

The Climate Contracting Knowledge Graph: Construction and Evaluation

Georgia Ray^{1,2}, Felix Cohen¹

¹The Chancery Lane Project, ²Imperial College London

Correspondence: georgia.ray@chancerylaneproject.org

Abstract

Contracts serve as a cornerstone mechanism of private environmental governance, and yet their form as unstructured text has meant a recent overhaul in their construction and iteration due to large language models (LLMs). As such, historical forms of knowledge sharing in the legal field must also change. We present the Climate Contracting Knowledge Graph, a publicly released resource that organises the model clauses, glossary terms, and guides of The Chancery Lane Project (TCLP) into a queryable property graph, linked to the laws and standards they invoke. Existing climate knowledge graphs target policy documents or climate-model metadata; none represent the contractual instruments through which private climate commitments are implemented. The graph is built with a lawyer-in-the-loop pipeline and achieves a statistically significant improvement in mean average precision over dense and lexical baselines. The graph, models, and a lightweight retrieval service are released publicly.

1 Introduction

Private environmental governance (PEG) is the private implementation of standards, monitoring, enforcement, and dispute resolution toward public ends (Vandenbergh, 2013). Vandenbergh & Johnson characterise contracting and other private enforcement mechanisms as “an important gap-filling effort given the limits of international, national and sub-national climate mitigation” (Vandenbergh and Johnson, 2021). Contracting, one component of PEG, has been shown to function as a lever in emissions reduction (Wang et al., 2024), and has been proposed as an enforcement mechanism for business-to-business climate action (Leonard, 2020). More recently AI companies have been called to provide environmental disclosures, and scholars have recognised contractual lock-in of en-

ergy supply for large-scale infrastructure projects powering AI (Luccioni et al., 2024, 2025).

In order to practice PEG, lawyers must be able to locate and deploy the relevant climate-aligned contracting content. However, as legal practice increasingly incorporates large language model (LLM)-based workflows, conventional mechanisms of information retrieval are being displaced, and the infrastructure for hosting professional content must adapt accordingly. Knowledge graphs (KGs) offer a mechanism for embedding structured legal and climate knowledge within LLM-based systems, and such graphs are increasingly being constructed across the climate domain: for policy documents (Dutia et al., 2025), climate model data (Ceriani et al., 2026), and mapping climate science (Carbon Brief, 2026). However, to this point, none address contractual instruments through which much climate action is implemented.

We present such a resource: a knowledge graph constructed from professionally drafted and maintained clauses, glossary terms, and guides for climate-aligned contracting, linked to the laws and standards they invoke. The graph is constructed via a lawyer-in-the-loop pipeline that favours task-specific models where feasible and LLMs otherwise. A retrieval ablation indicates employing the KG structure contributes a discriminative signal beyond dense and lexical retrieval alone, notably improving mean average precision (MAP). We further demonstrate the resource’s utility through potential applications and release a lightweight interface for integration with third-party LLMs via a Model Context Protocol (MCP) server.¹ The underlying data and models are also made publicly available.^{2,3}

¹<https://mcp.chancerylaneproject.org/>

²<https://huggingface.co/thechancerylaneproject/legal-sustainability-ner>

³<https://labs.chancerylaneproject.org/project/tclp-clause-library/>

2 Related Works

The peculiarities of a specific domain language in climate science and law, and the challenges those peculiarities pose to natural language processing (NLP) have long been recognised. Webersinke et al. (2021) show that general language models represent climate text poorly and release ClimateBERT, while Chalkidis et al. release LEGAL-BERT as the legalese equivalent (2020). Both of these BERT-based applications have enabled further exploration of climate and legal language in assessing corporate climate risk disclosures (Bingler et al., 2022), detecting and assessing net zero targets (Schimanski et al., 2023), and predicting legal judgements (Niklaus et al., 2021).

Meanwhile, KGs are acknowledged as a tool to ground LLMs in structured, verifiable knowledge, improving factual reliability, reducing hallucination, and promoting responsible use (Agrawal et al., 2023; Pan et al., 2024). LLMs are based on a transformer architecture, which encodes contextual meaning into the numerical (vector) representation of language (Vaswani et al., 2017). However, this approach carries a quadratic computational cost that constrains context length (Tay et al., 2023), and has been empirically shown to attend to the full context non-uniformly, typically focusing on the beginning and end (Liu et al., 2023). As such, uncommon relationships, like those specific to climate legal practice, or context provided only once, are liable to be under-weighted or missed. KGs address this shortcoming by supplying relationships explicitly and in a machine-retrievable manner with lower noise and higher density than raw text (Edge et al., 2024).

Surveys of KG construction describe a common pipeline of extraction, entity resolution, and refinement, and note that several of these stages still require human input (Zhong et al., 2024; Hofer et al., 2024). Ours follows the same pipeline, using lawyer annotations to fine-tune the extractor, LLMs for typing and entity matching, and expert review at the refinement stage.

Our evaluation adopts the information-retrieval perspective on knowledge graphs (Reinanda et al., 2020), meaning we evaluate the graph as a structure for finding and ranking relevant documents and entities, not as a system that performs multi-step logical inference over its edges.

3 Presentation of the knowledge graph

3.1 Source corpus

The corpus comprises 200 climate clauses, 76 glossary terms, and 11 guides authored and maintained by TCLP.⁴ Of the clauses, 18 currently have a “maintained” status and therefore have the richest metadata taxonomy. All clauses have jurisdiction and practice area tags, and some additionally carry sector tags. The entire corpus of TCLP content was included in creating the knowledge graph.

3.2 Construction pipeline

Data collection and annotation. To gather hand-annotations from non-technical legal professionals, we built a multi-user annotation platform, seeded by an out-of-the-box spaCy model. spaCy’s own default predictions across the whole corpus were the starting point for legal review. Entities were then approved, adjusted, or removed by legal annotators, who were also able to add entities that had not been tagged by the out-of-the-box system. Entity labels were selected from a pre-approved set agreed with our internal legal professionals. Annotators were trained on the annotating platform and entity tagging in two 45-minute training sessions, led by report authors.

New entities were most often added where the general-purpose model had no matching category or did not understand the relevance of a given entity in the climate law context. For example, ‘Paris’ was tagged by spaCy as a location. Our annotators revised it as to be part of the broader phrase ‘Paris Agreement’ and changed the label type to ‘Standard’. Chunks were prioritised for annotation using a greedy minimum set-cover logic: the smallest chunk set covering the most not-yet-covered entities was repeatedly calculated and then presented first. This process continued until all pre-seeded entities had been seen.

Ultimately, 18 annotators annotated 831 of the 1,266 total chunks from the corpus. Annotators could merge or normalise entities manually, by

⁴These TCLP-specific terms are defined institutionally. Clauses are climate-aligned legal clauses one can use in commercial contracts and legal documents (e.g. Consolidated Sustainability Clauses for Supply Chain Contracts or Climate Change Due Diligence Questionnaire for Suppliers). Glossary terms are standard definitions to help one draft climate-aligned contracts (e.g. Alternative fuel or Residual emissions). And guides are documents that provide simple guidance for common scenarios and opportunities to use climate-aligned contracting (e.g. Introduce climate provisions to contract parties or Deliver a climate transition plan).



Figure 1: Static visualisation of sub-component of knowledge graph. This sub-component is focused around the glossary term “climate professional.” The red dots are glossary terms, the purple dots are clauses, the brown dots are entities, light blue are guides, bright blue are laws or regulations, and darker blue are standards or frameworks. Individual node labels are not intended to be legible at this size. The figure is included to show the general form of the graph and the types of entity it contains. The visualisation is provided by the neo4j platform and viewer.

free typing the canonical form for an entity set but additional normalisation is performed by an LLM in graph construction (see Section 3.3).

Extraction and labelling. The results of the annotation process were used to further fine-tune the LEGAL-BERT baseline (Chalkidis et al., 2020), creating an extractor tailored to this task and expert judgement in this domain. The model was run against unseen test chunks, achieving an F1 of 0.69 with a precision of 0.70 and a recall of 0.68 on binary entity extraction. Augmenting this tailored model was a simple and strict keyword matcher for organisation-specific terms, which are crucial to TCLP but low-frequency in the training data. Traditional NER includes type labelling in addition to binary extraction. Our manually annotated data was labelled, but those labels were very sparse (there were only 46 instances of the ‘sector’ label and 195 of the ‘law’ label. That, as compared to over 3000 instances of both ‘defined term’ and ‘concept’ entailed strong class imbalance). Unsurprisingly, the NER model underperformed in learning these class imbalanced labels, particularly on the lowest-represented labels, and so the NER labels were not relied on. Instead, an LLM (Claude Haiku via OpenRouter) was prompted to label the extracted entities and was provided with domain-specific knowledge curated by the same expert lawyers regarding the proper usage of each label. Because the LLM is only tasked with labelling, not

extraction, resource efficiency is improved and hallucinated entities are avoided. Ideally, using the LLM for labelling would have been unnecessary but as the labels are not an ontological organiser in the graph, a semi-automated approach demoting the labelling component to non-lawyer-reviewed was a prioritisation choice.

3.3 Graph creation and schema

The extraction pipeline writes its output to an intermediate SQLite storage, allowing the graph to be built and changed incrementally. When content was altered, a hash checkpoint was created to for the changes and extraction rerun for the altered content.

Each source document is represented by a clause, glossary term, or guide node. Alongside any taxonomy fields included in the specific piece of content, every node includes a dense embedding (BAAI/bge-base-en-v1.5, 768 dimensions).

Entities also become independent nodes. Every extracted entity is already keyed in SQLite to the document it was extracted from. To create entity nodes, we read this database row by row and either 1) create a new node if the entity is not yet in the graph or 2) merge that instance of an entity with the existing node. For merging, a normalised name match is tried first; where that fails, an LLM (Claude Haiku, via OpenRouter) performs a second, fuzzy pass that links acronyms, abbreviations,

	Count
<i>Parameters</i>	
CO_OCCURS_WITH threshold ≥ 31 docs	
<i>Nodes</i>	
Clause documents	200
Glossary terms	76
Guides	11
Laws & regulations	86
Standards & frameworks	202
Entity nodes (total)	5,424
of which: Concept	3,214
of which: Defined Term	1,563
of which: Object	1,128
of which: Org	223
of which: Standard	189
of which: Person	189
of which: Party to Contract	282
of which: Law	59
<i>Edges</i>	
CONTAINED_IN	16,366
CO_OCCURS_WITH	66,924
SIMILAR_TO	10,307
REFERENCES_LAW	183
REFERENCES_STANDARD	521

Table 1: Properties of the KG. Current counts for node types and edges, further disaggregated into entity subtypes (based on labels; a single entity may carry more than one label so the sub-totals sum is more than the entity total). The CO_OCCURS_WITH threshold of ≥ 31 shared source documents corresponds to the 75th percentile of the empirical co-document distribution. All counts reflect the live graph at time of writing.

and minor variants to their canonical form. Fuzzy-merge decisions were also reviewed iteratively during development and used to refine the matching rules, although not with a formal evaluation process. Every entity is linked to each of its sources of provenance via a CONTAINED_IN edge.

Other edges include REFERENCES_LAW, REFERENCES_STANDARD, CO_OCCURS_WITH, and SIMILAR_TO. Figure 1 visualises the content linked by these different edges.

The threshold, node counts, and edge counts for the current build are reported in Table 1.

LSRF process. Once the KG was created, we realised the metadata information about laws, standards, regulations, and frameworks (LSRF) was incomplete as it had been derived from the manually maintained metadata.

We proceeded with a two-pronged approach for completeness and verification. As a first step, we graduated entities labelled as ‘law’ or ‘standard’ from an Entity node type to a LawOrRegulation or StandardOrFramework node type, accordingly. Secondly, we used a two-step LLM procedure in

which a first, smaller LLM (google/gemma-4-31b-it) was given a chunk of text and asked to do a first pass identifying whether any LSRFs were present. A more sophisticated LLM (Claude Haiku) was then asked to classify each as a law, a standard, or neither. Both outputs, from the entity graduation approach and the LLM approach, were then manually reviewed by a domain expert.

Using two methods provides a measure of cross-validation and improves coverage. Approval rates were comparable between the methods, especially when considering occurrence (77.8% LLM, 75.4% graduation). However, the LLM pipeline was stronger on distinct instruments (69.6% vs 58.9%) and recovered 70.9% of the graduation instruments while the reverse was only 31.0%. (see Figure 2). This pipeline, including the human review, is an effort at properly seeding the graph with many of the relevant LSRFs for this community. As the graph expands, the approval ratings and cross-validated coverage from the LLM offer the potential for it to be used automatedly without human review.

4 Retrieval evaluation

4.1 Ablation against taxonomy derived gold standard

An important part of validating the usefulness of a KG in the information retrieval tradition is to have some measure of retrieval, and demonstrate performance compared to other retrieval mechanisms. We assess different retrieval mechanisms against manually established connections between pieces of information. We consider two clause nodes relevant to one another if they share at least two tags pooled across the practice-area and sector facets.

To measure retrieval, we report mean average precision (MAP) and mean reciprocal rank (MRR) as primary metrics and recall as secondary, with Recall@23 taken at the mean relevant-set size.

We compare four retrieval systems, fusing their components with reciprocal rank fusion (RRF, $k=60$) (Cormack et al., 2009); BM25 over clause text (Robertson and Zaragoza, 2009); dense retrieval over the clause embeddings; embeddings fused with BM25; and embeddings fused with the entity signal, in which two clauses are scored by the Jaccard overlap of the entities they share along CONTAINED_IN edges.

A random-ordering baseline has an expected precision of 0.12.

Augmenting dense retrieval with the entity layer

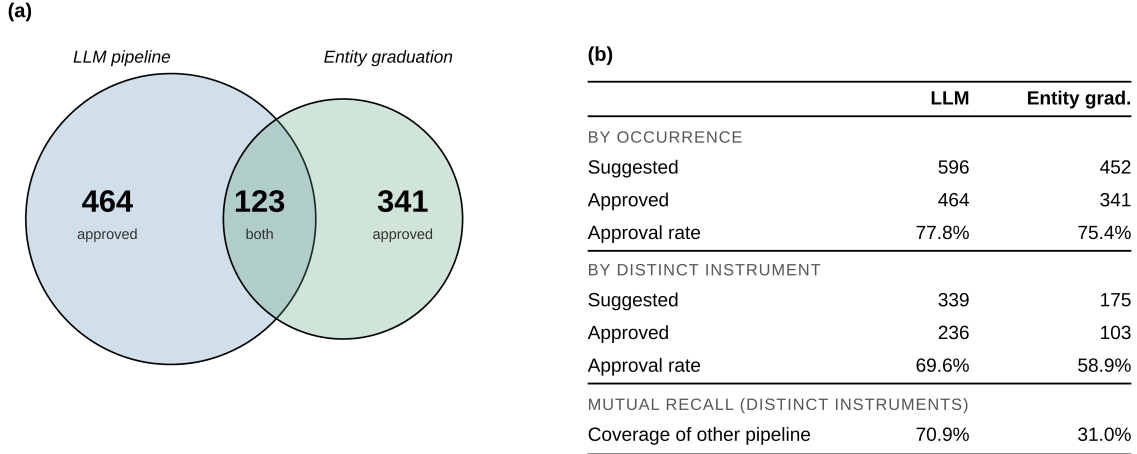


Figure 2: Coverage and overlap of LSRF identification pipelines. (a) Venn diagram showing approved occurrences (where occurrences are a node/clause pair rather than a unique node) for each pipeline. (b) Pipeline statistics by occurrence and by distinct instrument. Overlap was resolved using slug normalisation, parenthetical acronym extraction, and curated surface-form aliases, so that surface-form variants of the same instrument are counted as one match. Approval decisions reflect human review of pipeline-suggested instrument-clause links.

gives the strongest system on every metric, raising MAP from 0.465 to 0.513, a gain that is significant under a paired Wilcoxon signed-rank test ($p < 0.0001$). The entity-augmented system also significantly outperforms lexical fusion (embeddings + BM25; $p < 0.00001$), whereas BM25 fusion does not significantly improve on dense retrieval alone ($p = 0.062$). MRR is high and near-identical across the embedding-based systems (~ 0.81). This measure means a relevant clause is typically returned at rank one or two, no matter the system. The differences arise deeper in the ranking, which MAP and recall capture. The signal coming from the KG entities, therefore, is not reducible to lexical overlap. This finding indicates entities encode inter-document structure that neither dense nor lexical retrieval recovers. In short, using entities outperforms other methods of retrieval on all but one retrieval metric, suggesting there is an improvement in our ability to capture relevant information for these important private environmental governance questions using the KG.

4.2 Competency questions

On a more qualitative level, competency questions consider whether the graph can successfully answer the type of questions it is designed to answer (Grüniger and Fox, 1995).

We wrote a set of questions spanning the graph’s intended uses, and confirmed that each is answerable by a single graph query. We had four types of questions: finding related clauses, concept and def-

inition lookup, filtering by taxonomy, and lookup and aggregation. The full set and results can be seen in Appendix A.

5 Conclusions and Future work

We have introduced and publicly released the first, to our knowledge, knowledge graph for climate-aligned contracting, also demonstrating the application of a lawyer-in-the-loop pipeline for creating tools of this kind. Designed as a retrieval mechanism, retrieval ablation demonstrates that its entity layer significantly improves retrieval over dense and lexical baselines.

Potential applications of this graph include providing relevant information to practicing lawyers, as well as comparing instruments of private environmental governance to accepted standards, voluntary sustainability mechanisms, or public commitments.

A lightweight version of the retrieval interface is already released as an MCP server for integration with third-party LLMs, and as that tooling advances so too can this machine-readable climate-contracting knowledge graph advance with it.

Limitations

The principal limitation of this work is the availability of expert reviewers. We had hoped for three independent reviews at the annotation step but ultimately achieved only a single review of each chunk, so we cannot report inter-annotator agree-

ment. This scarcity of expert labels also left them too sparse to train a labelled entity extractor. As a workaround, we moved to binary entity extraction rather than labelled extraction and assigned types with an LLM. The fuzzy-merge decisions in Section 3.3 received the same single-reviewer assessment as the annotation step, and the results of that process are therefore also limited by the lack of inter-annotator review and full evaluation suite.

Similarly related to expert review, legal feedback on the results presented in Appendix A would have been an interesting extension of this project. As is, they serve only to demonstrate that these kinds of questions can be answered reasonably well, but we do not engage in a full conversation on how the legal community might receive them. This limitation is doubly challenging as experts would require not only expert legal knowledge, but also expert knowledge of the specific content used to create this KG. The same gap extends to the annotation process itself: we did not conduct qualitative interviews as to the annotators’ experience, including the usefulness of pre-seeded entities or annotation guidance, so that qualitative account is missing although acknowledged as potentially rich.

Expert time was also a constraint on maintained content. Only 18 of the clauses had been actively maintained, and only these carry the associated metadata, so we could not leverage that metadata for categorisation and exploration across the whole graph. Relatedly, we treated clauses as connected to one another in assessing the retrieval performance of the KG based on the practice areas or sectors they share. This treatment was a by-product of how previous maintenance had been performed, but more time and more access to legal expertise would have allowed for a curated test set based on what legal practitioners think should be related.

Finally, data availability is a structural limitation of the legal domain. Because the methods of private environmental governance are typically proprietary, future applications of this work will be constrained by access to appropriate legal data.

Acknowledgements

We would like to thank the legal staff at The Chancery Lane Project who were involved in annotating, metadata tagging, and legal guidance including Humzah Khan, Clarissa Lambourne, Simone Potter, Natasha Morgan, Meena Kamath, Becky Annison and Charlotte Goosey. We also received

annotation help from the excellent team at Lexis-Nexis.

We also thank the TCLP staff and the many pro bono lawyers and other professionals who have drafted, reviewed and maintained the clauses, glossary terms and guides since the project began. That body of work is the corpus on which this knowledge graph is built.

We further thank Philip Detje who prototyped the application of this work for tenders, and was involved in conversations surrounding this draft.

The authors are affiliated with The Chancery Lane Project, the non-profit whose clauses, glossary terms, and guides constitute the source corpus and the resource evaluated in this paper. The authors declare no other competing interests.

This work was completed with the aid of LLMs, including in code generation, report editing, and figure design.

We are grateful to funding received from the Patrick J. McGovern foundation which made this work possible.

References

- Garima Agrawal, Tharindu Kumarage, Zeyad Alghamdi, and Huan Liu. 2023. [Can knowledge graphs reduce hallucinations in LLMs?: A survey](#). In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL 2024)*, volume 1, pages 3947–3960.
- Julia Anna Bingler, Mathias Kraus, Markus Leippold, and Nicolas Webersinke. 2022. [Cheap talk and cherry-picking: What ClimateBert has to say on corporate climate risk disclosures](#). *Finance Research Letters*, 47:102776.
- Carbon Brief. 2026. [Project cosmos](#). <https://interactive.carbonbrief.org/cosmos/index.html>. Accessed June 25, 2026.
- Miguel Ceriani, Fiorela Ciroku, Alessandro Russo, Massimiliano Schembri, Fai Fung, Neha Mittal, Vito Trianni, and Andrea Giovanni Nuzzolese. 2026. [The climate change knowledge graph: Supporting climate services](#). <https://arxiv.org/pdf/2602.19786>. Accessed June 25, 2026.
- Ilias Chalkidis, Manos Fergadiotis, Prodromos Malakasiotis, Nikolaos Aletras, and Ion Androutsopoulos. 2020. [LEGAL-BERT: The muppets straight out of law school](#). In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 2898–2904.
- Gordon V. Cormack, Charles L. A. Clarke, and Stefan Büttcher. 2009. Reciprocal rank fusion outperforms

- Condorcet and individual rank learning methods. In *Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval*.
- Kaylan Dutia, Anne Sistema, Julie Saigusa, and Harrison Pim. 2025. Climate policy radar’s open knowledge graph. In *NeurIPS 2025 Workshop on Tackling Climate Change with Machine Learning*.
- Darren Edge, Ha Trinh, Newman Cheng, Joshua Bradley, Alex Chao, Apurva Mody, Steven Truitt, et al. 2024. [From local to global: A graph RAG approach to query-focused summarization](https://arxiv.org/pdf/2404.16130). <https://arxiv.org/pdf/2404.16130>. Accessed June 25, 2026.
- Michael Grüninger and Mark S. Fox. 1995. [Methodology for the design and evaluation of ontologies](#). In *Workshop on Basic Ontological Issues in Knowledge Sharing, IJCAI*, Montreal. Accessed July 14, 2026.
- Marvin Hofer, Daniel Obraczka, Alieh Saeedi, Hanna Köpcke, and Erhard Rahm. 2024. [Construction of knowledge graphs: Current state and challenges](#). *Information*, 15(8):509.
- Louis Leonard. 2020. [Under the radar: A coherent system of climate governance, driven by business](https://papers.ssrn.com/abstract=3598219). <https://papers.ssrn.com/abstract=3598219>. Accessed July 2, 2026.
- Nelson F. Liu, Kevin Lin, John Hewitt, Ashwin Paranjape, Michele Bevilacqua, Fabio Petroni, and Percy Liang. 2023. [Lost in the middle: How language models use long contexts](#). *Transactions of the Association for Computational Linguistics*, 12:157–173.
- Sasha Luccioni, Yacine Jernite, and Emma Strubell. 2024. [Power hungry processing: Watts driving the cost of AI deployment?](#) In *2024 ACM Conference on Fairness, Accountability, and Transparency (FAccT 2024)*, pages 85–99.
- Sasha Luccioni, Emma Strubell, and Kate Crawford. 2025. [From efficiency gains to rebound effects: The problem of Jevons’ paradox in AI’s polarized environmental debate](#). In *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency (FAccT 2025)*, volume 1, pages 76–88.
- Joel Niklaus, Ilias Chalkidis, and Matthias Stürmer. 2021. [Swiss-Judgment-Prediction: A multilingual legal judgment prediction benchmark](#). In *Proceedings of the 2021 Natural Legal Language Processing Workshop (NLLP 2021)*, pages 19–35.
- Shirui Pan, Linhao Luo, Yufei Wang, Chen Chen, Jipu Wang, and Xindong Wu. 2024. [Unifying large language models and knowledge graphs: A roadmap](#). *IEEE Transactions on Knowledge and Data Engineering*, 36(7):3580–3599.
- Ridho Reinanda, Edgar Meij, and Maarten De Rijke. 2020. [Knowledge graphs: An information retrieval perspective](#). *Foundations and Trends in Information Retrieval*, 14(4):1–159.
- Stephen Robertson and Hugo Zaragoza. 2009. [The probabilistic relevance framework: BM25 and beyond](#). *Foundations and Trends in Information Retrieval*, 3(4):333–389.
- Tobias Schimanski, Julia Bingler, Camilla Hyslop, Mathias Kraus, and Markus Leippold. 2023. [ClimateBERT-NetZero: Detecting and assessing net zero and reduction targets](#). In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP 2023)*, pages 15745–15756.
- Yi Tay, Mostafa Dehghani, Dara Bahri, and Donald Metzler. 2023. [Efficient transformers: A survey](#). *ACM Computing Surveys*, 55(6).
- Michael P. Vandenbergh. 2013. [Private environmental governance](#). *Cornell Law Review*, 99(1). Accessed June 25, 2026.
- Michael P. Vandenbergh and Bruce M. Johnson. 2021. [The role of private environmental governance in climate adaptation](#). *Frontiers in Climate*, 3:715368.
- Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. [Attention is all you need](#). In *Advances in Neural Information Processing Systems*. Accessed July 2, 2026.
- Yangyang Wang, Lan Fang, and Mingchun Jiang. 2024. [Optimal contract design for carbon emission reduction in a green supply chain: Monetary incentive vs. symbolic incentive](#). *Managerial and Decision Economics*, 45(5):3152–3162.
- Nicolas Webersinke, Mathias Kraus, Julia Anna Bingler, and Markus Leippold. 2021. [ClimateBert: A pretrained language model for climate-related text](#). *SSRN Electronic Journal*.
- Lingfeng Zhong, Jia Wu, Qian Li, Hao Peng, and Xindong Wu. 2024. [A comprehensive survey on automatic knowledge graph construction](#). *ACM Computing Surveys*, 56(4):1–62.

A Competency questions

Each question is one a practitioner might bring to the resource, paired with the graph operation that answers it and a representative result from the current build. The questions fall into four groups.

A.1 Finding related clauses

A1. Related clauses by shared entities *Question.* Which clauses are most closely related to a given clause on the basis of shared domain entities alone?

Answered by ranking clauses on the Jaccard overlap of the entities they share through CONTAINED_IN edges, with no embedding signal.

Querying “**The Net Zero Standard for Suppliers**” returns a coherent set of supply-chain and emissions clauses, led by “Target Product Carbon Footprint (Schedule for Consumer Goods Contracts)” and “Green Supplier Agreement Terms.”

A2. Substantively related clauses *Question. Which clauses are most related in substance, combining meaning with shared entities?*

Answered by reciprocal rank fusion ($k=60$) of embedding cosine similarity and entity overlap. Querying “**Carbon Neutral Agreement Clause**” returns semantic neighbours that need not share exact terminology, including “Ambitious Carbon Neutral Clause” and several green shareholders’ agreements.

A.2 Concept and definition lookup

A3. Concept lookup *Question. Which clauses mention a given concept?*

Answered by matching Entity nodes to clauses through CONTAINED_IN. “**Science Based Targets**” resolves to 31 clauses, “Scope 3 emissions” to 5, and “Net Zero” to 4, with distinct surface forms of a concept tracked separately.

A4. Defined term and glossary definition *Question. Where does a defined term appear, and what is its definition?*

Answered by traversing GlossaryTerm to Clause-Name through CONTAINED_IN. “**Net Zero Target**” returns its glossary definition together with the clauses that use it, such as “Board Paper Implementing Net Zero for SMEs.”

A.3 Filtering by taxonomy

A5. Taxonomy-scoped retrieval *Question. Which clauses match a combination of taxonomy filters?*

Answered by filtering on the slug-array properties of ClauseName. **Supply contracts whose climate solution is reducing greenhouse gas emissions** return two clauses: “Consolidated Sustainability Clauses for Supply Chain Contracts” and “Gain-Share Incentive for Reducing GHG Emissions.” When queried for a combination absent from the corpus, the same operation returns an empty set rather than a spurious match, so the graph reports coverage gaps faithfully.

A.4 Reference lookup and aggregation

A6. Reference lookup *Question. Which clauses invoke a given law or standard?*

Answered by REFERENCES_LAW and REFERENCES_STANDARD edges. The **Paris Agreement** is invoked by 45 clauses, the **GHG Protocol** by 32, and **ISO 14001** by 3.

A7. Reference aggregation *Question. Which instruments are most referenced across the corpus, and how do curated and machine-extracted references compare?*

Answered by aggregating reference edges by target and by provenance. The most-referenced instruments are the Paris Agreement (45), the GHG Protocol (32), and the Science Based Targets initiative (25), followed by a long tail of single-clause references. The large majority of these edges originate from the human-reviewed LSRF classifier rather than the original curated metadata.