Evolution of ESG-focused DLT Research: An NLP Analysis of the Literature

Anonymous ACL submission

Abstract

As Distributed Ledger Technologies (DLTs) rapidly evolve, their impacts extend beyond technology, influencing environmental and societal aspects. This evolution has increased publications, making manual literature analysis increasingly challenging. We address this with a Natural Language Processing (NLP)-based systematic literature review method to explore the intersection of Distributed Ledger Technology (DLT) with its Environmental, Social, and Governance (ESG) aspects. Our approach involves building and refining a directed citation network from 107 seed papers to a corpus of 24,539 publications and fine-tuning a transformer-based language model for Named Entity Recognition (NER) on DLT and ESG domains. Applying this model, we distilled the corpus to 505 key publications, enabling an inaugural literature review and temporal graph analysis of DLT's evolution in ESG contexts. Our contribution include an adaptable and scalable NLP-driven systematic literature review methodology and a unique NER dataset of 54,808 entities, tailored for DLT and ESG research. Our inaugural literature review demonstrates their applicability and effectiveness in analyzing DLT's evolution and impacts, proving invaluable for stakeholders in the DLT domain.

1 Introduction

Emerging technologies have seen increasing scrutiny in terms of energy consumption and broader ecological impacts, encompassing vital resources like water, precious metals, and synthetic compounds (Platt et al., 2021; Simone et al., 2022). This shift towards environmental consciousness emphasizes the need to evaluate technological advancements through their ecological footprint, including DLT. DLT promises record immutability and decentralization but faces challenges like high energy consumption in certain consensus algorithms, such as Bitcoin's Proof of Work (PoW) (Nakamoto, 2008), aimed to effectively prevent attackers from pretending to be many users simultaneously to increase their weight in the network, known as Sybil attacks. Therefore, DLT's advancements in security and immutability, alongside its complex and evolving applications, necessitate a sophisticated approach for analysis.

In this context, NLP, a sub-field of Artificial Intelligence (AI) and linguistics, emerges as a facilitator to delve into the growing number of publications in DLT, from academic articles to whitepapers. NLP focuses on certain human-related language tasks such as Question Answering (QA), NER, and text classification, among others. In this paper, we use NER to identify specific entities within the corpus of our dataset to illuminate gradual shifts in research emphasis and application of DLT. We build upon existing work that taxonomizes DLT to identify the entities. Our starting point is the hierarchical taxonomy of (Tasca and Tessone, 2019). Unlike previous systematic literature reviews that rely on citation measures and analysis of abstracts and keywords, our approach delves into the text of the body of the publications. This enables us to detect thematic shifts in key areas of research and industry publications (e.g., whitepapers) within the DLT field by mapping the publications' tokens to the components of the hierarchical taxonomy of DLT from (Tasca and Tessone, 2019) (see an example of mapping in Figure 2).

Our research has the following contributions:

1. A curated NER dataset composed of 54,808 named entities (see Table 2) for twelve DLT's taxonomy categories in the context of ESG (see Figure 2's a). To the best of our knowledge, this is the first NER dataset explicitly designed for DLT.[1]

2. A methodology and framework for executing

[1]The dataset will be made available

a NLP-driven systematic literature review at the intersection of domains, in this case, DLT and ESG research.[2]

3. Conducting what we believe is the first NLP-driven systematic literature for the DLT field that places a special emphasis on ESG aspects.

Additionally, our work represents a step for future research directions to improve further automated systematic literature review processes at scale, capable of capturing the intrinsic dependencies and evolution of concepts related to the intersection of fields.

2 Related work

Previous literature reviews have extensively explored blockchain applications in various sectors (Casino et al., 2019; Zheng et al., 2018). These reviews, however, differ in scope and depth compared to our systematic review, particularly in terms of article quantity and the manual nature of their analyses.

Studies have also focused on blockchain's role in decentralization and privacy, particularly in IoT (Conoscenti et al., 2016), and analyzed trends of centralization in decentralized systems like Bitcoin and Ethereum (Sai et al., 2021). (Spychiger et al., 2021) deconstructed 107 blockchain technologies using a specific taxonomy, emphasizing consensus mechanisms and cryptographic primitives. Our work, in contrast, provides a broader perspective on the evolution of DLT, including its ESG implications.

In the context of ESG, (Bilal et al., 2014; Mengelkamp et al., 2018; Poberezhna, 2018; Schulz and Feist, 2021; Wu et al., 2022; Jiang et al., 2022) have explored blockchain's potential in energy management, environmental sustainability, and transparent reporting. Our study extends these approaches by examining the intersection of ESG and DLT through a literature analysis.

Regarding NLP applications, studies have shown the use of advanced techniques for automated ESG scoring (Alik Sokolov et al., 2021) and opinion summarization (Dubey et al., 2023). Outside the DLT field, systematic literature reviews NLP-driven methodologies, such as in medical genomics, have been conducted (Alsheikh et al., 2022). These studies used database term searches and NLP models for abstract-based filtering, differing from our approach of building a corpus through directed citation graphs and full-text filtering using NLP.

[2]The repository with the code will be made available

3 Methodology

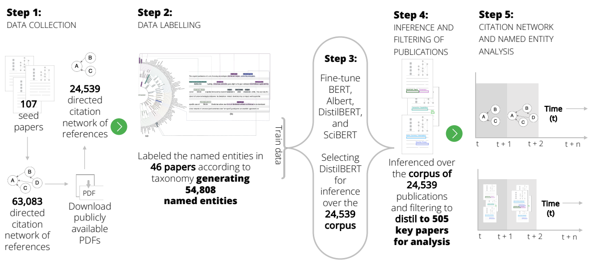


Figure 1: Methodology for the systematic literature review of ESG/DLT publications.

Ontologies, specifically hierarchical taxonomies, are pivotal in developing NER datasets for text mining (Spasic et al., 2005; Huang et al., 2020; Nabi et al., 2022; Chang et al., 2016; Mcentire et al., 2016; Alsheikh et al., 2022). For example, the GENIA corpus (Kim et al., 2003), a NER dataset of 2,000 biological abstracts, employs the GENIA ontology's hierarchical tree structure of 47 biological entities, including top-level categories like biological source, substance, and others, to facilitate text mining in biomedical literature. Similarly, the Human Phenotype Ontology is used for creating and expanding NER datasets in biomedicine (Lobo et al., 2017; Huang et al., 2020). (Alsheikh et al., 2022; Chang et al., 2016; Mcentire et al., 2016) further demonstrate the use of ontology-based NER datasets for domain-specific literature text mining.

Learning from these biomedical field precedents, our methodology for NLP-based text mining and filtering in the DLT field employs a hierarchical taxonomy (Tasca and Tessone, 2019) to annotate a NER dataset from 46 systematically reviewed publications of DLT's sustainability (Eigelshoven et al., 2020). Therefore, we demonstrate the generalizability and transferability of these biomedical field precedents by successfully applying some of their elements in our methodology, demonstrating their versatility across different domains.

Additionally, unlike keyword database searches, we construct our corpus using a directed citation graph from references (*citing to*) of 107 seed publications in the ESG/DLT domain intersection, fine-tuning a transformer-based language model for corpus filtering. We also perform temporal graph analysis to understand the evolution at the ESG and DLT intersection. Figure 1 summarizes our methodology.

4 Data Collection

The seed papers for our citation network were selected from two sources:

1. 89 papers from (Eigelshoven et al., 2020), reviewing sustainability in popular DLT consensus algorithms.

2. 18 recent publications (2018-2022) with at least three citations each, chosen to update the corpus with more current research relevant to DLT/ESG (Platt et al., 2021; Kohli et al., 2022; Nair et al., 2020; Ante and Fiedler, 2021; Sedlmeir et al., 2020; Fernando and Saravannan, 2021; Masood and Faridi, 2018; Ghosh and Das, 2020; Eshani et al., 2021; Cole and Cheng, 2018; Lucey et al., 2021; Sapkota and Grobys, 2020; Bada et al., 2021; Denisova et al., 2019; Schinckus et al., 2020; Sedlmeir et al., 2021; Powell et al., 2021; Alofi et al., 2022).

The key benefit of using seed papers to build a citation network for a systematic literature review is the ease of expanding and updating the literature review by adjusting the number of seed papers.

We limited our citation network to references made by the seed papers, ensuring thematic relevance to DLT/ESG. We restricted the expansion to two levels of references to avoid divergence from the theme. This led to a network with over 63,083 publications, from which 24,539 publicly available PDFs were retrieved using Semantic Scholar's database (see Figure 3's a).

4.1 Labeling

We manually annotated 46 papers using the brat tool (Stenetorp et al., 2012), following the taxonomy framework of (Tasca and Tessone, 2019). This taxonomy provides a hierarchical structure of blockchain components, with each principal component (e.g., `Consensus`) divided into subcomponents (e.g., `Gossiping`) and further into sub-sub-components if needed (e.g., `Local`). We introduced categories like `Blockchain_Name` to identify specific blockchains and the initial definition of `Security_Privacy` was expanded to label security threats (`Sybil attack`, `51% attack`, etc.) while a `Miscellaneous` category was added for ambiguous contexts (see Figure 2, Table 1), following the example of the CoNLL-2003 dataset for a similar category (Tjong et al., 2003). We further extended (Tasca and Tessone, 2019)'s taxonomy to identify sustainability-related concepts referred to in the ESG criterion (see Figure 2).

Group entities	Description
`Blockchain_Name`	The name of a blockchain system (E.g., Bitcoin, Ethereum, XRP Ledger), but also including other types of DLTs, such as Hedera, IOTA
`Consensus`	Rules and mechanisms to ensure the immutability of transaction records (E.g., Proof of Work, Proof of Stake, Blockchain, Hashgraph).
`Identifiers`	Information related to the token names, creators, purpose, and different names of a blockchain (E.g., Satoshi Nakamoto, Ripple, USDC, USDT).
`Security_Privacy`	Cryptographic methods to ensure data privacy and encryption in a blockchain ecosystem.
`ESG`	Entities relevant to the Environmental, Social, and Governance issues.
`Transaction_Capabilities`	Information related to the details of transactions, such as Data Structure in the Blockheader, Transaction Model, Server Storage, Block Storage, and Limits to Scalability.
`ChargingAndRewardingSystem`	Cost models for the operation and maintenance of blockchain systems.
`Extensibility`	Capabilities of Interoperability, Intraoperability, Governance, and Script Language of a blockchain ecosystem.
`Identity_Management`	Attributes to identify participants and their system access level.
`Native_Currency_Tokenisation`	Asset classes for transactions within a blockchain system (E.g., BTC, ETH, XRP, HBAR).
`Codebase`	Coding Language, Code License, and Software Architecture of a blockchain ecosystem (E.g. Solidity, Rust, MIT License, Apache License).
`Miscellaneous`	Miscellaneous entities that are ambiguous in a given context and are relevant for the DLT topic but are not captured by any of the above categories.

Table 1: List of 12 *ESG/DLT* groups of entity types based on the taxonomy from (Tasca and Tessone, 2017)

4.2 Text analysis/language processing

The label hierarchy within the taxonomy was pruned for class balance, where specific labels like `PoW` were replaced by broader categories like `Consensus` to maintain focus on primary taxonomy components (Figure 2). To improve NER model performance, which is sensitive to label consistency (Zeng et al., 2021; Jeong and Kang, 2023), we employed a systematic process for enhancing interlabeler consistency. This involved correcting inconsistent labeling of entities, such as "Sybil attack" sometimes categorized as `Consensus` and other times as `Security_Privacy`, following each labeler's approval and using programmatic cleaning to ensure consistency for non-context-dependent labels.

We applied text resampling for overlapping named entities that could fit into multiple categories, such as belonging to both `Blockchain_Name` and `Native_Currency_Tokenisation`. This process involves duplicating text and assigning distinct entities to each copy, thereby enhancing

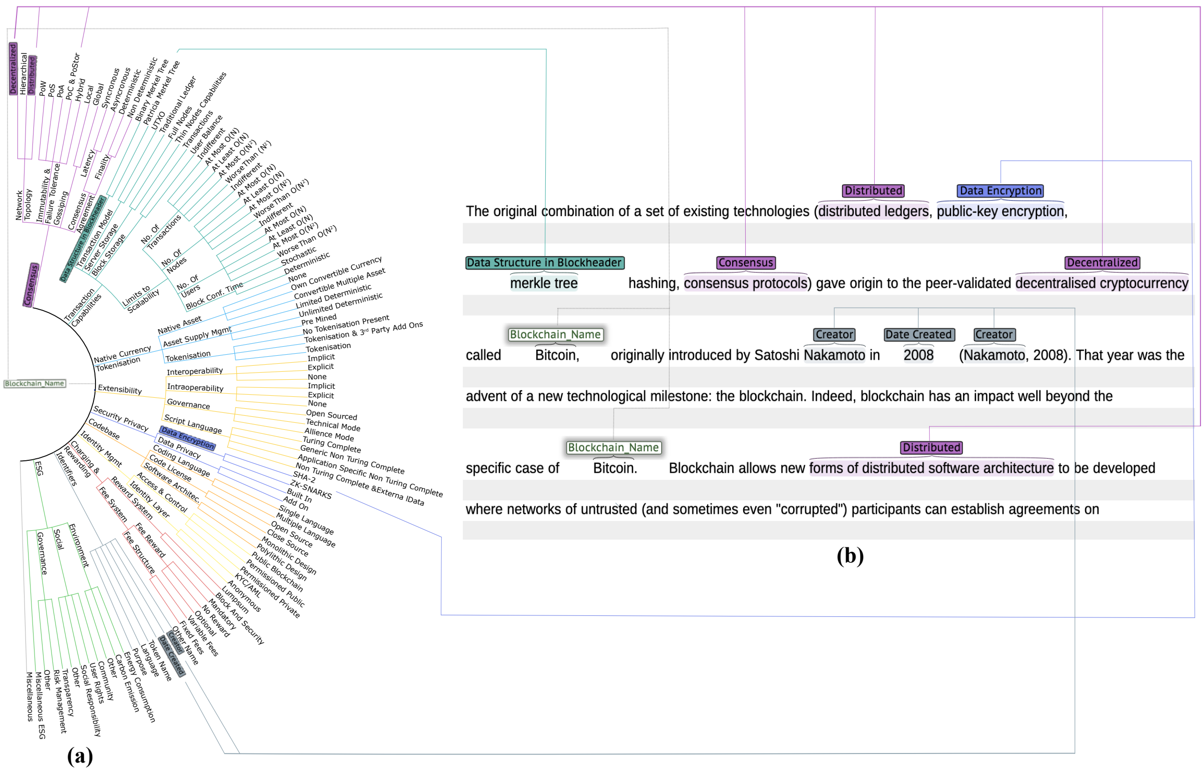


Figure 2: (a) The taxonomy of (Tasca and Tessone, 2019) extended with Blockchain_Name, ESG, and Miscellaneous (see 4.1) for the purpose of this research. (b) Example of parsed text with the taxonomy label associated with a span of text labeled. The labels used in the paragraph are highlighted in the taxonomy tree.

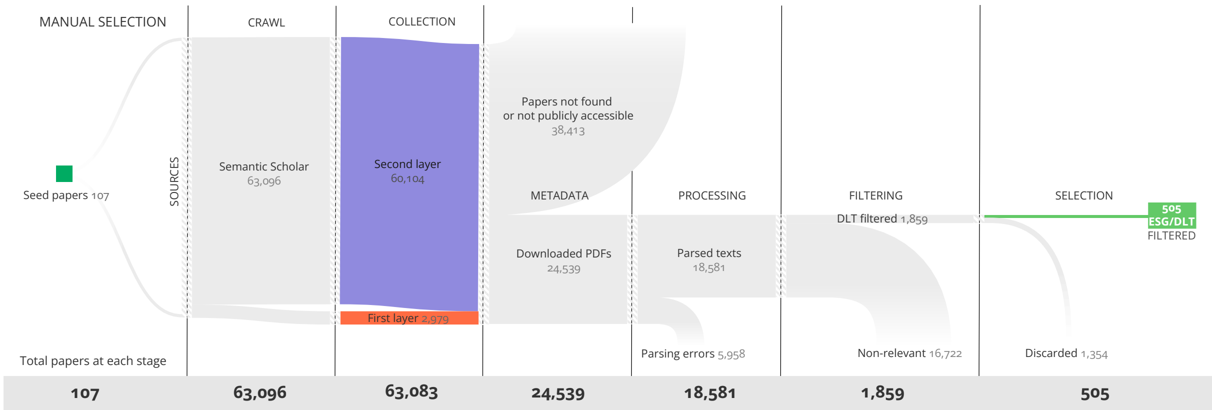
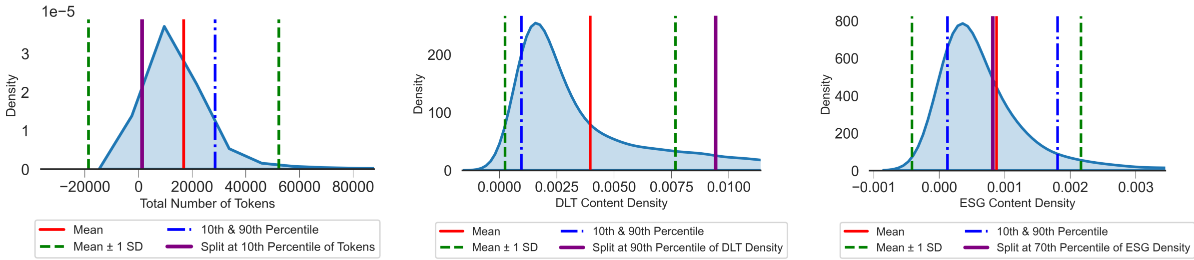


Figure 3: Processing pipeline for the collection and filtering of papers in the review. The total number of papers present at each stage of processing is shown. See Table 1 for the description of the labels in the corpus.

the capture of rare entities. This resampling strategy is beneficial, especially for datasets of modest size (Wang and Wang, 2022), improving model performance by accommodating diverse entity categories. Additionally, the duplication of training data has been found beneficial in enhancing a language model's ability to learn from limited examples (Muennighoff et al., 2023).

4.3 Mapping taxonomies using NLP

Recent advancements in NLP, including data acquisition (Bowman et al., 2015; Rajpurkar et al., 2016), model architecture development (Sutskever et al., 2014; Vaswani et al., 2017), and large-scale pre-training (Peters et al.; Devlin et al., 2019; Liu et al., 2019; Hoffmann et al., 2022; Radford et al., 2019; Touvron et al., 2023), have significantly propelled the field forward. For example, we considered Large Language Models (LLMs) for



(a) Filter for papers below the 10th percentile of total numbers of tokens.

(b) Filter for papers with a DLT content density above the 90th percentile.

(c) Filter for papers with an ESG content density above the 70th percentile.

Figure 4: Steps for the percentile-based filtering

NER tasks, inspired by the effectiveness of models like ChatGPT and GPT4 in zero-shot and few-shot learning scenarios (Li et al., 2023; Hu et al., 2023). However, despite their capabilities, (Li et al., 2023; Hu et al., 2023) noted that domain-specific NER tasks often perform better with supervised learning models than with current LLMs. Therefore, we adopted a supervised learning approach, fine-tuning transformer-based pre-trained language models such as BERT (Devlin et al., 2019), Albert(Lan et al.), DistilBERT (Sanh et al., 2019), and SciBERT (Beltagy et al., 2019). Our selection criterion for the final model was based on its performance in our NER task and efficiency at inference.

4.4 Filtering

We applied a percentile-based filtering process to the corpus of publications analyzed by our fine-tuned model. This method selects publications with substantial DLT and ESG content, using a threshold percentile to exclude marginally relevant papers. Seed papers were included to maintain foundational references. The filtering is represented as:

$$F = \{P_i : D(P_i) > T\} \cup S \quad (1)$$

Where F is the final set of papers, P_i is an individual paper, $D(P_i)$ is the ESG and DLT content density of a paper, T is the threshold percentile, and S is the set of seed papers. Content density $D(P_i)$ for each paper is calculated by the ratio of the number of DLT and ESG relevant named entities (N_{DLT} and N_{ESG}) to the total number of tokens $N(P_i)$ (Equation 2):

$$D(P_i) = \frac{N_{DLT}(P_i) + N_{ESG}(P_i)}{N(P_i)} \quad (2)$$

Our filtering methodology involved:

1. Excluding papers below the 10th percentile (see Figure 4a) in the total token count to avoid distortions due to PDF-to-text conversion issues or unusually short papers (e.g., below 100 tokens).

2. Computing DLT content density and retaining papers above the 90th percentile (see Figure 4b), ensuring a strong focus on DLT topics.

3. Filtering for at least the 70th percentile (see Figure 4c) in ESG content density to confirm relevance to ESG.

Finally, we manually reviewed the filtered publications to validate the accuracy of their ESG/DLT content density and relevance.

4.5 Network graphs

We analyzed the citation network, represented as $G(V, E)$ with papers as vertices V and citations as edges E. Using $G(V, E)$, we did temporal graph analysis with one-year time windows $W_1, W_2, \ldots, W_n$, as per the rolling window approach in (Hoadley et al., 2021; Steer et al., 2020, 2023). For each window W_i, we created a subgraph $G_i(V_i, E_i)$ and applied the HITS algorithm (Kleinberg, 2011) to determine temporal shifts in significant citations within the network.

Furthermore, we tracked the evolution of named entities in the citation network. Using lemmatization and programmatic grouping, we consolidated variations of similar entities (e.g., all forms of "Proof-of-Work" were unified under "PoW") to capture changes in entity prevalence accurately.

5 Evaluation

Figure 3 details the collection and filtering stages of our systematic literature review. The next sections provide more details of the results after applying our methodology (Figure 1).

Entity Category	Number of Entities
Blockchain_Name	5,358
Consensus	25,378
Transaction_Capabilities	4,729
Native_Currency_Tokenisation	2,671
Extensibility	1,752
Security_Privacy	4,838
Codebase	1,339
Identity_Management	1,305
ChargingAndRewardingSystem	1,531
Identifiers	1,511
ESG	3,468
Miscellaneous	928

Table 2: Number of named entities for each category in the dataset.

Text: *In this paper, the PoW consensus algorithm used in blockchains are analyzed in terms of difficulty, hash count, and probability of successful mining.* **Output:** *In this paper, the* $\langle Consensus \rangle$ *consensus algorithm used in* $\langle Identifiers \rangle$ *is analyzed in terms of* $\langle Consensus \rangle$, $\langle TransactionCapabilities \rangle$, *and* $\langle TransactionCapabilities \rangle$.
Text: *Generally, as the difficulty of mining increases, the mining time becomes longer because the target value must be found lower.* **Output:** *Generally, as the* $\langle NativeCurrencyTokenisation \rangle$, *the* $\langle ESG \rangle$ *because the target value must be found lower.*
Text: *As a result, the amount of electrical energy needed to process the work is immense.* **Output:** *As a result, the* $\langle ESG \rangle$ *to process the work is immense..*
Text: *It provides a distributed, immutable, transparent, secure, and auditable ledger.* **Output:** *It provides a* $\langle Consensus \rangle$, $\langle Consensus \rangle$, *transparent,* $\langle ESG \rangle$.
Text: *Blockchain was first introduced with the creation of Bitcoin back in 2008.* **Output:** *Blockchain was first introduced with the creation of* $\langle BlockchainName \rangle$ *back in* $\langle Identifiers \rangle$.

Table 3: Training examples for *ESG/DLT* labeling task.

5.1 Taxonomy labelling result

Our NER dataset organizes 54,808 named entities into a tree structure with 136 labels under 12 top-level categories (Figure 2, Table 2). This structure facilitated targeted analysis in our study. Table 3 provides examples from the dataset.

5.2 NLP result

We fine-tuned four models – bert-base-cased[3], albert-base-v2[4], distilbert-base-cased[5], and allenai/scibert_scivocab_cased[6] – using 5-fold cross-validation according to the titles of the publications to avoid a publication's data appearing in both the training and test set for each fold (see Table 3 for some samples of training data). Each model underwent 100 training epochs, 20 epochs per fold, with a learning rate of 5×10^{-5}, a training batch size of 32, and a validation batch size of 64. The maximum sequence length was set at 256 tokens.

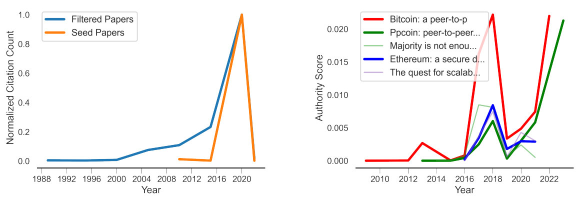
[3]https://huggingface.co/bert-base-cased
[4]https://huggingface.co/albert-base-v2
[5]https://huggingface.co/distilbert-base-cased
[6]https://huggingface.co/allenai/scibert_scivocab_cased

Model	Metrics				
	Fold	Precision	Recall	F1	Accuracy
BERT	1	0.43342	0.42502	0.42918	0.96443
	2	0.55149	0.58924	0.56974	0.96754
	3	0.57820	0.55315	0.56510	0.94566
	4	0.55809	0.58072	0.56918	0.94728
	5	0.58671	0.60786	0.59710	0.96414
	Mean	0.54158	0.55120	0.54606	0.95781
Albert	1	0.50650	0.34185	0.40820	0.96984
	2	0.57694	0.53416	0.55473	0.97207
	3	0.53164	0.43772	0.48013	0.95174
	4	0.52687	0.55281	0.53953	0.95577
	5	0.57680	0.60863	0.59229	0.97286
	Mean	0.54375	0.49503	0.51498	0.96446
DistilBERT	1	0.45704	0.38093	0.41553	0.96607
	2	0.55631	0.55364	0.55497	0.96713
	3	0.57937	0.54156	0.55983	0.94562
	4	0.55962	0.58406	0.57158	0.94747
	5	0.58537	0.61668	0.60062	0.96414
	Mean	0.54754	0.53537	0.54051	0.95809
SciBERT	1	0.46651	0.46432	0.46542	0.96983
	2	0.52566	0.61680	0.56760	0.96649
	3	0.55980	0.62262	0.58954	0.94162
	4	0.54930	0.64023	0.59129	0.94435
	5	0.57721	0.63896	0.60652	0.96544
	Mean	0.53570	0.59659	0.56407	0.95755

Table 4: Performance results after fine-tuning for BERT, Albert, DistilBERT, and SciBERT with the ESG/DLT NER dataset.

The evaluation results (Table 4) showed that SciBERT and BERT had the best performance. However, DistilBERT's efficiency made it more suitable for our large corpus of 24,539 publications. DistilBERT, being 60% faster than BERT, and likewise SciBERT, at inference and achieving 97% (Sanh et al., 2019) of BERT's performance, was selected for its balance between effectiveness and efficiency.

5.3 Citation network



(a) Filtered citation network evolution and seed papers using 1-year rolling window.

(b) Authority score evolution for top 5 papers in the citation network.

Figure 5: Temporal graph analysis.

In our citation network analysis, Nakamoto's Bitcoin whitepaper (Nakamoto, 2008) emerges as a central node (Figure 6), emphasizing its foundational impact on DLT research (Yli-Huumo et al., 2016; Spychiger et al., 2021). This network also prominently features other key innovations, including Hashcash (Back, 2002) as a precursor to Bitcoin, Ethereum's introduction of Smart Contracts in 2014 (Buterin), and PPCoin's 2012 development of Proof of Stake (PoS) (King and Nadal, 2012), indicating significant milestones in DLT evolution.

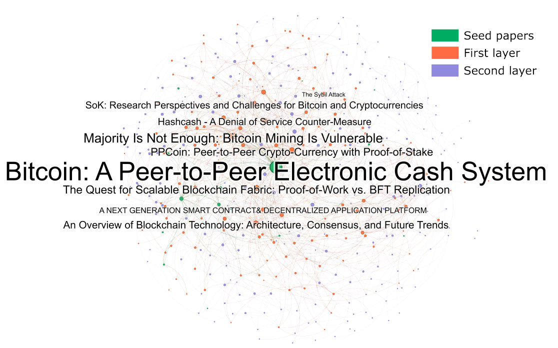


Figure 6: Density filtered network by layers and by the distribution of authority scores.

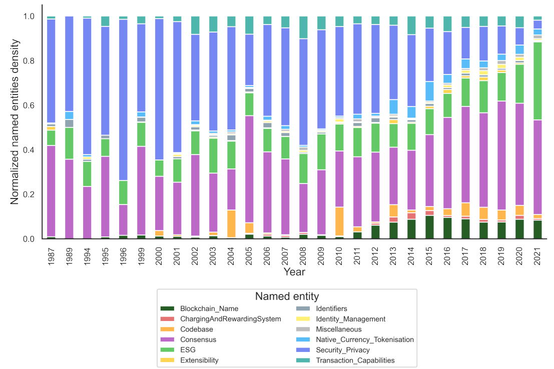
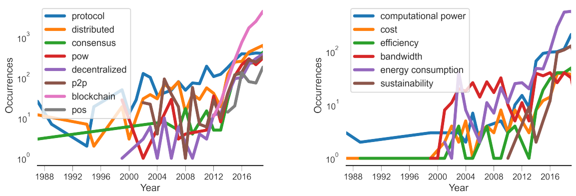


Figure 7: Publications showing the normalized ratio of labels for each part of a branch of the taxonomy.

Temporal graph analysis of the network (Figure 5a) reveals a publication surge between 2005 and 2010, aligning with Bitcoin's release and its subsequent influence on diverse DLT research areas, notably in Consensus mechanisms (Figure 7). Post-2010, the network saw a marked increase in citations, especially after 2015, reflecting the impact of seminal works like PPCoin and Ethereum's whitepapers (King and Nadal, 2012; Buterin). The decline in citations after 2020 is discussed in subsection 6.1.

6 Discussion



(a) Consensus' named entities evolution

(b) ESG's named entities evolution

Figure 8: Named entities evolution in the citation network after applying lemmatization.

Recent developments in LLMs, like Google DeepMind's Gemini[7], highlight the significance of our work in systematic literature review. Gemini's demonstration of a systematic literature review, where terms like "Chip" and "CRISPR-Cas9" are searched in publications' titles and abstracts to filter them[8] is akin to our methodology (section 3) of applying NER for field-specific filtering of literature, demonstrating the generalizability and applicability of our approach. However, Gemini faces limitations like potential hallucinations that could undermine its filtering of publications. On the other hand, despite that supervised learning approaches outperform current LLMs for NER tasks (Li et al., 2023; Hu et al., 2023), Gemini shows the potential of LLMs in few-shot learning for NER tasks. Additionally, as a commercial product, Gemini has limited accessibility, and its higher-performing models (Pro and Ultra versions) are not widely available. In contrast, our methodology (section 3) leverages a domain-specific labeled NER dataset and a fine-tuned language model to analyze full-text publications, not just their title and abstracts. This approach enhances the accuracy and depth of literature reviews. More importantly, our openly available methodology and NER dataset provide the NLP community and others the opportunity to build upon and improve systematic literature review processes at scale, ensuring more reliable filtered results.

In terms of our analysis of the citation network, Figure 5a and Figure 6 indicate that foundational publications, particularly those introducing Bitcoin, Ethereum, and other early blockchain technologies, have significantly influenced subsequent research. This is evident from their high citation counts and anchoring positions in the network (see Figure 6 and Figure 5b).

The increasing ESG content density within DLT research (Figure 7) highlights a shift in thematic interests, from an early focus on security and privacy driving adoption to a growing emphasis on tokenization, efficient and secure consensus algorithms (e.g., Byzantine Fault Tolerance), and blockchain architectures. Key historical developments include the emergence of PoW in the late

[7] https://www.youtube.com/watch?v=sPiOP_CB54A

[8] See https://youtu.be/sPiOP_CB54A?feature=shared&t=64 for the prompt used in the demonstration

1990s, as exemplified by (Back, 2002), and the rise of PoS-related entities around 2012, following (King and Nadal, 2012)'s work (Figure 8a).

Recent years have seen an academic shift towards ESG and consensus-related terms, reflecting an evolving focus on energy-efficient distributed systems, decentralization, and sustainable blockchain research (Figure 8b). This shift, coupled with the increasing prominence of terms like "decentralization", "blockchain", and "sustainability", underscores a multidisciplinary approach in the field. The sustained interest in PoW, along with explorations into PoS and other consensus mechanisms, highlights the field's adaptability to environmental and scalability challenges (Figure 8a). This evolution reflects a balance between technological advancements and societal ESG imperatives, demonstrating the academic community's holistic and forward-thinking approach to addressing blockchain technology's challenges and opportunities.

6.1 Limitations

Our literature review faces limitations, including potential biases in seed paper selection and a time lag in capturing recent publications, which may affect the comprehensiveness of our analysis. For instance, the choice of XRP's 2018 whitepaper (Chase and MacBrough, 2018) over the more cited 2014 edition (Schwartz et al.) could underestimate its influence in the citation network. Similarly, recent works like the 2018 Hedera whitepaper (Baird et al., 2018) are omitted due to unavailable citation data.

The retrospective approach of building the citation network predominantly from pre-2020 seed papers introduces a bias toward older publications, potentially overlooking newer research yet to achieve recognition (Figure 5a). While our methodology could theoretically filter *citations to* seed papers based on content density, our review focused solely on references within the seed papers, possibly limiting the thematic breadth.

A significant constraint of this study is the reliance on publicly available research. Despite starting with an extensive citation network of 63,083 references (Figure 3), the analysis was limited to 24,539 publications with accessible full texts, highlighting the challenges of limited public access to some academic publications. This limitation points to the need for broader accessibility in research, especially in rapidly evolving fields like DLT. On the other hand, we acknowledge the growing importance of non-traditional literature, such as whitepapers and industry publications, in offering more inclusive access to technological developments in DLT.

6.2 Future work

Future work, as outlined in subsection 6.1, should focus on integrating metadata from different whitepaper versions, like XRP's 2014 and 2018 editions (Schwartz et al.; Chase and MacBrough, 2018), and sourcing metadata from alternative databases for publications with missing information, such as Hedera's whitepaper (Baird et al., 2018).

Further research should also include regular updates to the taxonomy's named entity categories (refer to Table 1), expanding training data by annotating more seed papers, and exploring various language model architectures.

7 Conclusion

The expanding scientific corpus and rising significance of non-traditional literature, including whitepapers and academic preprints, emphasize the growing need for assisted analytical methods. Our research demonstrates the efficacy of using NLP for conducting systematic literature reviews on a large scale, particularly within the rapidly evolving DLT field.

Our key contributions include the creation of the first NER dataset focused on DLT and ESG and a scalable and adaptable NLP-driven systematic literature review methodology. Additionally, we have conducted an inaugural systematic literature review using this dataset and methodology, demonstrating their practical applicability and effectiveness in analyzing DLT's technological evolution and impacts, serving as valuable resources for researchers, policymakers, and practitioners.

Moreover, this research represents a step toward improving automated literature review processes. Compared to commercial options, our openly available methodology and NER dataset allow the NLP community and researchers in related fields to build upon and improve systematic literature review processes at scale to meet evolving research needs.

References

Alik Sokolov, Jonathan Mostovoy, Jack Ding, and Luis Seco. 2021. Building Machine Learning Systems for

Automated ESG Scoring. *The Journal of Impact and ESG Investing*, 1(3):39–50.

Akram Alofi, Mahmoud A Bokhari, Rami Bahsoon, and Robert Hendley. 2022. Optimizing the Energy Consumption of Blockchain-Based Systems Using Evolutionary Algorithms: A New Problem Formulation. *IEEE Transactions on Sustainable Computing*, 7(4):910–922.

Ammar J. Alsheikh, Sabrina Wollenhaupt, Emily A. King, Jonas Reeb, Sujana Ghosh, Lindsay R. Stolzenburg, Saleh Tamim, Jozef Lazar, J. Wade Davis, and Howard J. Jacob. 2022. The landscape of GWAS validation; systematic review identifying 309 validated non-coding variants across 130 human diseases. *BMC Medical Genomics 2022 15:1*, 15(1):1–21.

Lennart Ante and Ingo Fiedler. 2021. Bitcoin's Energy Consumption and Social Costs in Relation to its Capacity as a Settlement Layer. *SSRN Electronic Journal*.

Adam Back. 2002. Hashcash - A Denial of Service Counter-Measure. *Http://Www.Hashcash.Org/Papers/Hashcash.Pdf*, (August):1–10.

Abigael Okikijesu Bada, Amalia Damianou, Constantinos Marios Angelopoulos, and Vasilios Katos. 2021. Towards a Green Blockchain: A Review of Consensus Mechanisms and their Energy Consumption. *2021 17th International Conference on Distributed Computing in Sensor Systems (DCOSS)*, pages 503–511.

Leemon Baird, Mance Harmon, and Paul Madsen. 2018. Hedera: A Public Hashgraph Network & Governing Council.

Iz Beltagy, Kyle Lo, and Arman Cohan. 2019. SciBERT: A Pretrained Language Model for Scientific Text. *EMNLP-IJCNLP 2019 - 2019 Conference on Empirical Methods in Natural Language Processing and 9th International Joint Conference on Natural Language Processing, Proceedings of the Conference*, pages 3615–3620.

Kashif Bilal, Saif Ur Rehman Malik, Osman Khalid, Abdul Hameed, Enrique Alvarez, Vidura Wijaysekara, Rizwana Irfan, Sarjan Shrestha, Debjyoti Dwivedy, Mazhar Ali, Usman Shahid Khan, Assad Abbas, Nauman Jalil, and Samee U Khan. 2014. A taxonomy and survey on Green Data Center Networks. *Future Generation Computer Systems*, 36:189–208.

Samuel R Bowman, Gabor Angeli, Christopher Potts, and Christopher D Manning. 2015. A large annotated corpus for learning natural language inference. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 632–642, Lisbon, Portugal. Association for Computational Linguistics.

Vitalik Buterin. Ethereum: A Next-Generation Smart Contract and Decentralized Application Platform.

Fran Casino, Thomas K Dasaklis, and Constantinos Patsakis. 2019. A systematic literature review of blockchain-based applications: Current status, classification and open issues.

Meiping Chang, Man Chang, Jane Z. Reed, David Milward, Jinghai James Xu, and Wendy D. Cornell. 2016. Developing timely insights into comparative effectiveness research with a text-mining pipeline. *Drug Discovery Today*, 21(3):473–480.

Brad Chase and Ethan MacBrough. 2018. Analysis of the XRP Ledger Consensus Protocol.

Ryan Cole and Liang Cheng. 2018. Modeling the Energy Consumption of Blockchain Consensus Algorithms. *2018 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)*, pages 1691–1696.

Marco Conoscenti, Antonio Vetro, and Juan Carlos De Martin. 2016. Blockchain for the Internet of Things: A systematic literature review. In *Proceedings of IEEE/ACS International Conference on Computer Systems and Applications, AICCSA*.

Valeriia Denisova, Alexey Mikhaylov, and Evgeny Lopatin. 2019. BLOCKCHAIN INFRASTRUCTURE AND GROWTH OF GLOBAL POWER CONSUMPTION. *International Journal of Energy Economics and Policy*, 9(4):22–29.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. {BERT}: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North {A}merican Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Akash Dubey, Kartik Jain, and K Kalaiselvi. 2023. Smart Patient Records using NLP and Blockchain. In *Proceedings - 5th International Conference on Smart Systems and Inventive Technology, ICSSIT 2023*, pages 663–667. Institute of Electrical and Electronics Engineers (IEEE).

Felix Eigelshoven, André Ullrich, and Benedict Bender. 2020. Public Blockchain – a Systematic Literature Review on the Sustainability of Consensus Algorithms. *Twenty-Eigth European Conference on Information Systems (ECIS2020), Marrakesh, Morocco.*, (June):1–17.

Ghosh Eshani, Das Rajdeep, Raj Shubhankar, and Das Baisakhi. 2021. An Analysis of Energy Consumption of Blockchain Mining and Techniques to Overcome It. *Lecture Notes on Data Engineering and Communications Technologies*, 62:783–792.

Yudi Fernando and Rubakanthan Saravannan. 2021. Blockchain Technology: Energy Efficiency and Ethical Compliance. *Journal of Governance and Integrity*, 4(2):88–95.

Eshani Ghosh and Baisakhi Das. 2020. A Study on the Issue of Blockchain's Energy Consumption. *Advances in Intelligent Systems and Computing*, 1065:63–75.

D Hoadley, M Bartolo, R Chesterman, A Faus, W Hernandez, B Kultys, A P Moore, E Nemsic, N Roche, J Shangguan, B Steer, K Tylinski, and N West. 2021. A Global Community of Courts? Modelling the Use of Persuasive Authority as a Complex Network. *Frontiers in Physics*, 9:665719.

Jordan Hoffmann, Sebastian Borgeaud, Arthur Mensch, Elena Buchatskaya, Trevor Cai, Eliza Rutherford, Diego de Las Casas, Lisa Anne Hendricks, Johannes Welbl, Aidan Clark, Tom Hennigan, Eric Noland, Katie Millican, George van den Driessche, Bogdan Damoc, Aurelia Guy, Simon Osindero, Karen Simonyan, Erich Elsen, Jack W Rae, Oriol Vinyals, and Laurent Sifre. 2022. Training Compute-Optimal Large Language Models.

Yan Hu, Iqra Ameer, Xu Zuo, Xueqing Peng, Yujia Zhou, Zehan Li, Yiming Li, Jianfu Li, Xiaoqian Jiang, and Hua Xu. 2023. Zero-shot Clinical Entity Recognition using ChatGPT.

Ming Siang Huang, Po Ting Lai, Pei Yen Lin, Yu Ting You, Richard Tzong Han Tsai, and Wen Lian Hsu. 2020. Biomedical named entity recognition and linking datasets: Survey and our recent development. *Briefings in Bioinformatics*, 21(6):2219–2238.

Minbyul Jeong and Jaewoo Kang. 2023. Consistency enhancement of model prediction on document-level named entity recognition. *Bioinformatics*, 39(6).

Lanxin Jiang, Yu Gu, Wenjun Yu, and Jun Dai. 2022. Blockchain-based Life Cycle Assessment System for ESG Reporting. *SSRN Electronic Journal*.

J. D. Kim, T. Ohta, Y. Tateisi, and J. Tsujii. 2003. GENIA corpus - A semantically annotated corpus for bio-textmining. *Bioinformatics*, 19(SUPPL. 1).

S King and Scott Nadal. 2012. PPCoin: Peer-to-Peer Crypto-Currency with Proof-of-Stake.

Jon M Kleinberg. 2011. Authoritative sources in a hyperlinked environment. In *The Structure and Dynamics of Networks*, volume 9781400841, pages 514–542.

Varun Kohli, Sombuddha Chakravarty, Vinay Chamola, Kuldip Singh Sangwan, and Sherali Zeadally. 2022. An Analysis of Energy Consumption and Carbon Footprints of Cryptocurrencies and Possible Solutions. *Digital Communications and Networks*, 9(1):79–89.

Zhenzhong Lan, Mingda Chen, Sebastian Goodman, Kevin Gimpel, Piyush Sharma, Radu Soricut, and Google Research. ALBERT: A LITE BERT FOR SELF-SUPERVISED LEARNING OF LANGUAGE REPRESENTATIONS.

Xianzhi Li, Xiaodan Zhu, Zhiqiang Ma, Xiaomo Liu, and Sameena Shah. 2023. Are ChatGPT and GPT-4 General-Purpose Solvers for Financial Text Analytics? An Examination on Several Typical Tasks.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. {RoBERTa}: {A} Robustly Optimized {BERT} Pretraining Approach. *CoRR*, abs/1907.1.

Manuel Lobo, Andre Lamurias, and Francisco M. Couto. 2017. Identifying Human Phenotype Terms by Combining Machine Learning and Validation Rules. *BioMed Research International*, 2017.

Brian M Lucey, Samuel Vigne, Yizhi Wang, and Larisa Yarovaya. 2021. An Index of Cryptocurrency Environmental Attention (ICEA). *Monetary Economics: International Financial Flows*.

Faraz Masood and Arman Rasool Faridi. 2018. Consensus Algorithms in distributed ledger technology for open environment.

Robin Mcentire, Debbie Szalkowski, James Butler, Michelle S. Kuo, Meiping Chang, Man Chang, Darren Freeman, Sarah McQuay, Jagruti Patel, Michael McGlashen, Wendy D. Cornell, and Jinghai James Xu. 2016. Application of an automated natural language processing (NLP) workflow to enable federated search of external biomedical content in drug discovery and development. *Drug Discovery Today*, 21(5):826–835.

Esther Mengelkamp, Benedikt Notheisen, Carolin Beer, David Dauer, and Christof Weinhardt. 2018. A blockchain-based smart grid: towards sustainable local energy markets. In *Computer Science - Research and Development*, volume 33, pages 207–214. Springer Verlag.

Niklas Muennighoff, Alexander M. Rush, Boaz Barak, Teven Le Scao, Aleksandra Piktus, Nouamane Tazi, Sampo Pyysalo, Thomas Wolf, and Colin Raffel. 2023. Scaling Data-Constrained Language Models.

Zubair Nabi, Ramzan Talib, Muhammad Kashif Hanif, and Muhammad Awais. 2022. Contextual Text Mining Framework for Unstructured Textual Judicial Corpora through Ontologies. *Computer systems science and engineering*, 43(3):1357–1374.

Rajit Nair, Sweta Gupta, Mukesh Soni, Piyush Kumar Shukla, and Gaurav Dhiman. 2020. WITHDRAWN: An approach to minimize the energy consumption during blockchain transaction. *Materials Today: Proceedings*.

Satoshi Nakamoto. 2008. Bitcoin: A peer-to-Peer Electronic Cash System.

Matthew E Peters, Mark Neumann, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. Deep contextualized word representations. pages 2227–2237.

Moritz Platt, Johannes Sedlmeir, Daniel Platt, Jiahua Xu, Paolo Tasca, Nikhil Vadgama, and Juan Ignacio Ibanez. 2021. The Energy Footprint of Blockchain Consensus Mechanisms Beyond Proof-of-Work. In *Proceedings - 2021 21st International Conference on Software Quality, Reliability and Security Companion, QRS-C 2021*, pages 1135–1144. Institute of Electrical and Electronics Engineers Inc.

Anna Poberezhna. 2018. Addressing Water Sustainability With Blockchain Technology and Green Finance. In *Transforming Climate Finance and Green Investment with Blockchains*, pages 189–196. Elsevier.

Loreen M Powell, Michalina Hendon, Andrew Mangle, and Hayden Wimmer. 2021. Awareness of Blockchain Usage, Structure, & Generation of Platform's Energy Consumption: Working Towards a Greener Blockchain. *Issues in Information Systems*, 22(1):114–123.

Alec Radford, Jeff Wu, Rewon Child, D Luan, Dario Amodei, and Ilya Sutskever. 2019. Language Models are Unsupervised Multitask Learners.

Pranav Rajpurkar, Jian Zhang, Konstantin Lopyrev, and Percy Liang. 2016. SQuad: 100,000+ questions for machine comprehension of text. In *Conference on Empirical Methods in Natural Language Processing*, pages 2383–2392, Stroudsburg, PA, USA. Association for Computational Linguistics.

Ashish Rajendra Sai, Jim Buckley, Brian Fitzgerald, and Andrew Le Gear. 2021. Taxonomy of centralization in public blockchain systems: A systematic literature review. *Information Processing and Management*, 58(4).

Victor Sanh, Lysandre Debut, Julien Chaumond, and Thomas Wolf. 2019. DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. *ArXiv*.

Niranjan Sapkota and Klaus Grobys. 2020. Blockchain consensus protocols, energy consumption and cryptocurrency prices. *Journal of Energy Markets*, 13(4):117–139.

Christophe Schinckus, Canh Phuc Nguyen, and Felicia Chong Hui Ling. 2020. CRYPTO-CURRENCIES TRADING AND ENERGY CONSUMPTION. *International Journal of Energy Economics and Policy*, 10(3):355–364.

Karsten Schulz and Marian Feist. 2021. Leveraging blockchain technology for innovative climate finance under the Green Climate Fund. *Earth System Governance*, 7.

David Schwartz, Noah Youngs, and Arthur Britto. The Ripple Protocol Consensus Algorithm.

Johannes Sedlmeir, Hans Ulrich Buhl, Gilbert Fridgen, and Robert Keller. 2020. The Energy Consumption of Blockchain Technology: Beyond Myth. *Business and Information Systems Engineering*, 62(6):599–608.

Johannes Sedlmeir, Hans Ulrich Buhl, Gilbert Fridgen, and Robert Keller. 2021. Recent Developments in Blockchain Technology and their Impact on Energy Consumption. *ArXiv*, 43(6):391–404.

Luca Di Simone, Barbara Petracci, and Mariacristina Piva. 2022. Economic Sustainability, Innovation, and the ESG Factors: An Empirical Investigation. *Sustainability*, 14(4).

Irena Spasic, Sophia Ananiadou, John McNaught, and Anand Kumar. 2005. Text mining and ontologies in biomedicine: Making sense of raw text. *Briefings in Bioinformatics*, 6(3):239–251.

Florian Spychiger, Paolo Tasca, and Claudio J Tessone. 2021. Unveiling the Importance and Evolution of Design Components Through the "Tree of Blockchain". *Frontiers in Blockchain*, 3:613476.

Ben Steer, Naomi Arnold, † Cheick, Tidiane Ba, Renaud Lambiotte, Haaroon Yousaf, Lucas Jeub, Fabian Murariu, Shivam Kapoor, Pedro Rico, Rachel Chan, Louis Chan, James Alford, Richard G Clegg, Felix Cuadrado, Matthew Russell Barnes, Peijie Zhong, John N Pougué Biyong, and Alhamza Alnaimi. 2023. Raphtory: The temporal graph engine for Rust and Python.

Benjamin Steer, Felix Cuadrado, and Richard Clegg. 2020. Raphtory: Streaming analysis of distributed temporal graphs. *Future Generation Computer Systems*, 102:453–464.

Pontus Stenetorp, Sampo Pyysalo, Goran Topíc, Tomoko Ohta, Sophia Ananiadou, and Jun'ichi Tsujii. 2012. BRAT: AWeb-based tool for NLP-Assisted text annotation. In *EACL 2012 - Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics*, pages 102–107.

Ilya Sutskever, Oriol Vinyals, and Quoc V Le. 2014. Sequence to Sequence Learning with Neural Networks. In Z Ghahramani, M Welling, C Cortes, N D Lawrence, and K Q Weinberger, editors, *Advances in Neural Information Processing Systems 27*, pages 3104–3112. Curran Associates, Inc.

Paolo Tasca and Claudio J Tessone. 2017. Taxonomy of Blockchain Technologies. Principles of Identification and Classification.

Paolo Tasca and Claudio J Tessone. 2019. A Taxonomy of Blockchain Technologies: Principles of Identification and Classification. *Ledger*, 4:1–39.

Erik F Tjong, Kim Sang, and Fien De Meulder. 2003. Introduction to the CoNLL-2003 Shared Task: Language-Independent Named Entity Recognition.

Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, Aurelien Rodriguez, Armand Joulin, Edouard Grave, and Guillaume Lample. 2023. LLaMA: Open and Efficient Foundation Language Models.

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 I Guyon, U V Luxburg, S Bengio, H Wallach, R Fergus, S Vishwanathan, and R Garnett, editors, *Advances in Neural Information Processing Systems 30*, pages 5998–6008. Curran Associates, Inc.

Xiaochen Wang and Yue Wang. 2022. Sentence-Level Resampling for Named Entity Recognition. *NAACL 2022 - 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Proceedings of the Conference*, pages 2151–2165.

Wei Wu, Yelin Fu, Zicheng Wang, Xinlai Liu, Yuxiang Niu, Bing Li, and George Q Huang. 2022. Consortium blockchain-enabled smart ESG reporting platform with token-based incentives for corporate crowdsensing. *Computers and Industrial Engineering*, 172.

Jesse Yli-Huumo, Deokyoon Ko, Sujin Choi, Sooyong Park, and Kari Smolander. 2016. Where Is Current Research on Blockchain Technology?—A Systematic Review. *PLoS ONE*, 11(10).

Qingkai Zeng, Mengxia Yu, Wenhao Yu, Tianwen Jiang, and Meng Jiang. 2021. Validating Label Consistency in NER Data Annotation. pages 11–15.

Zibin Zheng, Shaoan Xie, Hong Ning Dai, Xiangping Chen, and Huaimin Wang. 2018. Blockchain challenges and opportunities: A survey. *International Journal of Web and Grid Services*, 14(4):352–375.