Sustainability in Blockchain Development: A BERT-Based Analysis of Ethereum Developers Discussions

M. Vaccargiu[1,3], S. Aufiero[2], S. Bartolucci[2], R. Neykova[3], R. Tonelli[1], G. Destefanis[3]
[1]University of Cagliari, Italy - [2]University College London, UK - [3]Brunel University London, UK
{matteo.vaccargiu,roberto.tonelli}@unica.it
{sabrina.aufiero.22,s.bartolucci}@ucl.ac.uk
{rumyana.neykova,giuseppe.destefanis}@brunel.ac.uk

ABSTRACT

Blockchain technology faces significant challenges related to sustainability, including issues with optimisation, as well as high energy and gas consumption—factors that developers may sometimes neglect. We introduce a methodology to analyse the key sustainability topics discussed by Go-Ethereum developers, using thematic analysis of their issues and comments from Github. Our approach uses the BERT model to conduct an in-depth topic analysis, enabling us to study the underlying themes and trends in developer's conversations regarding energy use and sustainability. We assess the sustainability of the identified topics using the five dimensions outlined in the Sustainability Awareness Framework (SusAF): economic, social, individual, environmental, and technical. Our goal is to shed light on how much attention developers pay to sustainability and energy consumption issues. The findings from this qualitative analysis aim to encourage technologists to incorporate these considerations into their future projects, in order to achieve better outcomes in terms of sustainability and reduced consumption.

KEYWORDS

Blockchain, Development, BERT, Sustainability, Energy

ACM Reference Format:
M. Vaccargiu[1,3], S. Aufiero[2], S. Bartolucci[2], R. Neykova[3], R. Tonelli[1], G. Destefanis[3]. 2024. Sustainability in Blockchain Development: A BERT-Based Analysis of Ethereum Developers Discussions. In *28th International Conference on Evaluation and Assessment in Software Engineering (EASE 2024), June 18–21, 2024, Salerno, Italy.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3661167.3661194

1 INTRODUCTION

Since its inception, blockchain technology has witnessed growth and widespread adoption, attributed to its key features such as security, scalability, transparency, immutability, and traceability [26]. These features have driven the adoption of blockchain technology in a wide range of industries. However, the surge in popularity of cryptocurrencies, especially Bitcoin, along with other blockchain implementations, has raised significant concerns regarding their environmental footprint. This is primarily due to the increased greenhouse gas emissions and substantial energy consumption associated with their operation [16]. Such environmental concerns have sparked ongoing debates within both scientific communities and industry practices about the overall sustainability of blockchain and distributed ledger technologies [7]. A particular point of contention is the extensive energy requirements driven by the computational processes they employ, notably the Proof-of-Work (PoW) algorithm. This algorithm, commonly used in cryptocurrency mining, demands considerable computing resources [10], highlighting the urgent need for a reassessment of the energy efficiency of these technologies [13]. Initiatives such as the Paris Agreement[1] and the United Nations' 2030 Agenda for Sustainable Development[2] highlight the need for innovative solutions. In this context, research into blockchain's potential to contribute to these global sustainability efforts by enhancing its own sustainability practices is gaining momentum[27].

The interest in understanding how developers prioritise sustainability concerns, address them in their projects, and implement relevant measures in software and application development has led to this study. Topic modelling offers a powerful means to sift through textual data to discover the main themes of discussion. This research explores a dataset comprising issues and comments from Go-Ethereum[3] developers. The Go-Ethereum command-line interface enables developers to operate full Ethereum nodes, mine cryptocurrency, and deploy smart contracts. We focused our analysis on this platform because Ethereum transitioned from a PoW to a Proof-of-Stake consensus mechanism, known as the Merge event, which allows us to track changes in developers' interest and discussions related to sustainability. Our aim is to introduce a reproducible approach for analysing the content of discussions among developers, assessing the sustainability of the topics they engage with, and tracking any developmental trends over time. This study is guided by the following research questions:

- RQ1: Are sustainability issues a topic of discussion among Go-Ethereum developers?
- RQ2: How do sustainability-related discussions evolve over time?

To explore these research questions, we applied Bidirectional Encoder Representations from Transformers (BERT), which helped us analysing key discussion topics through their keywords. Next, we interpreted these keywords using both Chat-GPT 3.5 and manual review. We then identified discussions around sustainability for

EASE 2024, June 18–21, 2024, Salerno, Italy

ACM ISBN 979-8-4007-1701-7/24/06
https://doi.org/10.1145/3661167.3661194

[1]https://unfccc.int/process-and-meetings/the-paris-agreement
[2]https://sdgs.un.org/goals
[3]https://geth.ethereum.org/

each issue highlighted by the model, based on the sustainability framework outlined by Duboc et al. [8].

The paper is organised as follows: in Section 2 we present the current literature, the methodology followed to answer the research questions and inference procedures is in 3, the results are presented in Section 4; and finally, in Section 6, conclusions and future research developments in this area are deduced. The code used for this study is available at this *Code Link*.

2 RELATED WORKS

General analysis related to energy consumption and sustainability in Ethereum is provided in [1, 2, 23] where the authors discuss platform consumption, possible solutions and challenges to improve it, and the transition from proof-of-work to proof-of-stake protocol. An investigation about agile blockchain-oriented software development principles and sustainability software design principles was conducted by Pinna et al.[22], where they present a new Agile method for the development of blockchain-oriented systems that includes sustainability awareness practices within the development phases, in particular in the requirements and the acceptance tests. Eligüzel[9] presents a study on the relationship between blockchain technology and sustainability through a descriptive literature review, using the topic modelling and clustering method of latent semantic analysis, which is a social spider optimisation technique, on the corpus of 1069 articles extracted from Scopus. Another work was carried out by Liu et al.[17], where 759 articles extracted from Web of Science related to blockchain technology in sustainable financial field were analysed by keyword analysis, bi-clustering algorithms, and strategic coordinate analysis so as to explore the hot topics in this field and predict the trend of future sustainable development. Other works of blockchain topic analysis are more focused on using BERT-based models to analyse generic blockchain-related topics with data extracted from social media such as Reddit[12] or based on abstracts from USPTO patents [15], highlighting in both cases the benefits of using an NLP-based BERT textual analysis approach to examine technological knowledge and relationships within the field of blockchain technology.

3 METHODOLOGY

This study develops a reliable and repeatable NLP-based approach for analysing discussion topics, with a focus on identifying and providing insights into the themes most frequently addressed by Ethereum developers in GitHub issues and comments, especially those related to sustainability. In this section, we offer a detailed look at the dataset used and explain the method we employed to extract and analyse the topics.

The dataset employed in this research was obtained from GitHub and centers on issues and comments related to Go-Ethereum spanning from 2014 up to May 2023. Our analysis uses a total of 15,954 issues and 50,023 comments. For every issue, the dataset captures several pieces of information, including an ID, the name of the author, the count of comments, the date of the first posting, the date of the latest update, and the full text of the issue. In a similar manner, each comment is detailed by the author's ID, the ID of the associated issue, the date it was created, and the text of the comment itself. To better prepare the dataset for our analysis, we conducted data pre-processing to ensure data quality. We removed NaN values and stop words to reduce noise and improve model performance. A manual sampling of 600 comments revealed that code snippets were primarily found in issue bodies rather than comments. To minimise the impact of non-natural language content on topic modelling, we merged the comments with their corresponding issues into a singular column, capturing the complete context of the discussions while reducing the influence of code snippets on the model's performance. Figure 1 shows the year-by-year breakdown of issues

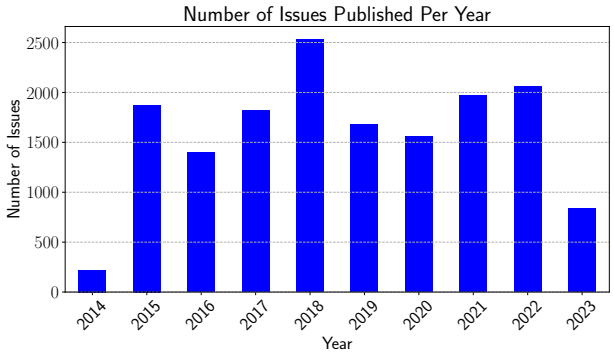

Figure 1: Number of issues published per year from 2014 to 2023.

posted on GitHub. Omitting 2014, the study's start year, and 2023, for which data is only available until May, the yearly distribution of issues is fairly even. The number of issues annually varied from 1,398 in 2016 to 2,061 in 2022, with the highest activity observed in 2018, featuring 2,534 issues. Figure 2 shows the distribution of the lifespan, in years, of closed issues. The median time to resolution is 6 days, showing that half of the issues were addressed within a week. Additionally, 75% of the issues were closed in 83 days or fewer. Out of the total 15,954 issues, only 241 are still *open*. These open issues generally last much longer than the ones that have been closed, with the median open time showing that half of them stay unresolved for at least 125 days. This indicates that the issues that remain open tend to be either more complex or contentious.

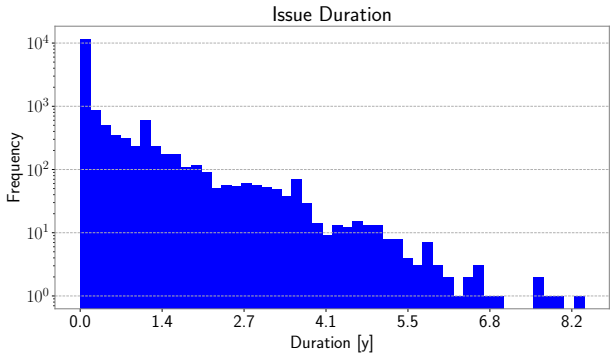

Figure 2: Lifespan of issues marked as *closed* in the dataset.

To each issue, we can assign a measure of popularity based on the number of comments associated with it: this distribution is

plotted in Figure 3. Seventy-five percent of the issues receive four

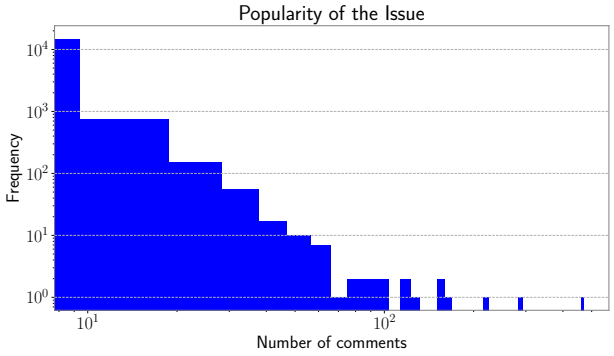

Figure 3: Number of comments per issues.

comments or less, highlighting that most discussions are brief. The issue with the highest engagement, titled "Rinkeby Faucet Down," obtained 470 comments before it was closed on March 19, 2022, after 186 days.

This issue focuses on occasions when the service that distributes free test Ether to developers is down, which disrupts testing processes. The extensive discussion on this issue may point to broader concerns about technical sustainability, including challenges related to network congestion, maintenance, and ease of access. The data reveals that although a large portion of issues are settled swiftly and without extensive discussion, a small number of issues require significant attention from the community and more time to address.

We analysed user activity within the dataset to highlight the five most active contributors based on the number of issues they initiated. The top contributors, with user IDs 129561, 142290, 5959481, and 6264126, opened 1208, 863, 669, 455, and 426 issues, respectively. These individuals are also among the top commentators, highlighting their substantial engagement with the platform. We calculated the "user lifespan", defined as the period from their initial contribution (either posting an issue or making a comment) to their last recorded activity. On average, users remain active for 200 days, yet the median lifespan is just 1 day, suggesting that most users post a single comment and then cease further interaction with the platform. This metric shows the fleeting nature of participation for many users within this ecosystem [20, 21].

Figure 4 shows the step followed for the dataset generation, topic modelling and topic interpretation. The latter two are explained in detail in the next two subsections.

3.1 Topic modelling

Topic modelling uses natural language processing (NLP) techniques and probabilistic algorithms to delineate topics from a corpus of text. In this research, we used the BERT (Bidirectional Encoder Representations from Transformers) model [11].

Our choice was guided by BERT's advanced capability to handle the distinctions of language context, its adeptness at managing short texts, and its minimal dependency on hyperparameter tuning. By using class-based TF-IDF (c-TF-IDF) [3], BERT efficiently processes and associates words with their relevant topics, significantly improving the results' clarity and ease of interpretation. This methodology produces topics identified by keywords, with each keyword linked to a likelihood score, which facilitates interpreting the data. We employed *BERTopic*, which uses BERT's contextual embeddings to analyse topics within texts. This model's architecture enables it to convert text into uniform hidden representations. Our initial trials, applying the full dataset to identify discussion themes, produced ambiguous results with more than 160 varied topics. This led us to adopt a semi-supervised zero-shot method for more precise topic identification. A semi-supervised zero-shot approach is a hybrid technique that combines the advantages of both supervised and unsupervised learning but does not require explicit examples of every category for training. Instead, it uses known keywords for some data points to infer the classification of unlabelled data, even for categories not seen during training. Employing this technique facilitated a more directed analysis, allowing us to guide the model through the incorporation of issue titles and the specific topics pertinent to our study, thus enhancing the model's ability to classify discussions into relevant and previously undefined topics.

We applied *KeyBERT*[4], as discussed by Khan et al. [14], a tool developed specifically for extracting keywords using BERT.

This process involves feeding the text from issues and comments into KeyBERT, which then identifies and returns a set of keywords for each document. These keywords are selected based on the phrases within a document that most closely match the document's overall content. The resulting collection of keywords forms a vocabulary that is used as input for CountVectorizer, a component of Sklearn, aiding the BERTopic model in its identification of topics. As described by Tingting Ma et al.[18], employing a Zero-Shot classification model with unlabelled data enhances the accuracy of topic extraction. The model was created using the following candidate topics of interest: "consumption", "consensus", "cost", "efficiency", "energy", "fee", "gas", "green", "management", "merge", "mining", "optimization", "pos", "pow", "proof-of-stake", "proof-of-work", "protocols", "scalability", "sustainability". These were chosen based on the literature reviews provided in Section 1 and Ethereum's transition from proof-of-work to proof-of-stake, called The Merge [5], which significantly reduced energy consumption and gas emissions [19]. For each experiment, the embedding model used was *BAAI/bge-small-en* [6] and for the representative model we used *facebook/bart-large-mnli* [7]. To evaluate the effectiveness of our topic model, we used a method that combines topic interpretation with calculating the coherence score c_v as outlined by Roder et al. [25]. The coherence score yielded a value of c_v equal to 0.66, which is generally considered to indicate a satisfactory level of coherence. However, the model generated a total of 165 topics, a figure considered excessively high. The emergence of numerous small clusters led to a subsequent experiment with the $min\_topic\_size = 30$ parameter set, aiming to establish a minimum size for each topic cluster. In this revised model, the c_v coherence value of 0.67 was maintained, but the number of topics decreased to 58, making the results easier to interpret. A manual review of

[4]https://maartengr.github.io/KeyBERT/api/keybert.html
[5]https://ethereum.org/en/roadmap/merge/
[6]https://huggingface.co/BAAI/bge-small-en
[7]https://huggingface.co/facebook/bart-large-mnli

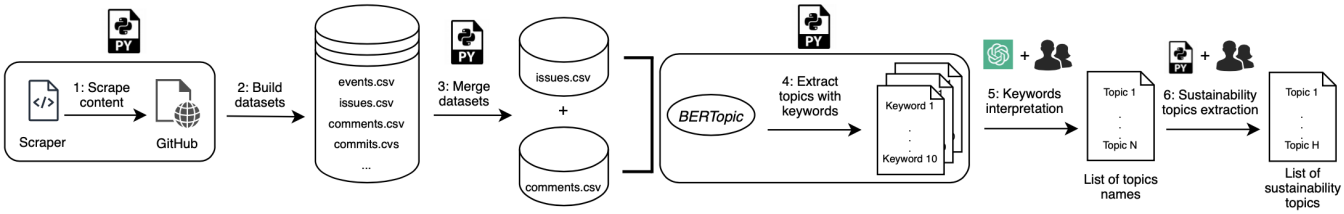

Figure 4: Research Methodology Applied To Topic Modelling Analysis

the topics validated the model's findings, showing them to be consistent.

3.2 Topic interpretation

In the process of topic interpretation, BERT generates a list of 10 keywords for each topic, along with a probability score for each keyword.

This score indicates the relevance of the specific word to the given topic. The interpretation was carried out using a dual approach: initially, Chat-GPT 3.5 analysed the keywords and their respective probabilities to identify the subject of discussion. This automated interpretation was then validated through a manual review by the authors [24].

The use of Chat-GPT 3.5 for topic interpretation was motivated by its ability to understand the context and meaning of the keywords, similar to how a human would interpret them. We tailored a prompt that instructed the tool to generate topic labels based on the provided keywords and their probabilities, ensuring a focused and relevant interpretation process. While keywords alone may not always capture the full scope of a topic, they serve as a strong foundation for both human and machine interpretation.

Colavito et al.[6] demonstrated the effectiveness of GPT-like models for automated labeling tasks of issues without the need for fine-tuning. The study showed substantial agreement between GPT-like models and human annotators, suggesting that these models can be used to reduce the costs associated with manual annotation.

To ensure the accuracy of the Chat-GPT 3.5 interpretations, we also conducted a manual review process. This involved examining the assigned labels and comparing them with the keywords and their probabilities. The manual validation step allowed us to confirm that the labels accurately reflected the underlying topics.

After analysing the full list of topics, we focused on isolating those pertinent to sustainability. This selection was based on the labels assigned to each topic by Chat-GPT and the top 10 words for each topic generated by BERTopic. A topic was flagged as relevant to sustainability if its labels or keywords matched any of the predefined subjects of interest in the zero-shot model. Following the identification of all topics associated with sustainability, we conducted a thorough manual review to validate the findings.

4 RESULTS

In this section, we will discuss the findings derived from the topic extraction and sustainability evaluation. These were conducted according to the guidelines detailed in Subsection 3.2 and within the framework established by Duboc et al. [8].

Topic labels	Size	SusAF measures
Gas Price and Transaction Fees	355	Economic Environmental
Database State Management and Trie	321	Individual Technical
P2P Network and Ethereum Build Process	319	Social Technical
Account Security and Keystore Management	274	Individual Technical
Swarm Network and Manifest Management	234	Social Technical
Tracing and Debugging with Tracers	188	Technical
Ethereum Source and Path Management	118	Technical
Website Deployment and Management with Netlify	88	Technical
Docker Image Management and Process Handling	84	Technical
EVM (Ethereum Virtual Machine) and Stack Management	83	Technical
Source and Path Management for Ethereum	80	Technical
Light Ethereum Subprotocol (LES) Client Logic and Capacity	73	Technical
Peer-to-Peer Protocol and Node Discovery	70	Technical
Bootnode and Peer Management	69	Social Technical
Peer Sync and Latency Management	66	Social Technical
Repository Management and Issue Handling	52	Technical
Whisper Protocol and Message Handling	48	Technical
Framework and Library Management	46	Technical
Benchmark Analysis and Optimisation	46	Environmental Technical
Blockchain Status and Transaction Logs	37	Technical
Light Mode and Peer Management	36	Social Technical
Crypto Benchmarking and Performance Analysis	33	Economic Technical
Community Support and Question Tracking	32	Social Technical

Table 1: Sustainability topics interpretation results based on the topic keywords and SusAF

The Sustainability Awareness Framework (SusAF) delineates five dimensions of sustainability: economic, social, individual, environmental, and technical. The subsequent subsections will explore the topics identified within these five domains, detailing how the discussions among Go-Ethereum developers correspond to and support each sustainability dimension.

In Table 1, we outline the topics identified as relevant to sustainability by the model, with these findings later confirmed through manual verification. The column labelled "SusAF measure" indicates the specific dimensions of sustainability each topic pertains to, which we will explore in greater depth in the next section. After setting aside issues classified as "Undefined", we found that 2830 out of the 8665 total issues are connected to sustainability, accounting for 32.66%.

4.1 Sustainability evaluation framework

Here we provide an analysis of the topics derived from our experiments. Subsequently, we examine the findings in light of the five sustainability dimensions defined by the Sustainability Awareness Framework (SusAF) [8], itself inspired by the Karlskrona Manifesto for Sustainability Design [5].

4.2 General consideration

Sustainability is a key factor in blockchain design and development. The analysis of developer discussions, detailed in Table 1, indicates that sustainability is either a primary or secondary focus in 22 out of 58 topics.

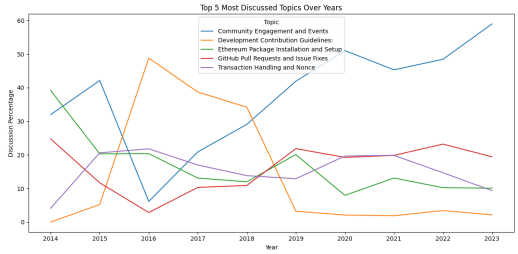

Figure 5: Percentage of issues per topic over the annual amount

Figure 5 presents the top five topics as a percentage of all annual issues, highlighting the main subjects of conversation among developers, with a strong emphasis on the technical management of the platform.

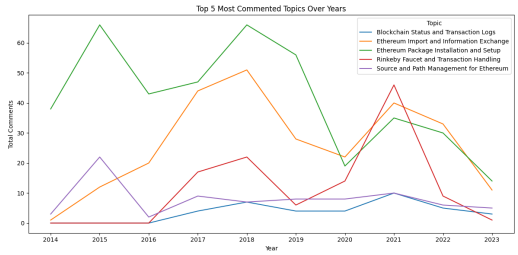

Figure 6: Percentage of comments per topic over the annual amount

Figure 6 displays the top five topics based on their share of the total annual comments, reflecting the community's engagement with certain issues. Similar to the previous figure, the most commented-on topics by developers are technical in nature. For example, "Blockchain Status and Transaction Logs" has been considered as relevant to sustainability at the technical level by our framework. Conversations on these subjects contribute to overseeing blockchain functionality and the efficiency of transactions.

The graph in Figure 7 presents the ratio of discussions on sustainability related topics to the overall number of topics tackled within the same year. The focus on sustainability among Ethereum developers reached its highest in 2021, with 2023 being a close second. This spike in discussion coincides with the Ethereum Merge on September 15, 2022, when Ethereum shifted from a proof-of-work to a proof-of-stake consensus mechanism, slashing its energy consumption by approximately 99.95%. This uptick in sustainability

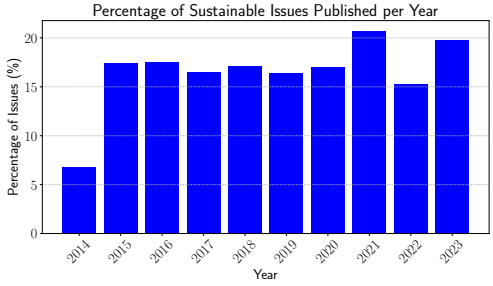

Figure 7: Barplot of percentage of sustainable issues published over the total published each year

conversations likely occurred as the community prepared for the Merge by discussing its implementation and continued post-Merge to assess its impact on efficiency and ongoing operation. The dip in sustainability-related discussions in 2022, the year of the Merge, could be attributed to the predominance of technical issues related to the transition, which may have steered the community's focus away from broader sustainability topics. Finally Table 1 shows how the extracted topics affect the five sustainability measures defined by the Framework of Duboc et al. [8]. As was to be expected, the major contribution that discussions among developers provide concerns the technical sustainability of the platform. Nevertheless, often indirectly, these topics also provide social, individual and economic improvements among developers as well as environmental improvements by reducing energy consumption and optimising emissions.

To summarise our results, the first question, **RQ1**, asked: **Are sustainability issues a topic of discussion among Go-Ethereum developers?** Our findings reveal that *out of the 58 topics identified, 22 were related to sustainability, indicating that sustainability is indeed a subject of discussion among Go-Ethereum developers.*

The second question, **RQ2**, asked: **How do sustainability-related discussions evolve over time?** The analysis shows that *the proportion of sustainability-related issues peaked in 2021, closely followed by 2023, suggesting an increased focus on sustainability topics in recent years, particularly around the time of the Ethereum Merge event in September 2022.*

5 THREATS TO VALIDITY

In conducting this study, we acknowledge a range of threats to its validity that span external, internal, and construct validity concerns. **External validity threats** relate to the applicability of our findings beyond the examined context. The focus on GitHub and specifically the Go-Ethereum project discussions might not fully represent the broader dialogue on sustainability within the blockchain developer communities at large. Relying solely on publicly available discussions from GitHub may not capture the entire spectrum of sustainability concerns among all Ethereum developers, as some discussions may take place on other platforms or in private settings.

Internal validity threats stem from the study's methodology. Employing a BERT-based method for topic extraction and analysis offers robustness but might miss subtle nuances in sustainability discussions. The combination of a semi-supervised zero-shot approach and manual validation introduces a potential for bias, influenced

by the researchers' interpretations and perspectives. Such biases could affect the study's ability to accurately represent the range and depth of sustainability topics. Additionally, using the number of comments as a measure of issue popularity may conflate prolonged issues with genuinely popular ones[4]. An alternative metric, such as the number of unique authors commenting on an issue, could provide a more accurate representation of developer engagement and interest.

Construct validity threats: the accuracy of our study's measurement methods may not fully capture the complexity of blockchain sustainability. Relying on the SusAF framework for categorising discussions could overlook emerging sustainability issues unique to blockchain's rapid evolution. Limiting our analysis to discussions up until May 2023 risks the findings' relevance due to the fast-paced developments in blockchain technology and sustainability practices, affecting the study's validity over time. Interpretation bias, due to subjective manual reviews and automated keyword extraction, may affect our study's replicability. Recognising this need for cautious interpretation, future work should expand to various platforms and employ diverse methods for robust validation of findings.

6 CONCLUSION AND FUTURE WORKS

Blockchain technology's widespread use has spotlighted sustainability and consumption concerns. In this study we developed a systematic method to analyse how developers discuss these issues on GitHub, particularly within the Ethereum community. Our findings reveal that while developers mainly focus on technical details, their discussions also touch on sustainability from various angles, including economic, social, individual, environmental, and technical aspects. The conversation around gas prices and cost efficiency emerges as a central theme, intersecting with all sustainability dimensions. For future works, we aim to improve our topic extraction method and focus into a more detailed yearly analysis of sustainability discussions among developers.

7 ACKNOWLEDGMENTS

This work was partially funded by MUR, issue D.M. 351/2022 Borse di Dottorato di Ricerca di Interesse Nazionale in "Blockchain e Distributed Ledger Technology", under the NRRP.

REFERENCES

[1] Ameena Arshad, Faisal Shahzad, Ijaz Ur Rehman, and Bruno S. Sergi. 2023. A systematic literature review of blockchain technology and environmental sustainability: Status quo and future research. *International Review of Economics Finance* 88 (2023), 1602–1622. https://doi.org/10.1016/j.iref.2023.07.044

[2] Rameez Asif and Syed Raheel Hassan. 2023. Shaping the future of Ethereum: Exploring energy consumption in Proof-of-Work and Proof-of-Stake consensus. *Frontiers in Blockchain* 6 (2023), 1151724.

[3] Prafulla Bafna, Dhanya Pramod, and Anagha Vaidya. 2016. Document clustering: TF-IDF approach. In *2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT)*. IEEE, 61–66.

[4] Setu Kumar Basak, Lorenzo Neil, Bradley Reaves, and Laurie Williams. 2023. What Challenges Do Developers Face About Checked-in Secrets in Software Artifacts?. In *2023 IEEE/ACM 45th International Conference on Software Engineering (ICSE)*. IEEE, 1635–1647.

[5] Christoph Becker, Ruzanna Chitchyan, Leticia Duboc, Steve Easterbrook, Birgit Penzenstadler, Norbert Seyff, and Colin C Venters. 2015. Sustainability design and software: The karlskrona manifesto. In *2015 IEEE/ACM 37th IEEE International Conference on Software Engineering*, Vol. 2. IEEE, 467–476.

[6] Giuseppe Colavito, Filippo Lanubile, Nicole Novielli, and Luigi Quaranta. 2024. Leveraging GPT-like LLMs to Automate Issue Labeling. In *2024 IEEE/ACM 21st International Conference on Mining Software Repositories (MSR)*. IEEE.

[7] Advait Deshpande, Katherine Stewart, Louise Lepetit, and Salil Gunashekar. 2017. Distributed Ledger Technologies/Blockchain: Challenges, opportunities and the prospects for standards. *Overview report The British Standards Institution (BSI)* 40 (2017), 40.

[8] Leticia Duboc, Birgit Penzenstadler, Jari Porras, Sedef Akinli Kocak, Stefanie Betz, Ruzanna Chitchyan, Ola Leifler, Norbert Seyff, and Colin C Venters. 2020. Requirements engineering for sustainability: an awareness framework for designing software systems for a better tomorrow. *Requirements Engineering* 25 (2020), 469–492.

[9] Nazmiye Eligüzel. 2023. An analysis of the integration of sustainability concepts into blockchain technology. *International Journal of Applied Methods in Electronics and Computers* 11, 3 (Sep. 2023), 158–164. https://doi.org/10.58190/ijamec.2023.43

[10] Eshani Ghosh and Baisakhi Das. 2020. A Study on the Issue of Blockchain's Energy Consumption. In *Proceedings of International Ethical Hacking Conference 2019*, Mohuya Chakraborty, Satyajit Chakrabarti, and Valentina E. Balas (Eds.). Springer Singapore, Singapore, 63–75.

[11] Maarten Grootendorst. 2022. BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv preprint arXiv:2203.05794* (2022).

[12] Giacomo Ibba and Matteo Vaccargiu. 2023. Analysis of Users' Most Discussed Topics and Trends on Blockchain Technologies and Smart Contracts. In *2023 IEEE International Conference on Software Analysis, Evolution and Reengineering (SANER)*. IEEE, 865–873.

[13] Md Rafiqul Islam, Muhammad Mahbubur Rashid, Mohammed Ataur Rahman, Muslim Har Sani Bin Mohamad, et al. 2022. A comprehensive analysis of blockchain-based cryptocurrency mining impact on energy consumption. *International Journal of Advanced Computer Science and Applications* 13, 4 (2022).

[14] Uddin MI Roman M Alharbi A Alosaimi W Almalki J Alshahrani SM. Khan MQ, Shahid A. 2022. Impact analysis of keyword extraction using contextual word embedding. *PeerJ Computer Science* (2022). https://doi.org/10.7717/peerj-cs.967

[15] Brian Tae-Seok Kim and Eun-Jung Hyun. 2023. Mapping the Landscape of Blockchain Technology Knowledge: A Patent Co-Citation and Semantic Similarity Approach. *Systems* 11, 3 (2023). https://doi.org/10.3390/systems11030111

[16] Varun Kohli, Sombuddha Chakravarty, Vinay Chamola, Kuldip Singh Sangwan, and Sherali Zeadally. 2023. An analysis of energy consumption and carbon footprints of cryptocurrencies and possible solutions. *Digital Communications and Networks* 9, 1 (2023), 79–89. https://doi.org/10.1016/j.dcan.2022.06.017

[17] Yunmei Liu, Shuai Zhang, Min Chen, Yenchun Wu, and Zhengxian Chen. 2021. The Sustainable Development of Financial Topic Detection and Trend Prediction by Data Mining. *Sustainability* 13, 14 (2021). https://doi.org/10.3390/su13147585

[18] Tingting Ma, Jin-Ge Yao, Chin-Yew Lin, and Tiejun Zhao. 2021. Issues with Entailment-based Zero-shot Text Classification. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, Chengqing Zong, Fei Xia, Wenjie Li, and Roberto Navigli (Eds.). Association for Computational Linguistics, Online, 786–796. https://doi.org/10.18653/v1/2021.acl-short.99

[19] P. Rajitha Nair and D. Ramya Dorai. 2021. Evaluation of Performance and Security of Proof of Work and Proof of Stake using Blockchain. In *2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV)*. 279–283. https://doi.org/10.1109/ICICV50876.2021.9388487

[20] Marco Ortu, Giuseppe Destefanis, Tracy Hall, and David Bowes. 2023. Fault-insertion and fault-fixing behavioural patterns in Apache Software Foundation Projects. *Information and Software Technology* 158 (2023), 107187.

[21] Marco Ortu, Tracy Hall, Michele Marchesi, Roberto Tonelli, David Bowes, and Giuseppe Destefanis. 2018. Mining communication patterns in software development: A github analysis. In *Proceedings of the 14th international conference on predictive models and data analytics in software engineering*. 70–79.

[22] Andrea Pinna, Gavina Baralla, Michele Marchesi, and Roberto Tonelli. 2021. Raising sustainability awareness in agile blockchain-oriented software engineering. In *2021 IEEE International Conference on Software Analysis, Evolution and Reengineering (SANER)*. IEEE, 696–700.

[23] Kaihua Qin and Arthur Gervais. 2018. An overview of blockchain scalability, interoperability and sustainability. *Hochschule Luzern Imperial College London Liquidity Network* (2018), 1–15.

[24] Emil Rijcken, Floortje Scheepers, Kalliopi Zervanou, Marco Spruit, Pablo Mosteiro, and Uzay Kaymak. 2023. Towards Interpreting Topic Models with ChatGPT. In *The 20th World Congress of the International Fuzzy Systems Association*.

[25] Michael Röder, Andreas Both, and Alexander Hinneburg. 2015. Exploring the space of topic coherence measures. In *Proceedings of the eighth ACM international conference on Web search and data mining*. 399–408.

[26] Gautami Tripathi, Mohd Abdul Ahad, and Gabriella Casalino. 2023. A comprehensive review of blockchain technology: Underlying principles and historical background with future challenges. *Decision Analytics Journal* 9 (2023), 100344. https://doi.org/10.1016/j.dajour.2023.100344

[27] Matteo Vaccargiu, Andrea Pinna, Roberto Tonelli, and Luisanna Cocco. 2023. Blockchain in the Energy Sector for SDG Achievement. *Sustainability* 15, 20 (2023). https://doi.org/10.3390/su152014843