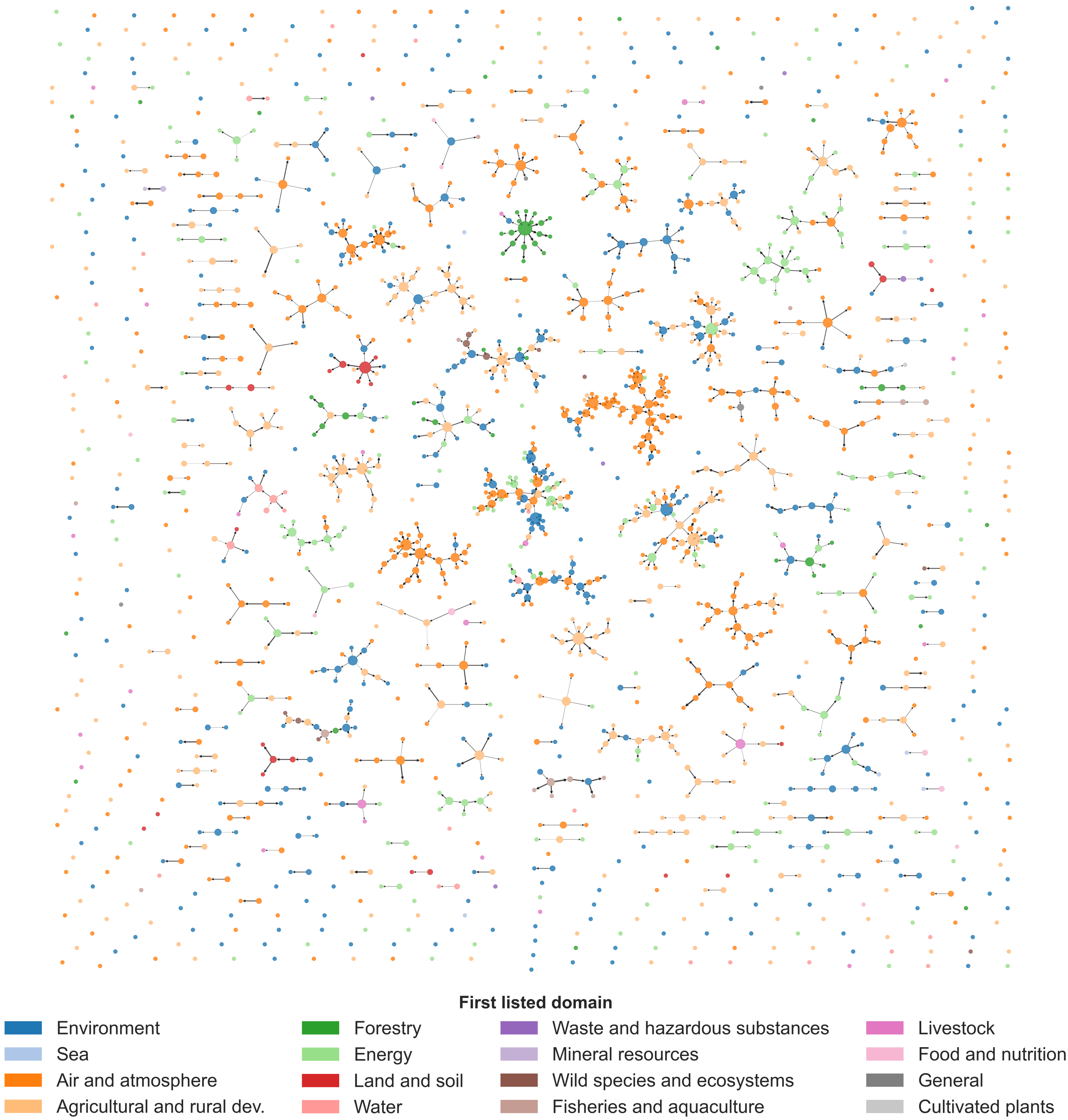


Figure 22: FAOLEX network for similarity threshold > 0.75

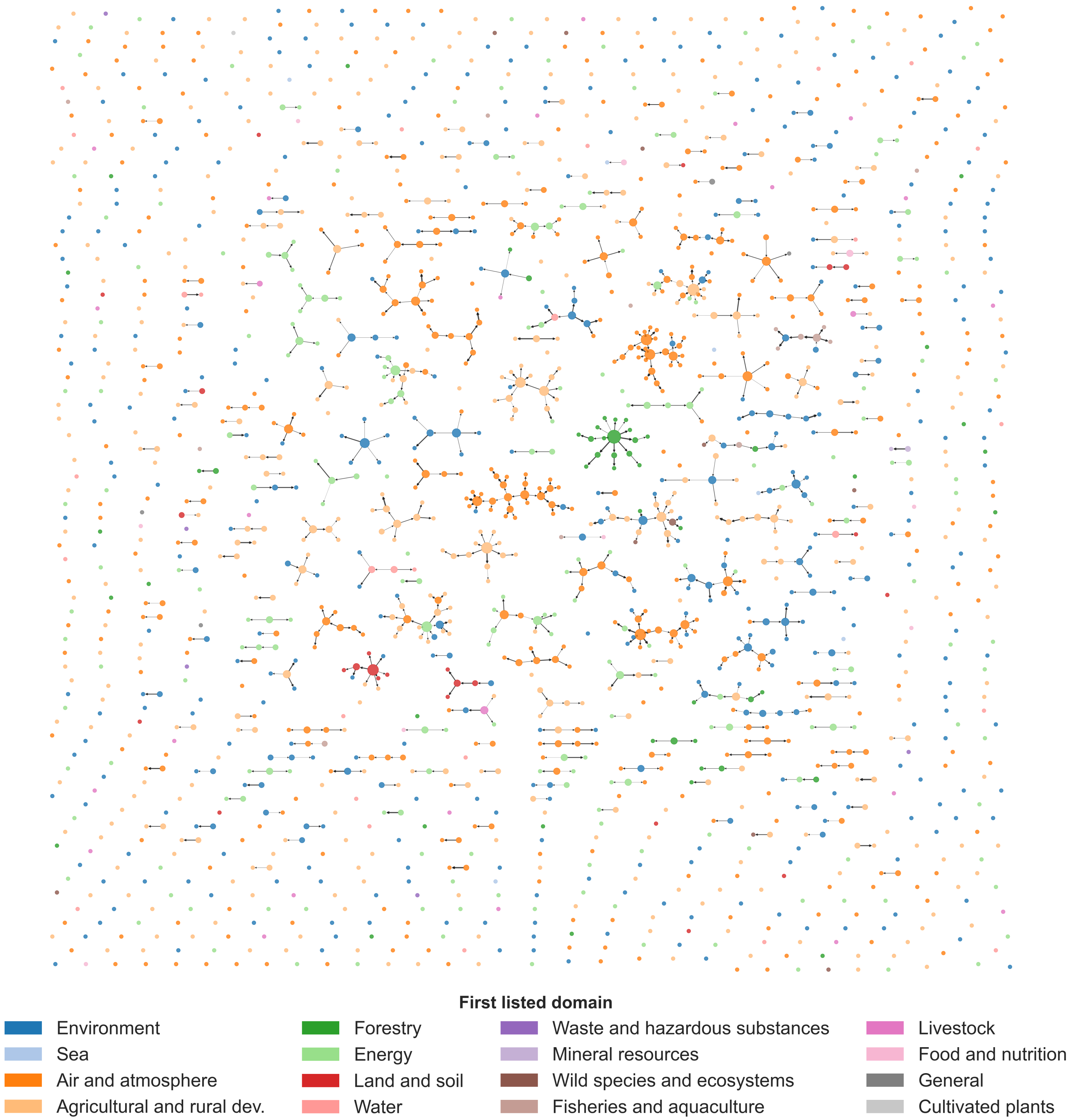


Figure 23: FAOLEX network for similarity threshold > 0.80

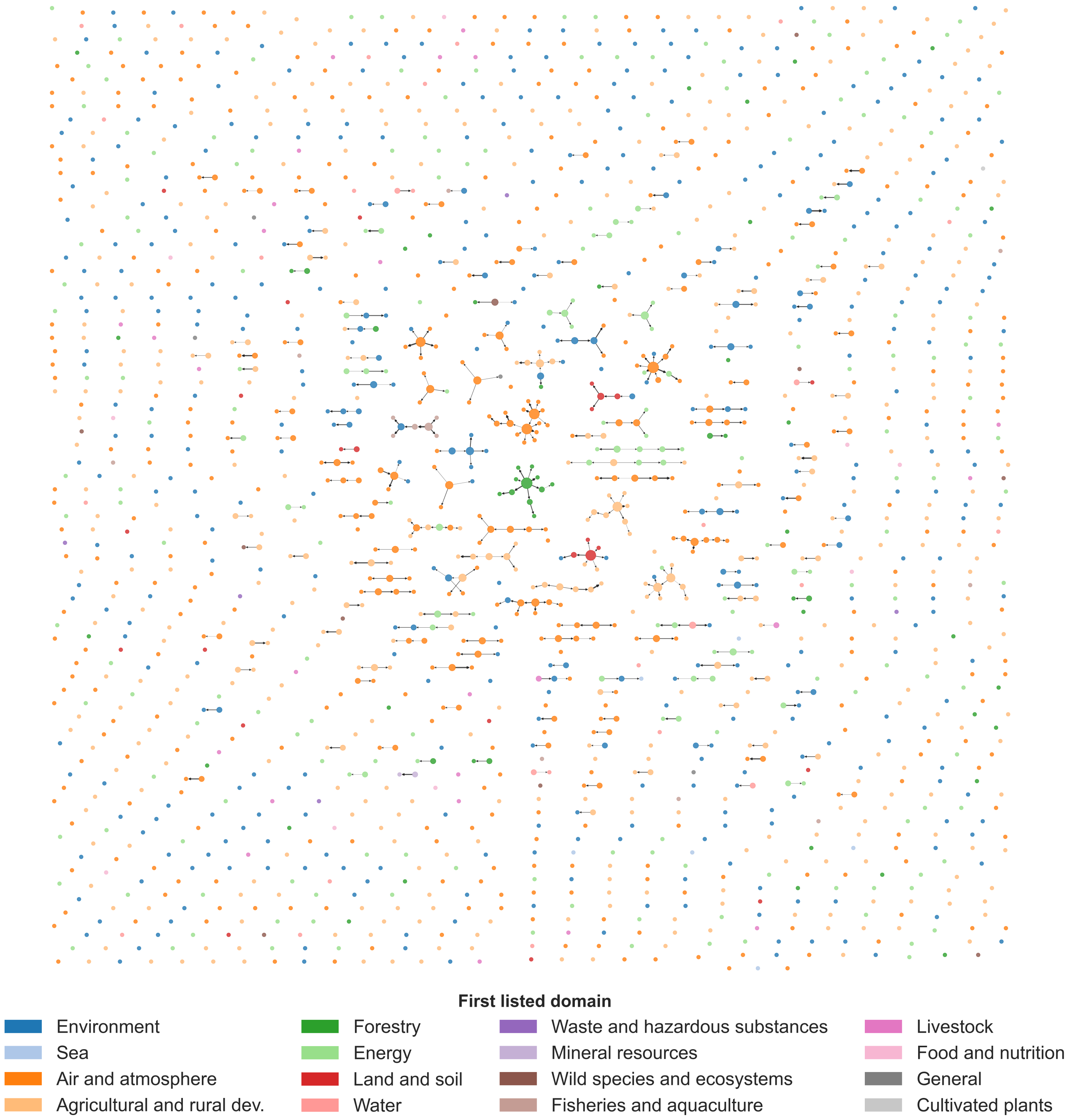


Figure 24: FAOLEX network for similarity threshold > 0.85

F Datasets

Directives Testset

Date d/m/y	Country	Policy name	Shorthand name	Type	Consolidated version?	Associated EU legislation	Policy area	PDF recovered	URL
05/06/2019	European Union	Directive (EU) 2019/904 of the European Parliament and of the Council of 5 June 2019 on the reduction of the impact of certain plastic products on the environment	Single-use plastics directive	Directive	FALSE	Waste Framework directive;	Plastics	18/02/2025	https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32019L0904
11/10/2024	Denmark	LBK nr 1093 af 11/10/2024 Bekendtgørelse af lov om miljøbeskyttelse	Miljøbeskyttels	Law	TRUE	Single-use plastics directive; Landfill of waste directive; Habitats directive	Multiple	18/02/2025	https://www.retsinformation.dk/eli/lta/2024/1093
04/09/2023	Denmark	BEK nr 1173 af 04/09/2023 Bekendtgørelse om forbud mod markedsføring af visse engangsplastprodukter m.v. og om krav til visse andre engangsplastprodukter	Engangsplastbe	Delegated Act	FALSE	Single-use plastics directive	Plastics	18/02/2025	https://www.retsinformation.dk/eli/lta/2023/1173
21/11/2024	Ireland	WASTE MANAGEMENT ACT 1996	Waste Management Act	Law	TRUE	Single-use plastics directive; Landfill of waste directive;	Multiple	18/02/2025	https://revisedacts.lawreform.ie/eli/1996/act/10/rev
15/10/2021	Ireland	S.I. No. 516/2021 - European Union (Single Use Plastics) (No. 2) Regulations 2021	Single Use Plastics Regulation (No. 2) 2021	Delegated Act	FALSE	Single-use plastics directive	Plastics	18/02/2025	https://www.irishstatutebook.ie/eli/2021/si/516/ma
26/04/1999	European Union	Council Directive 1999/31/EC of 26 April 1999 on the landfill of waste	Landfill of waste directive	Directive	TRUE		Waste	18/02/2025	https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32019L0904
21/11/2019	Denmark	BEK nr 1253 af 21/11/2019 Bekendtgørelse om deponeringsanlæg	Deponeringsbek	Delegated Act	FALSE	Landfill of waste directive	Waste	18/02/2025	https://www.retsinformation.dk/eli/lta/2019/1253
19/09/2024	Ireland	ENVIRONMENTAL PROTECTION AGENCY ACT 1992	Environmental Protection Agency Act	Law	TRUE	Landfill of waste directive;	Waste	18/02/2025	https://revisedacts.lawreform.ie/eli/1992/act/7/revis
14/07/2003	Ireland	PROTECTION OF THE ENVIRONMENT ACT 2003	Protection of the Environment Act	Law	FALSE	Landfill of waste directive;	Waste	18/02/2025	https://www.irishstatutebook.ie/eli/2003/act/27/ena
12/07/2004	Ireland	S.I. No. 395/2004 - Waste Management (Licensing) Regulations 2004	Waste (Licensing) Regulations 2004	Delegated Act	FALSE	Landfill of waste directive;	Waste	18/02/2025	https://www.irishstatutebook.ie/eli/2004/si/395/ma

Date d/m/y	Country	Policy name	Shorthand name	Type	Consolidated version?	Associated EU legislation	Policy area	PDF recovered	URL
21/05/1992	European Union	Council Directive 92/43/EEC of 21 May 1992 on the conservation of natural habitats and of wild fauna and flora	Habitats directive	Directive	TRUE		Biodiversity	18/02/2025	https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:31992L0043
28/06/2024	Denmark	LBK nr 927 af 28/06/2024 Bekendtgørelse af lov om naturbeskyttelse	Naturbeskyttels	Law	TRUE	Habitats directive;	Biodiversity	18/02/2025	https://www.retsinformation.dk/eli/lta/2024/927
21/08/2023	Denmark	BEK nr 1098 af 21/08/2023 Bekendtgørelse om udpegning og administration af internationale naturbeskyttelsesområder samt beskyttelse af visse arter	Habitatbekendt	Delegated Act	FALSE	Habitats directive;	Biodiversity	18/02/2025	https://www.retsinformation.dk/eli/lta/2023/1098
16/05/2024	Ireland	Wildlife Act 1976	Wildlife Act	Law	TRUE	Habitats directive;	Biodiversity	18/02/2025	https://revisedacts.lawreform.ie/eli/1976/act/39/rev
01/06/2024	Ireland	Wildlife (Amendment) Act 2000	Wildlife (Amendment) Act 2000	Law	TRUE	Habitats directive;	Biodiversity	18/02/2025	https://revisedacts.lawreform.ie/eli/2000/act/38/rev
21/09/2011	Ireland	S.I. No. 477/2011 - European Communities (Birds and Natural Habitats) Regulations 2011	Birds and Natural Habitats Regulations 2011	Delegated Act	FALSE	Habitats directive;	Biodiversity	18/02/2025	https://www.irishstatutebook.ie/eli/2011/si/477/ma
11/12/2018	European Union	Directive (EU) 2018/2001 of the European Parliament and of the Council of 11 December 2018 on the promotion of the use of energy from renewable sources (recast)	Renewable energy directive	Directive	TRUE		Renewable Energy		https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32018L2001
06/09/2024	Denmark	LBK nr 1031 af 06/09/2024 Bekendtgørelse af lov om fremme af vedvarende energi	VE-loven	Law	TRUE	Renewable energy directive;	Renewable Energy	18/02/2025	https://www.retsinformation.dk/eli/lta/2024/1031
20/06/2024	Denmark	BEK nr 773 af 20/06/2024 Bekendtgørelse om kontaktpunkt, VE-tilladelsesprocessen og områder til fremme af VE	VE-områder bekendtgørelsen	Delegated Act	FALSE	Renewable energy directive;	Renewable Energy	18/02/2025	https://www.retsinformation.dk/eli/lta/2024/773
31/12/2024	Ireland	ELECTRICITY REGULATION ACT 1999 REVISED Updated to 31 December 2024	Electricity Regulation Act	Law	TRUE	Renewable energy directive;	Renewable Energy	18/02/2025	https://revisedacts.lawreform.ie/eli/1999/act/23/re
25/09/2020	Ireland	S.I. No. 365/2020 - European Union (Renewable Energy) Regulations 2020	Renewable Energy Regulations 2020	Delegated Act	FALSE	Renewable energy directive;	Renewable Energy	18/02/2025	https://www.irishstatutebook.ie/eli/2020/si/365/ma
25/02/2022	Ireland	S.I. No. 76/2022 - European Union (Renewable Energy) Regulations 2022	Renewable Energy Regulations 2022	Delegated Act	FALSE	Renewable energy directive;	Renewable Energy	18/02/2025	https://www.irishstatutebook.ie/eli/2022/si/76/mad
15/07/2022	Ireland	S.I. No. 350/2022 - European Union (Renewable Energy) Regulations (2) 2022	Renewable Energy Regulations (2) 2022	Delegated Act	FALSE	Renewable energy directive;	Renewable Energy	18/02/2025	https://www.irishstatutebook.ie/eli/2022/si/350/ma

Date d/m/y	Country	Policy name	Shorthand name	Type	Consolidated version?	Associated EU legislation	Policy area	PDF recovered	URL
27/04/2016	European Union	Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)	GDPR	Regulation	TRUE	Unrelated	Unrelated testset	16/03/2025	https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A02016R0679-20160504
06/01/2025	Denmark	Bekendtgørelse af lov om ligestilling af kvinder og mænd	Ligestillingslove	Law	TRUE	Unrelated	Unrelated testset	16/03/2025	https://www.retsinformation.dk/eli/lta/2025/5
27/09/2021	Other	Vietnam's solar and wind power success: Polic other ASEAN countries	Academic lit. - Vietnam RE Policy	Academic litterature	FALSE	Unrelated	Unrelated testset	16/03/2025	https://doi.org/10.1016/j.esd.2021.09.002
01/10/2024	Other	Vegetarian Giouvetsi	Recipe - Climate friendly foods	Recipe	FALSE	Unrelated	Unrelated testset	16/03/2025	https://climate-friendly-cooking.com/recipe/giouvetsi-vegetarian/
30/05/1984	Other	Black Company Vol. 1, Chap. 4: Whisper	Fiction chapter - BlackCompanyVol1Chap4	Fiction	FALSE	Unrelated	Unrelated testset	16/03/2025	
27/02/2025	Other	En "grøn industrirevolution" skal få Europa på fode igen. Men planen bliver mødt med anklager om "økonomisk vanvid"	News - Green EU Politics	News	FALSE	Unrelated	Unrelated testset	16/03/2025	https://www.altinget.dk/eu/artikel/en-groen-industrirevolution-skal-faa-europa-paa-fode-igen-bla
16/03/2025	Other	Utrolig bedrift: Mads Pedersen vinder den grønne trøje	News - Sports	News	FALSE	Unrelated	Unrelated testset	16/03/2025	https://ekstrabladet.dk/sport/cykling/utrolig-bedrift-mads-pedersen-vinder-den-groenne-troeje/10560536

Adaptation Strategies

Date d/m/y	**Country**	**Policy name**	**Shorthand name**	**Type**	**PDF recovered**
31/03/2008	Denmark	Strategi for tilpasning til klimaændringer i Danmark	AS (2008)	Strategy	10/04/2025
31/12/2012	Denmark	Sådan håndterer vi skybrud og regnvand - Handlingsplan forklimasikring afDanmark	AS (2012)	Strategy	10/04/2025
23/10/2007	European Union	DIRECTIVE 2007/60/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 23 October 2007 on the assessment and management of flood risks	Floods directive (2007)	Directive	10/04/2025
29/06/2007	European Union	GREEN PAPER FROM THE COMMISSION TO THE COUNCIL, THE EUROPEAN PARLIAMENT, THE EUROPEAN ECONOMIC AND SOCIAL COMMITTEE AND THE COMMITTEE OF THE REGIONS Adapting to climate change in Europe - options for EU action	Green paper (2007)	Policy	10/04/2025
01/04/2009	European Union	WHITE PAPER Adapting to climate change: Towards a European framework for action	White paper (2009)	Policy	10/04/2025
16/04/2013	European Union	COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT, THE COUNCIL, THE EUROPEAN ECONOMIC AND SOCIAL COMMITTEE AND THE COMMITTEE OF THE REGIONS An EU Strategy on adaptation to climate change	AS (2013)	Strategy	10/04/2025
24/02/2021	European Union	COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT, THE COUNCIL, THE EUROPEAN ECONOMIC AND SOCIAL COMMITTEE AND THE COMMITTEE OF THE REGIONS Forging a climate-resilient Europe - the new EU Strategy on Adaptation to Climate Change	AS (2021)	Strategy	10/04/2025
31/07/2013	United Kingdom	The National Adaptation Programme Making the country resilient to a changing climate	AS (2013)	Strategy	10/04/2025
31/07/2018	United Kingdom	The National Adaptation Programme and the Third Strategy for Climate Adaptation Reporting Making the country resilient to a changing climate July 2018	AS (2018)	Strategy	10/04/2025
18/07/2018	United Kingdom	The Third National Adaptation Programme (NAP3) and the Fourth Strategy for Climate Adaptation Reporting	AS (2023)	Strategy	10/04/2025

Green Taxonomies

Date d/m/y	Author	Author Code	Policy name	Shorthand name	Type	Region	Comments	PDF recovered	URL
10/11/2021	Association of Southeast Asian Nations	ASEAN	ASEAN TAXONOMY FOR SUSTAINABLE FINANCE VERSION 1	Sustainable Taxonomy V1 (2021)	Final	South-east Asia		10/04/2025	https://asean.org/asean-sectoral-bodies-release-asean-taxonomy-for-sustainable-finance-version-1/
09/06/2023	Association of Southeast Asian Nations	ASEAN	ASEAN TAXONOMY FOR SUSTAINABLE FINANCE VERSION 2	Sustainable Taxonomy V2 (2023)	Final	South-east Asia		10/04/2025	https://asean.org/wp-content/uploads/2023/03/ASEAN-Taxonomy-Version-2.pdf
20/12/2024	Association of Southeast Asian Nations	ASEAN	ASEAN TAXONOMY FOR SUSTAINABLE FINANCE VERSION 3 - FOR TRANSPORTATION AND CONSTRUCTION SECTORS	Sustainable Taxonomy V3 (2024)	Final	South-east Asia			https://www.theacmf.org/sustainable-finance/publications/asean-taxonomy-version-3
30/06/2023	Thailand	TH	Thailand Taxonomy Phase 1	Taxonomy Phase 1 (2023)	Final	South-east Asia		10/04/2025	https://www.bot.or.th/en/financial-innovation/sustainable-finance/green/Thailand-Taxonomy.html
27/10/2024	Thailand	TH	Thailand Taxonomy Phase 2 (draft)	Taxonomy Phase 2 (2024 draft)	Draft	South-east Asia	Own merging	10/04/2025	https://www.bot.or.th/en/financial-innovation/sustainable-finance/green/Thailand-Taxonomy/public-hearing-thailand-taxonomy-phase2.html
20/01/2022	Indonesia	ID	TAKSONOMI HIJAU INDONESIA Indonesia Green Taxonomy Edisi 1.0 - 2022 / Edition 1.0 - 2022	Green Taxonomy 1 (2022)	Final	South-east Asia		10/04/2025	https://www.ojk.go.id/id/berita-dan-kegiatan/info-terkini/Pages/Taksonomi-Hijau-Indonesia-Edisi-1—2022.aspx
20/02/2024	Indonesia	ID	TAKSONOMI UNTUK KEUANGAN BERKELANJUTAN INDONESIA	Sustainable Finance Taxonomy (2024)	Final	South-east Asia		10/04/2025	https://keuanganberkelanjutan.ojk.go.id/keuanganb untuk-keuangan-berkelanjutan-indonesia

Date d/m/y	Author	Author Code	Policy name	Shorthand name	Type	Region	Comments	PDF recovered	URL
10/02/2025	Indonesia	ID	TAKSONOMI UNTUK KEUANGAN BERKELANJUTAN INDONESIA Versi 2	Sustainable Finance Taxonomy v2 (2025)	Final	South-east Asia		10/04/2025	https://www.ojk.go.id/id/Publikasi/Roadmap-dan-Pedoman/Sektor-Jasa-Keuangan/Keuangan-Berkelanjutan/Pages/Buku-Taksonomi-untuk-Keuangan-Berkelanjutan-Indonesia-(TKBI)-Versi-2.aspx
30/04/2021	Malaysia	MY	Climate Change and Principle-based Taxonomy	CC Taxonomy (2021)	Final	South-east Asia		10/04/2025	https://www.bnm.gov.my/documents/20124/938039 based+Taxonomy.pdf
12/12/2022	Malaysia	MY	PRINCIPLES-BASED SUSTAINABLE AND RESPONSIBLE INVESTMENT TAXONOMY FOR THE MALAYSIAN CAPITAL MARKET	Sustainable Investment Taxonomy (2022)	Final	South-east Asia		10/04/2025	https://www.sc.com.my/api/documentms/download 5d7d-4c66-8638-caec92c209c1
03/12/2023	Singapore	SG	Singapore-Asia Taxonomy for Sustainable Finance 2023 edition	Sustainable Finance Taxonomy (2023)	Final	South-east Asia		10/04/2025	https://eurocham.org.sg/wp-content/uploads/2024/01/Singapore-Asia-Taxonomy-Dec-2023.pdf
30/12/2020	Bangladesh	BD	Sustainable Finance Policy for Banks and Financial Institutions	Sustainable Finance Policy (2020)	Final	South Asia		10/04/2025	https://www.bb.org.bd/mediaroom/circulars/gbcrd
06/05/2022	Sri Lanka	LK	Sri Lanka Green Finance Taxonomy	Green Finance Taxonomy (2022)	Final	South Asia		10/04/2025	https://www.cbsl.gov.lk/en/sl-green-finance-taxonomy
17/12/2019	Mongolia	MN	Mongolian Green Taxonomy	Green Taxonomy (2019)	Final	East Asia			https://www.toc.mn/en/publication/mongol-ulsyn-nogoon-taksonomi
14/02/2019	China	CN	Green Industry Guidance Catalogue (2019)	GIGC (2019)	Final	East Asia	Full document	10/04/2025	https://www.ncsti.gov.cn/kjdt/tzgg/202109/P02021

Date d/m/y	Author	Author Code	Policy name	Shorthand name	Type	Region	Comments	PDF recovered	URL
16/03/2023	China	CN	Green Industry Guidance Catalogue (2023 draft)	GIGC (2023)	Draft	East Asia	**Excluded - causes trouble in completions. Own merging of documents and preamble**	10/04/2025	https://yyglxxbsgw.ndrc.gov.cn/htmls/article/arti
02/02/2024	China	CN	Green Low-Carbon Transition Industry Guidance Catalogue (2024 edition)	GIGC (2024)	Final	East Asia	Own merging of documents and preamble	10/04/2025	https://www.gov.cn/zhengce/zhengceku/202403/co
15/12/2015	China	CN	Green Bond Endorsed Project Catalogue (2015 edition)	GBEPC (2015)	Final	East Asia	Own merging of documents and preamble	10/04/2025	https://policy.asiapacificenergy.org/node/2675; https://www.icmagroup.org/assets/documents/Regu Bonds/Preparation-Instructions-on-Green-Bond-Endorsed-Project-Catalogue-2015-Edition-by-EY.pdf
02/04/2021	China	CN	Green Bond Endorsed Project Catalogue (2021 edition)	GBEPC (2021)	Final	East Asia	Full document	10/04/2025	https://www.climatebonds.net/files/files/the-Green-Bond-Endorsed-Project-Catalogue-2021-Edition-110521.pdf
30/12/2021	South Korea	KR	Korean Green Classification System Guidelines (K-TAXONOMY)	Sustainable Finance Taxonomy (2021)	Final	East Asia		10/04/2025	https://me.go.kr/home/web/policy_data/read.do?m
03/05/2024	Hong Kong	HK	Hong Kong Taxonomy for Sustainable Finance	Sustainable Finance Taxonomy (2024)	Final	East Asia		10/04/2025	https://brdr.hkma.gov.hk/eng/doc-ldg/docId/20240503-3-EN

Date d/m/y	Author	Author Code	Policy name	Shorthand name	Type	Region	Comments	PDF recovered	URL
03/06/2022	China;EU	CN;EU	International Platform on Sustainable Finance Common Ground Taxonomy - Climate Change Mitigation	Common Ground Taxonomy (2022)	Report	IO	Own merging of report and tables	10/04/2025	https://finance.ec.europa.eu/system/files/2022-06/220603-international-platform-sustainable-finance-common-ground-taxonomy-instruction-report_en.pdf;https://finance.ec.europa.eu/system/06/220603-international-platform-sustainable-finance-common-ground-taxonomy-table-activities_en.pdf
18/06/2020	European Union	EU	REGULATION (EU) 2020/852 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 18 June 2020 on the establishment of a framework to facilitate sustainable investment, and amending Regulation (EU) 2019/2088	Taxonomy Regulation (2020)	Final	Europe		10/04/2025	https://eur-lex.europa.eu/eli/reg/2020/852/oj/eng
04/06/2021	European Union	EU	COMMISSION DELEGATED REGULATION (EU) 2021/2139 of 4 June 2021 supplementing Regulation (EU) 2020/852 of the European Parliament and of the Council by establishing the technical screening criteria for determining the conditions under which an economic activity qualifies as contributing substantially to climate change mitigation or climate change adaptation and for determining whether that economic activity causes no significant harm to any of the other environmental objectives	Taxonomy DA1 Climate1 (2021)	Final	Europe		10/04/2025	https://eur-lex.europa.eu/eli/reg_del/2021/2139/oj/eng
06/07/2021	European Union	EU	COMMISSION DELEGATED REGULATION (EU) 2021/2178 of 6 July 2021 supplementing Regulation (EU) 2020/852 of the European Parliament and of the Council by specifying the content and presentation of information to be disclosed by undertakings subject to Articles 19a or 29a of Directive 2013/34/EU concerning environmentally sustainable economic activities, and specifying the methodology to comply with that disclosure obligation	Taxonomy DA2 Disclosure (2021)	Final	Europe		10/04/2025	https://eur-lex.europa.eu/eli/reg_del/2021/2178/oj/eng
09/03/2022	European Union	EU	COMMISSION DELEGATED REGULATION (EU) 2022/1214 of 9 March 2022 amending Delegated Regulation (EU) 2021/2139 as regards economic activities in certain energy sectors and Delegated Regulation (EU) 2021/2178 as regards specific public disclosures for those economic activities	Taxonomy DA3 Climate2 (2022)	Final	Europe			https://eur-lex.europa.eu/eli/reg_del/2022/1214/oj/eng

Date d/m/y	Author	Author Code	Policy name	Shorthand name	Type	Region	Comments	PDF recovered	URL
27/06/2023	European Union	EU	COMMISSION DELEGATED REGULATION (EU) 2023/2486of 27 June 2023supplementing Regulation (EU) 2020/852 of the European Parliament and of the Council byestablishing the technical screening criteria for determining the conditions under which aneconomic activity qualifies as contributing substantially to the sustainable use and protection ofwater and marine resources, to the transition to a circular economy, to pollution prevention andcontrol, or to the protection and restoration of biodiversity and ecosystems and for determiningwhether that economic activity causes no significant harm to any of the other environmentalobjectives and amending Commission Delegated Regulation (EU) 2021/2178 as regards specificpublic disclosures for those economic activities	Taxonomy DA4 Environment (2023)	Final	Europe		10/04/2025	https://eur-lex.europa.eu/eli/reg_del/2023/2486/oj/eng
21/09/2021	Russia	RU	Order of the Government of the Russian Federation of February 21, 2021 No. 1587 "About approval of the criteria of the projects of sustainable (including green) development in the Russian Federation and requirements to the system of verification of instruments of financing of sustainable development in the Russian Federation	Green Project Taxonomy (2021)	Final	Europe		10/04/2025	http://actual.pravo.gov.ru/content/content.html#p
31/12/2021	Kazakhstan	KZ	Classification (taxonomy) of "green" projects to be financed through green bonds and "green" loans	Green Project Taxonomy (2021)	Final	Central Asia		10/04/2025	https://adilet.zan.kz/rus/docs/P2100000996
29/04/2022	Georgia	GE	Sustainable Finance Taxonomy for Georgia (2022)	Sustainable Finance Taxonomy (2023)	Final	Central Asia		10/04/2025	https://nbg.gov.ge/en/page/sustainable-finance-tax
03/03/2023	Canada	CA	Taxonomy Roadmap Report Mobilizing Finance for Sustainable Growth by Defining Green and Transition Investments Sustainable Finance Action Council September 2022	Taxonomy Roadmap Report (2023)	Report	North America	Not a draft just a report	10/04/2025	https://www.canada.ca/en/department-finance/programs/financial-sector-policy/sustainable-finance/sustainable-finance-action-council/taxonomy-roadmap-report.html
12/08/2024	Costa Rica	CR	Taxonomía de Finanzas Sostenibles	Sustainable Finance Taxonomy (2024)	Final	Latin America		10/04/2025	https://www.sugeval.fi.cr/Informacion-inversionista

Date d/m/y	Author	Author Code	Policy name	Shorthand name	Type	Region	Comments	PDF recovered	URL
11/03/2023	Mexico	MX	Taxonomía Sostenible De México (Primera Edición Marzo 2023)	Sustainable Taxonomy (2023)	Final	Latin America		10/04/2025	https://www.sbfnetwork.org/wp-content/uploads/2023/03/1131_Mexico_Mexican_S 2023.pdf
04/06/2024	Dominican Republic	DO	Taxanomía Verde República Dominicana	Green Taxonomy (2024)	Final	Latin America		10/04/2025	https://bvearmb.do/handle/123456789/4851
30/03/2024	Panama	PA	TAXONOMÍA DE FINANZAS SOSTENIBLES DE PANAMÁ PRIMERA EDICIÓN: MARZO 2024	Sustainable Finance Taxonomy (2024)	Final	Latin America		10/04/2025	https://www.superbancos.gob.pa/taxonomia-finanzas-sostenibles/documentos
12/07/2023	UNEP	LA	Common Framework for Sustainable Finance Taxonomies for Latin America and the Caribbean	Sustainable Finance Taxonomies Framework	Report	Latin America		10/04/2025	https://www.unep.org/resources/report/common-framework-sustainable-finance-taxonomies-latin-america-and-caribbean
30/03/2022	Colombia	CO	TAXONOMÍA VERDE DE COLOMBIA	Green Taxonomy (2022)	Final	Latin America		10/04/2025	https://www.taxonomiaverde.gov.co/
30/12/2024	Chile	CL	Sistema de Clasificación o Taxonomía de Actividades Económicas Medioambientalmente Sostenibles de Chile (T-MAS)	Sustainable Taxonomy (2024 draft)	Draft	Latin America		10/04/2025	https://cms.hacienda.cl/ciudadana/assets/documen
01/04/2022	South Africa	SA	South African Green Finance Taxonomy (1st EDITION)	Green Finance Taxonomy (2022)	Final	Sub-Saharan Africa		10/04/2025	https://sustainablefinanceinitiative.org.za/working-groups/taxonomy-working-group/
30/10/2024	Rwanda	RW	Rwanda Green Taxonomy	Green Taxonomy (2024)	Final	Sub-Saharan Africa	Summary merged with 3 annexes	10/04/2025	https://www.minecofin.gov.rw/rwandagreentaxonon
31/10/2024	Australia	AU	Australian Sustainable Finance Taxonomy V0.1 Public Consultation Paper (second consultation)	Sustainable Finance Taxonomy (2nd draft Oct 2024)	Draft	Oceania		10/04/2025	https://www.asfi.org.au/taxonomy-public-consultation