Corpus-based analysis of human values in blockchain and constitutions

ADITYA JOSHI, ERIC LEASE MORGAN, AGNIESZKA MARCZAK-CZAJKA, ANDREY KUEHLKAMP, ELLEN JOYCE, SWAPNA K. RADHA, EMILIA JUSTYNA POWELL, and JAREK NABRZYSKI, University of Notre Dame, USA

The examination of human values integrated in technologically advanced systems can provide novel perspectives on the broader societal impact of the adoption of those systems. Many blockchain networks are developed and operated through the collaboration of people from countries with diverse social and political ideologies. We describe a data-driven approach to study the human values in blockchains, and their relationships with human values in constitutions from around the world. We introduce a novel collection of datasets, comprising annotated human values references from representative blockchain documents, and unique constitutions from 194 sovereign states. We conduct an initial exploratory dataset analysis to establish baselines to motivate further research in the area. Our analysis demonstrates that the human values in blockchains exhibit centrism in their alignment with geopolitical ideologies.

CCS Concepts: • **Social and professional topics → Socio-technical systems**; • **Applied computing → Sociology**.

Additional Key Words and Phrases: Human values, blockchain, constitution

ACM Reference Format:
Aditya Joshi, Eric Lease Morgan, Agnieszka Marczak-Czajka, Andrey Kuehlkamp, Ellen Joyce, Swapna K. Radha, Emilia Justyna Powell, and Jarek Nabrzyski. 2025. Corpus-based analysis of human values in blockchain and constitutions. *Distrib. Ledger Technol.* X, X, Article 111 (July 2025), 25 pages. https://doi.org/XXXXXXX.XXXXXXX

1 Introduction

Blockchain technology continues to expand, and is projected to grow from USD 27.84 billion in 2024 to USD 825.93 billion by 2032 [36]. The increased acceptance of blockchain raises curiosity to inquire whether some of its tenets are at odds with the human values enshrined in domestic legal systems, and in particular national constitutions. This invites a deeper examination of why it is important to study the human values pertinent to blockchains. To answer this, we should delve deeper and examine the core principles that motivate blockchain's networked computing design.

The idea of a data structure similar to a blockchain was first described in Satoshi Nakamoto's paper on Bitcoin [52], which subsequently led to the development of Nakamoto consensus for establishing the necessary *trust* mechanism for performing financial transactions without the oversight of third-parties [72]. Today, blockchains are commonly described as decentralized databases that chronologically record every transaction, which are validated and maintained by networks of computers around the world [65]. Blockchains have democratized computing by enabling the

Authors' Contact Information: Aditya Joshi, ajoshi2@nd.edu; Eric Lease Morgan; Agnieszka Marczak-Czajka; Andrey Kuehlkamp; Ellen Joyce; Swapna K. Radha; Emilia Justyna Powell, epowell4@nd.edu; Jarek Nabrzyski, naber@nd.edu, University of Notre Dame, Notre Dame, Indiana, USA.

ACM 2769-6480/2025/7-ART111
https://doi.org/XXXXXXX.XXXXXXX

development of automated systems based on decentralized, peer-to-peer network models [82]. The blockchain ethos has inspired visions of a Decentralized Society (DeSoc) that can resist wealth concentration and vulnerability of governance to financial attacks [54].

Many large blockchain networks, such as Bitcoin and Ethereum, are community driven projects by individual collaborators from around the world. We seek to understand whether blockchain design choices in these projects are influenced by a collaborator's sentiments towards human values, assuming that those sentiments are also reflected in the constitutions of the sovereign states where the collaborator has ever domiciled. As an example of the global spread of blockchain developer communities, according to an analysis shared in GitHub by Electric Capital [26], the top five regions home to blockchain developers in 2022 were: US (29%), Europe (29%), Asia (13%), UK (6 %) and India (6%).

In our assessment, we analyze the extent to which sovereign state constitutions uphold human values. A constitution of any sovereign state provides a backdrop—however aspirational—that contains institutional setup for governance, establishes limitations on the state and expectations for the citizenry. Many constitutions contain references to values, whether social, economic, political, or the values associated with the well-being of an individual. In this paper, we focus on analyzing human values [1] represented in constitutions from 194 sovereign states. Correspondingly for blockchains, we retrieve human values connoted in a variety of data sources, including articles, textbooks, blogs, etc. A note about our terminology – henceforth in this paper, the single-word "value" implies "human values" unless otherwise emphasized for clarity. Also, the words "country" and "nation" imply a "sovereign state" for brevity.

Previous studies on values in the context of blockchain have examined ethical concerns [70], economic implications [31, 56], or cultural influences [64]. Our study of human values in blockchain and constitutions is motivated by the need for more interdisciplinary studies that explore the relationships between human concerns within political contexts and technology innovations.

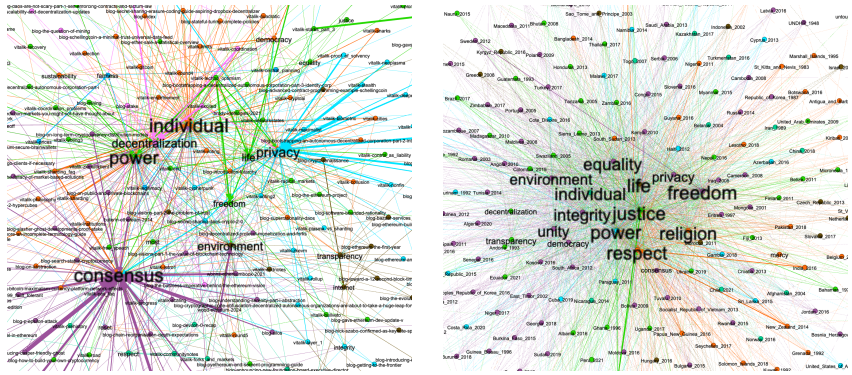
Contributions: (1) Publicly available annotated datasets of human values created using constitutions and blockchain corpora. These datasets can be used for training AI models. (2) Novel application of existing Named Entity Recognition (NER) modeling techniques for an interdisciplinary study that bridges political science (constitutions) and digital technology (blockchain).

Research Goals: (1) To develop a methodology for identifying human values in blockchains and country constitutions, and (2) To compare human values in blockchains and country constitutions, for analyzing the following:

- **RQ1**: Which human values are most commonly referenced in blockchains and country constitutions?
- **RQ2**: Which human values do blockchains share with country constitutions?
- **RQ3**: Are the human values in blockchains uniformly represented worldwide, Eurocentric, Afrocentric, Americentric, or Asiacentric?
- **RQ4**: How democratic are the human values incorporated in blockchains?
- **RQ5**: Do the human values in blockchains align with certain geopolitical ideologies?

The remainder of this article is organized as follows. We begin with summarizing seminal works on human values to establish a definition of human values that forms the basis of our study. We briefly cover key related work before describing the dataset development process. Subsequently, we elaborate on model development and application for human values identification. We analyze the questions that motivate our research in the light of our results. Finally, we conclude with remarks on our modeling limitations and future research directions.

[1] See the discussion in section 2 for our definition of human values



(a) Blockchain (b) Constitutions

Fig. 1. **Human values in blockchain and constitutions corpus**. Prominent human values are shown protruding from source documents in the background. The font size of a word connoting a human value is proportional to its frequency in the corpus.

2 What are Human Values?

Human values drive selective orientations in our interests, pleasures, likes, preferences, duties, moral obligations, desires, wants, goals, needs, aversions and attractions [62]. In this paper, we use "human values" to underscore that these values pertain to human behavior at individual and societal levels. The notion of values functioning as links between self and society to explain social behavior, has been previously elaborated by Rokeach, M. [61]. Note that values differ from norms and attitudes, both of which are narrower in scope, and usually refer to specific actions, objects, or situations [67].

To build our perspective on values, we draw from revered sources. Cambridge dictionary defines the plural noun values as "the beliefs people have, especially about what is right and wrong and what is most important in life, that control their behaviour" [24]. Social science scholars, however, have provided several interpretations of values based on differing rationales and terminology. For instance, previous studies describe values as representations of universal human requirements – namely, the biologically based needs of the organism, the social interactional requirements for interpersonal coordination, and the social institutional demands for group welfare and survival [68]. These diverse perspectives provide a foundation for developing a broad and cohesive understanding of values.

Nevertheless, to compare human values across different domains – blockchain (technology) and constitutions (political science) — we must attempt a difficult task to adopt or develop a description of human values that most concisely unites the perspectives on values appropriate to the two domains. In an earlier work on eliciting a common interpretation, Cheng and Fleischmann summarized the key differences among scholars in describing values [16]. They noted that previous works described values using collective terms, such as, "enduring belief," "principles," etc. We call these terms *concepts and notions*. Drawing on these literature, we pose the following definition of human values as applicable to our research goals:

> Human values are concepts and notions that represent ideals, or standards, influencing people to develop aspirations and orientations, and to follow a course of action towards realizing the desiderata.

3 Related Work

The intersection of technology, ethics, and governance has long been a fertile ground for scholarly inquiry. Several previous studies have explored the ethics of blockchain, and the social and economic values fostered by blockchain adoption [2, 14, 70]. Following a linguistic analysis driven approach, Inwood and Zappavigna developed a corpus of whitepapers of blockchain start-ups, and analyzed the values manifested in their design concepts [43]. This corpus-based approach provided the empirical basis to discuss ideological and political positions expressed in those documents.

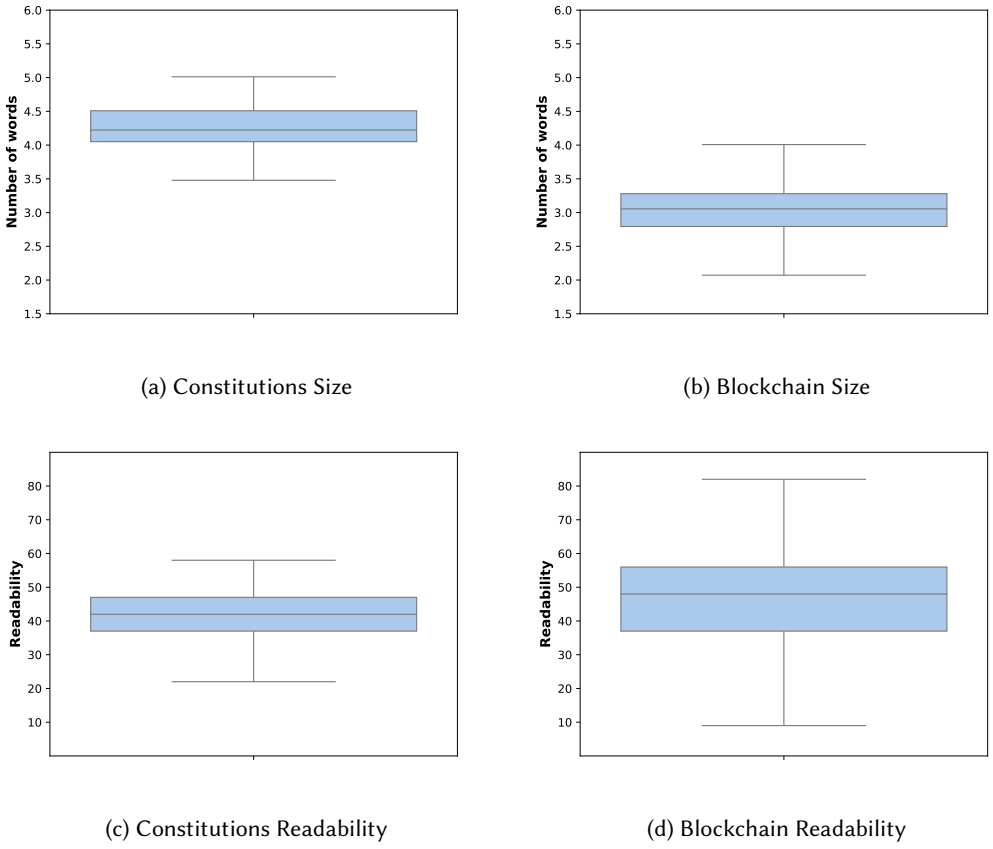
Blockchain technologies have been scrutinized for their ethical and governance implications. Walch, A. [78] discussed the evolution of blockchain lexicons and their implications for legal frameworks, and also emphasized challenges in decentralized governance systems. De Filippi, P. and Wright, A. [21] provided a foundational exploration of how blockchain challenges traditional legal structures, particularly in embedding ethical considerations into code. These works underscore the tension between decentralization and accountability, which is crucial to understanding how human values manifest in blockchain systems.

The concepts of values alignment, human preferences, values assessment, and values driven reward models have been extensively applied to develop data-driven approaches for evaluating software and AI systems. The seminal works by Rokeach, M. [61] and Schwartz, S.H. [67] provided foundational framework to study human values in many scientific disciplines. Building on these foundations, several datasets have been created to assess AI systems in the context of ethics and morality. For example, the ETHICS dataset [41] enables assessing a language model's outputs to express basic moral sentiments and ethical judgments. Similarly, the SCRUPLES [49] dataset contains descriptive judgments and anecdotes to assess machine ethics, and AI system's understanding of ethical and moral norms. Another notable dataset, SOCIAL-CHEM-101 [35] was inspired by descriptive ethics [46], and organizes norms as rules-of-thumb (RoTs) over real-life situations. This dataset facilitates the training of AI models to better understand social norms and acceptability.

Human values in constitutions have also been explored extensively. Shulztiner, D. and Carmi, G.E. [71] delve into the role of dignity and justice as foundational principles in national constitutions, emphasizing how these values guide governance and legal frameworks. Elkins, Ginsburg, and Melton [28] provide a comparative perspective on constitutional endurance, offering insights into how values evolve and persist across different socio-political contexts. Further, Feld, L.P. and Voigt, S. [33] analyze judicial independence and its relationship to economic growth, highlighting how constitutions encode values like fairness, transparency, and accountability.

In socio-technical systems, value-sensitive design (VSD) has emerged as a critical framework for embedding human values into technology. Friedman, B., Kahn, P. H., and Borning, A. [39] introduced VSD to emphasize the importance of integrating ethical considerations into the design of technological systems. Gabriel, I. [40] extended this discussion to artificial intelligence, and proposed methodologies for aligning technological outputs with societal values. These concepts are directly applicable to this study's methodology for identifying and prioritizing values in blockchain documents. Binns, R. [9] highlighted the parallels between fairness in machine learning and political philosophy, offering insights into how systems designers can account for diverse value systems. Jasanoff, S. [44] explored how ethical considerations shape technology's societal impact, which is a perspective that resonates with the goals of this study.

The scholarship on domestic constitutions because of its truly interdisciplinary nature hails from several disciplines including comparative law and political science. Interestingly, most of this literature focuses on political, economic, and social dimensions of constitutional provisions. Granted, much work has been done on the presence of human values in constitutions in order to assess whether, and if so, how countries incorporate international human rights into their domestic



(a) Constitutions Size

(b) Blockchain Size

(c) Constitutions Readability

(d) Blockchain Readability

Fig. 2. **Dataset sizes and readability scores**. The constitutions documents contain orders of magnitude more words than the blockchain documents in our datasets. The sizes are displayed in Common Log Scale for easier comparison. The blockchain dataset has better readability score than the constitutions dataset.

legal systems. Furthermore, a substantial body of the scholarship explores effects of these and other constitutional provisions on within-state societal and legal dynamics, and countries' subsequent behavior in the domestic as well as international spheres [11, 17, 30, 73]. Much of the literature in recent years has also considered whether constitutions provide mere window dressing and how to design constitutions that generate greater legal protection. Yet, the topic of human values – beyond human rights – remains somewhat unexplored.

4 Dataset

In this paper, we introduce datasets [2] for studying the human values across blockchain and country constitutions. The datasets include annotated human values identified in blockchain documents,

[2]The datasets are available at:

- constitutions - https://doi.org/10.5281/zenodo.14230231
- blockchain - https://doi.org/10.5281/zenodo.14236844
- human values - https://doi.org/10.5281/zenodo.14237012

and constitutions from 194 countries. This section elaborates the dataset creation process, and descriptive statistics.

The process used to create these datasets was very similar to each other: (1) Download documents of interest and establish distinct collections. (2) Associate each item in each collection with author- and title-like values. (3) Use the Distant Reader Toolbox [51] to transform the collections into data sets. The Distant Reader Toolbox (DRT) is used for reading and analyzing large text corpora at scale. The DRT can be used to extract, transform and summarize text content from URLs, or a set of files. By definition, data sets are computable, and the DRT can create computational models using these data sets in a number of ways, including but not limited to: rudimentary counts and tabulations, full-text indexing, concordances, semantic indexing, topic modeling, network graphs, and to some degree large-language models.

4.1 Dataset Characteristics

Each data set includes a number of things:

- The original documents.
- Plain text derivatives of the original documents.
- Sets of extracted features (bibliographics, parts-of-speech, named-entities, computed keywords, etc.) in the form of tab-delimited files.
- An SQLite database of the extracted features.

Appendix A.2 contains statistics on the frequencies of unigrams, bigrams, and computed keywords from the corpus of constitutions, blockchains and combined datasets. See Appendix A.2 for details on the frequency characteristics of the dataset.

4.2 Blockchain Dataset

The blockchain dataset was created using 3,086 unique documents comprising 4,979,903 total words. The documents include technical articles, blogs and project specifications of blockchain core platforms (Ethereum and Bitcoin), decentralized exchange (Uniswap [1]), governance (Aave [38]) and non-financial applications (Hyperledger [23], Filecoin [34] and Arweave [80]). The DRT's readability score for the blockchain dataset was 53 (0 denotes impossible and 100 denotes easy to read). Figure 2 (b) and Figure 2 (d) summarize these metrics.

4.3 Constitutions Dataset

The constitution dataset was created from 194 unique sovereign state constitutions, comprising of 4,702,380 total words. More often than not, the constitution in its written form is the heart of the constitutional regime of a state. Thus, in general written constitutions constitute a mainstay of an overwhelming majority of domestic legal systems, and usually it is not difficult to identify them. Our data comprising constitutional texts of all written domestic constitutions come from the Constitute project [27, 29] launched by the Comparative Constitutions Project (CCP) and Google Ideas in 2013 [74].

The DRT's readability score for the constitutions dataset was 40 (0 denotes impossible and 100 denotes easy to read). Figure 2 (a) and Figure 2 (c) summarize these metrics and compare them with the blockchain dataset.

4.4 Human Values Dataset

We further process the constitutions and blockchain datasets to identify human values in each of these datasets. Our process was as follows: (1) Create two sets – blockchain and constitutions – of human annotated values (gold annotations set) in samples of documents, (2) Rank the annotated

Table 1. **Blockchain documents used for creating gold annotations**

Title	Description	Why chosen?
Ethereum whitepaper [13]	Introductory paper on the Ethereum blockchain by its founder.	The paper incorporates the values that inspired original design, philosophy and vision of Ethereum
Bitcoin: A Peer-to-Peer Electronic Cash System [52]	An introduction to Bitcoin as a digital currency based alternative to traditional financial systems	The paper discusses the advantages and values enabled by Bitcoin
Ethereum: A secure decentralised generalised transaction ledger [81]	Overview of Ethereum's architecture, consensus mechanism, and potential applications	The paper mentions the values embedded in technical specifications
Ideologies and imaginaries in blockchain communities: The case of Ethereum [12]	Analysis of the complex interplay between technology, ideology, and community in the context of Ethereum	The article delves into the underlying ideologies and visions that drive the Ethereum community
Mastering ethereum: building smart contracts and dapps [5]	Details on the Ethereum architecture, applications and technology impact	The book describes Ethereum's capabilities and highlights its unique values

values from the two sets in priority order, (3) Create a Named Entity Recognition (NER) model to identify additional human values, apply the model to our corpus, and output measurements, and (4) Articulate generalizations garnered from the measurements.

4.4.1 Annotation Process. To annotate human values in blockchain and constitution documents, we formed two focus groups based on the specific skill sets and professional backgrounds of the participating volunteers. The annotators from the two focus groups met to agree upon the value labels and definitions. For creating the gold annotations set with blockchain documents, the human annotators included four graduate students and two faculty members. The group members had five to twenty years of experience in computer science and blockchain industry. The group identified sentences containing human values in five documents related to blockchain. The documents express perspectives of a variety of authors in different formats, such as technical articles, whitepapers, and books. Table 1 contains descriptions of these documents.

For creating the gold annotations set with constitutional texts, the annotation process was conducted by two groups consisting of four undergraduate students each, all with backgrounds in political science. The two groups were responsible for annotating a complete set of 194 world constitutions, ensuring a comprehensive analysis across a diverse range of legal frameworks. Following the initial annotation phase, both groups participated in a focus group discussion to refine and finalize the list of human values identified.

4.4.2 Prioritization. After identifying the human values in blockchain and constitutions, the respective groups were asked to develop ordered lists of the most relevant values, ranked by priority. For constitutions, human values were ranked according to their frequencies of occurrence

Table 2. **Human values ordered by priority**. Values common between constitutions and blockchain are highlighted in yellow.

Domain	Values ordered by priority
Constitutions	power, freedom, order, unity, respect, justice, equality, security, individual, religion, environment, life, community, family, peace, competence, dignity, tradition, confidence, privacy, discipline, trust, conscience, safety, autonomy, courage, morality, honor, belonging, reputation, transparency, leisure, well-being, efficiency, leadership, stability, fulfillment, mercy, love, ethics, prosperity, harmony, truth, obedience, honesty, tolerance, loyalty, creativity, understanding, productivity, self-determination, wisdom, happiness, destiny, decency, prestige, tranquility, hope, prudence, appreciation, authenticity, consciousness, openness, self-reliance, self-sufficiency, excellence, goodwill, gratitude, perfection, ambition, forgiveness, uniqueness, benevolence
Blockchain	security, trust, transparency, integrity, privacy, equality, decentralization, consensus, fairness, democracy, power, sustainability, justice, freedom, diversity, self-governance, compliance, autonomy, scalability, truth, accountability, anonymity, faithfulness, participation, community, responsibility, reliability, efficiency, independence, innovation, objectivity, cooperation, safety, honesty, stability, respect, resilience, inclusivity, standardization, verifiability, collaboration, traceability, self-enforcing, ethos, validity, solidarity, reputation, opportunity, improvement, unity, flexibility, openness, traceacility, auditability, control, sufficiency, incentivization, egalitarian

within all constitutions. For blockchain, each group member was required to independently create different prioritized list of values, based on the member's domain knowledge, and accounting the frequencies of occurrence of those values. We aimed to leverage domain expertise to overcome gaps in using only frequency-based ranking for the two datasets, given the orders of difference of vocabulary sizes between constitutions and blockchain datasets (see Figure 2). Section 5.3 contains a related discussion on bias. The different prioritized lists were consolidated to construct a single list for blockchain values, using the following methodology:

(1) Suppose there are K different ranked sets of human values.
(2) For each value, assign its lowest rank across the K sets as its "priority" score.
(3) Order all the human values using the assigned priority scores.
(4) Move to the top the values that are common across the K sets, starting with the highest priority score.
(5) Assign higher priorities to few randomly selected low frequency (rare) values, and place them near the middle of the combined lists. The resulting two lists are shown in Table 2.

5 Model for Identifying Human Values

In this section, we describe our semi-supervised learning approach to develop a dataset of human values by utilizing an NER and extraction model on constitution and blockchain datasets. Before describing the modeling process, we discuss how the model can reduce the errors of identifying polysemous words that do not represent a human value in the given context.

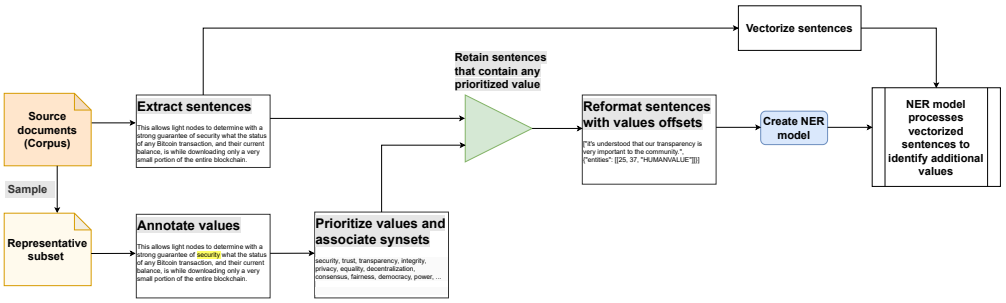


Fig. 3. **The overview of our model for identifying human values.** We process annotated and prioritized values utilizing WordNet synset definitions for sense disambiguiation. The processed outputs train a Named-Entity Recognition (NER) model for values identification.

5.1 Human Values Disambiguation

A word can convey multiple meanings depending on context. In computational linguistics, this is called polysemy, and the multiple meanings of a word are also called *senses* [55], such that each sense depends on the context in which it is used. The term context refers to the neighboring words in a piece of text [20]. To account for word context, several methods to computationally disambiguate different senses of words have been developed [53]. Nonetheless, novel language models for Word Sense Disambiguation (WSD) continue to be developed to fit the requirements of specific questions under study.

Not every human value word needs to be disambiguated from its other senses. For the words that need disambiguation, contextual information can be utilized by linking values with affects [32, 45, 61, 67], that is, the feelings infused by those values. In the sentence, "People who value independence become despair when they are helpless to protect it" [67], the word "independence" functions as a human value because it can be linked to the affect, "feeling despair." Without such contextual links, the word independence may not connote a human value.

Can a word's sense be more accurately determined by changing the context? Many supervised and knowledge-based WSD approaches have utilized contextual information from single sentences [79], a few sentences [3], topic modeling the entire document [15], or by enriching the knowledge base with external resources such as Wikipedia [57]. Conventionally, for general English words a moderate context window size between 4 and 10 was recommended [48]. More recently, WSD architectures built on top of language models like BERT [22] have performed better at capturing meaning, by including every similar context and semantic collocation information [66].

However, complex language models are computationally expensive and increase the carbon footprint [4]. Thus, a viable alternative towards greener AI is to develop simpler WSD models based on semi-supervised approach using a small set of contexts annotated with correct senses. In our study, the WSD task requires disambiguating a pre-compiled list of words conveying human values, given an inventory of senses, which is, WordNet [50]. Next, the task identifies additional human values in unlabeled contexts. Given the specificity of target words (human values, used as nouns), and the context sources (constitutions and blockchain), we trained a simpler model for named-entity extraction, and quantified the accuracy of its output in the human values dataset.

5.2 Modeling Process

5.2.1 Create Dictionaries of Human Values Synsets. We aim to make the NER model capable of identifying human values even for words that have ambiguous senses. To supervise the model training, we prepared examples of what senses connote human values, using standardized definitions of ambiguous words to differentiate between the human values and non values senses. We applied the definition of human values (see Section 2) to make these distinctions. For each word in our prioritized sets, we queried WordNet to extract the definitions (called glosses in WordNet) and unique identifiers (synset IDs) that allude to human values rather than something else. Thus, we obtained two dictionaries of synsets - one for blockchain and another for constitutions - formatted as JSON streams. Appendix A.1 contains few examples of the prioritized values associated with synsets.

5.2.2 Extract Context and Values. We used the Natural Language Toolkit (NLTK) [10] to extract all the sentences from our dataset, and created a subset of the higher prioritized human value words. We then looped through each sentence, and retained the sentences containing at least one of our prioritized human value words. We applied the Lesk algorithm [47] (as implemented in the NLTK) to identify the relevant synonym in the human value word's synset. If that synonym was in our dictionary of synsets, then the sentence was reformatted into a new JSON stream denoting the offset in the sentence where the value was located and the sentence itself. This JSON form is needed by the next step in our process. Appendix A.1 contains a few examples of the extracted sentences. Figure 3 illustrates the different components of our values identification model.

5.2.3 Model Training. We adopted the NER based modeling approach because it is a widely recognized methodology, specially suitable for identifying key elements in text, which is the central requirement of our research. Also, NER modeling allows prioritizing named entities, which fits with our values prioritization process. In particular, we used spaCy [42] for NER modeling because it inherently supports tokenization and POS tagging, along with providing high performance.

Our NER model development using spaCy entailed: (1) dividing our semi-supervised set of JSON into training and testing sets, (2) transforming each JSON record into a spaCy object, and (3) training. Each sentence was tokenized, vectorized, and associated with the given human value, using the training set. The model then used the test set to predict whether an input (vectorized) sentence contained a human value, and output the specific value. The training steps were repeated over several iterations, and the model with the highest accuracy was retained.

The set of prioritized values included 113 unique items from the constitutions and blockchain sets. For the model created using all the prioritized values, the accuracy was lower than 60%. However, when we used the top 10% of values in each set, the accuracy scores were very high (>90%). To find a balance between accuracy and representation, we included 10 of the highest prioritized values from each set (20 total items). This enabled the model to maintain and accuracy of 75%. In other words, using about 18% of 113 prioritized values, the model's accuracy was never lower than 75%. The performance metrics for the final epoch of the NER model training were: F-score (73.0), Precision (66.63) and Recall (80.94). See Appendix A.3 for a discussion of the model's precision with the validation dataset.

5.3 Modeling Bias

5.3.1 Representation. Our values prioritization process was prone to representation bias, which occurs when some values are underrepresented. Although representation bias can be reduced by collecting data for sufficiently including various subgroups [69], several practical limitations, such

as time [75], constrain the amount of data that can be collected in a research study. Moreover, constitutions and blockchain documents in our corpus were composed differently. Most constitutions express the dominant views of the national communities in sovereign states, whereas the blockchain documents express the perspectives of individual authors, based on their own understanding and research.

For constitutions, our frequency-based values prioritization posed a low risk of representation bias because the datasource included constitutions from almost all the countries in the world. For blockchain, our prioritization process utilized the knowledge of domain experts to compensate for values underrepresented in the source documents (see Table 1). Our approach is inspired by the domain experts driven technique of conditional upsampling for representation debiasing, presented by Bhattacharya et al. [8].

5.3.2 Interpretation. Additionally, our NER model used for polysemy resolution can cause bias by leaning towards particular interpretation of values. We used stop words to exclude items that definitely do not represent human values. We also used the Lesk algorithm based synset definitions to remove sentences that do not connote human values. Therefore, our model is inherently at least as biased as the Lesk algorithm, which was used for selecting the most relevant synset definitions.

5.3.3 Institutional Perspectives. We annotated English translations of the original collection of constitutions provided by the Constitute project [18]. Due to language and culture specific nuances, some values referenced in the original language may not translate appropriately to English. Also, text does not always represent the ground reality of values upheld by governments. Furthermore, sub constitutional legal systems constitute an important part of the legal framework besides the constitution. In other words, a constitution provides only a partial, but important picture of the legal landscape [76]. Hence, our analysis solely based on annotated constitutional text is not free from these biases.

6 Analysis

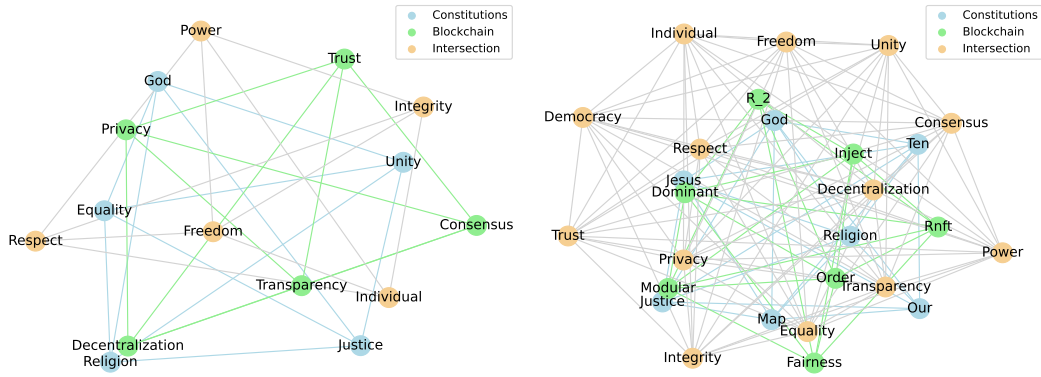
In this section, we utilize the datasets to examine the research questions raised in section 1. We show how to obtain insights directly from the datasets, and also inform the readers about potential limitations.

6.1 Prevalence

RQ1: *Which human values are most commonly referenced in blockchains and country constitutions?*

Based on the dataset, we found that *justice* was most frequent among the human values identified by the model in the constitutions dataset (Figure 5). This implies that constitutions are most concerned with people getting equal treatment from the governments, which can be extended to the people enjoying equal opportunities and rights. Though different societies understand the concept of justice differently, many constitutions repeatedly refer to justice. After all, reliance on law in all aspects of state governance and constitutionally anchored respect for the rule of law are of highest importance to the citizenry. In contrast, *consensus* is referenced the most in blockchain dataset. Consensus is a core feature in blockchain, and the mechanism by which several computer nodes in a decentralized network agree on the order of transactions.

We expected the model to learn the characteristics of human values and contextual semantics, and subsequently identify values mostly consistent with the prioritized sets created by annotators (see Table 2). Evidently, the model's output of the most frequent values in constitutions and blockchains does not exactly match the items and their orders in the annotators' lists of highest priority values. For blockchain, where annotations were based on 5 documents, 6 of the output are in the top 10



(a) Top 10 highest occurring values.

(b) Top 20 highest occurring values.

Fig. 4. **Human values specific to, and shared among blockchain and constitutions sets.** Elements within each set are linked. The top ten (left figure) most occurring values have fewer values in common (the **Intersection** set) compared to the top twenty (right figure) most occurring values. As the model is not perfectly accurate, words such as *internet* and *tual* were identified as human values.

Table 3. **Comparing Distributional Statistics of Datasets**. The test statistics and p-values (in parenthesis) indicate that the blockchain-constitutional values overlap data are distributed differently than each V-Dem index, namely, Electoral, Liberal, Participatory and Egalitarian. The distributions are independent both before and after discretization, as discussed in Section 6.2

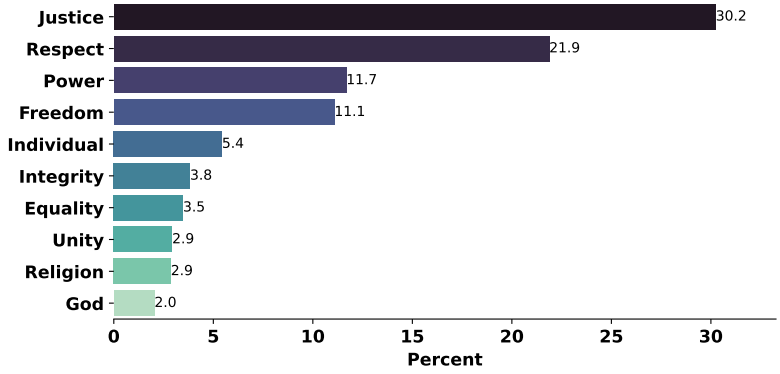
Statistic	Electoral	Liberal	Participatory	Egalitarian
Kolmogorov–Smirnov	0.395 (2.56e-14)	0.48 (3.13e-21)	0.495 (1.29e-22)	0.477 (1.1e-20)
Chi-Square	8.80 (0.012)	6.947 (0.031)	1.809 (0.404)	6.16 (0.046)

highest priority values, and 8 of the output are in top 20 highest priority values. Moreover, the model demonstrated the ability to generalize by identifying values that the annotators missed. Post training, the model found 1,330 uses in blockchain and 16 uses in constitutions of the ambiguous word *trust* applied in the human values sense. *Trust* was ranked at number 22 out of 79 by annotators, and number 2 out of 61 in the prioritized sets for constitutions and blockchain respectively.

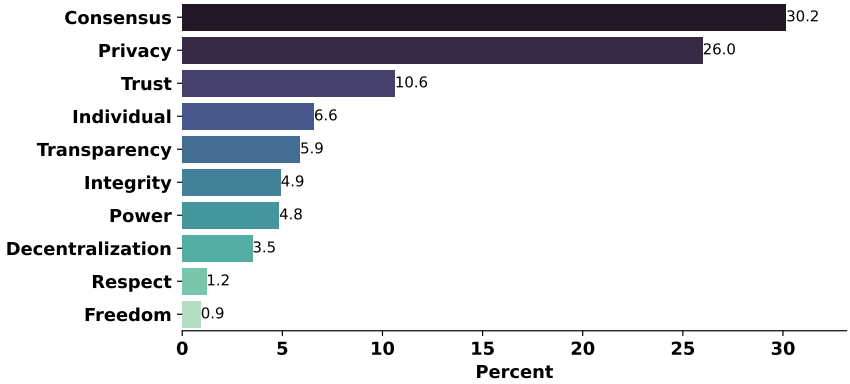
Next, consider the word *individual* which connotes a human value when it represents a person as opposed to a collective. The word was ranked 9 out of 79 by annotators in the prioritized set for constitutions, but not prioritized for blockchain. The model found 984 and 824 human values connotations of *individual* in constitutions and blockchain respectively. In constitutions, *individual* is frequently found in the context of individual rights. The model was able to identify equivalent usage of *individual* in blockchain, despite it not being prioritized by the annotators.

Furthermore, the model also identified values that annotators missed. For example, the model identified 370 and 10 human values connotations of *God* in constitutions and blockchain datasets.

RQ2: *Which human values do blockchains share with country constitutions?*



(a) Top ten most frequent human values in constitutions dataset



(b) Top ten most frequent human values in blockchain dataset

Fig. 5. **Top ten values ordered by proportions in (a) constitutions compared to (b) blockchain.** The proportions in each dataset are computed by dividing the number of occurrences of a value by the total number of values identified in that dataset.

Though in different manners, both blockchains and constitutions constitute documents that contain values important to individuals that participate in each respective grouping, namely, blockchain communities, and citizens of a particular country. After all, most individuals who belong to a group of any nature – political, social, etc. – have greater affinity with in-group values [63].

For the values shared among blockchain and world constitutions, there were four in the top ten most occurring values, namely, power, freedom, individual and life. Figure 4 illustrates the highest frequency values shared among blockchain and constitutions. The top twenty most occurring items contain more common values between the blockchain and constitution datasets. The model occasionally identified words such as *internet* and *tual* as human values, reflecting the inherent challenges in achieving perfect accuracy, as discussed in section 5.2.3.

Values are manifested in the governance mechanisms in blockchains and constitutions in diverse ways. As these domains are fundamentally different it can be difficult to draw exact parallels. Governance mechanisms in blockchains have different binding structures than those in constitutions.

Consider *power*, which is one of the most frequent values in constitutions as well as blockchain. The governance mechanisms of power differ between the two domains. Our blockchain dataset

contains the following sentence where the word power is applied in the context of voting in a Decentralized Autonomous Organization (DAO): "During this time, token holders can cast their votes directly on-chain, with their voting power proportional to the amount of AAVE, stkAAVE, or aAAVE they hold or have been delegated on Ethereum mainnet." Yes, it is correct that voting is a mechanism to exercise power in both DAOs and democratic governments. However, in DAOs, voting power can be distributed asymmetrically among the participants depending on their influence, specifically, voting power is proportional to the amount of tokens owned. In contrast, democratic governments usually put measures in place to foster equality of voting power. This exemplifies that from the perspective of actual realization, there is limited credibility in saying that power is a shared value between blockchain and constitutions.

6.2 Affiliations

RQ3: *Are the human values in blockchains uniformly represented worldwide, Eurocentric, Afrocentric, Americentric, or Asiacentric?*

Figure 6(top) illustrates the proportions of human values each country shares with constitutions from all the countries in our dataset. We find low proportions of human values represented in the older constitutions from sovereign states such as the United States, Canada and Australia. Comparatively, Figure 6(bottom) displays the proportion of values in blockchain that intersect with the values in different country constitutions. Overall, blockchain, which has developed recently (after the year 2000), represents human values across most countries worldwide with proportions greater than 0.5. The spread does not align with any particular region in the world.

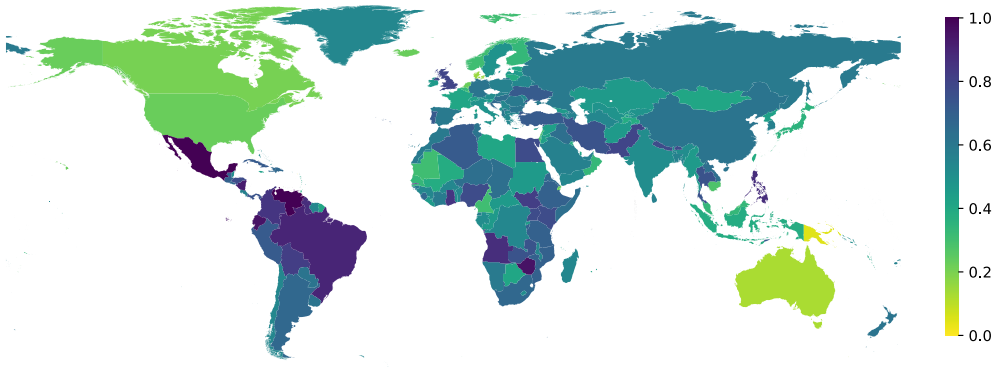
RQ4: *How democratic are the values incorporated in blockchains?*

Every democracy is different because there are many aspects of a democratic regime and its performance. By way of illustration, a country may have a fully functioning system for representative governance, but be plagued with a corrupt judiciary.

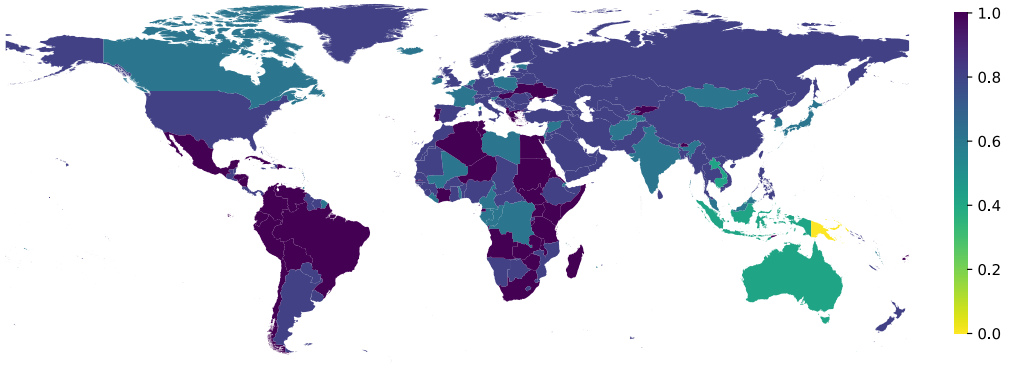
The Varieties of Democracy (V-Dem) [19] project provides various measures constructed from expert-based ratings on a wide range of indicators related to different aspects of democracy, then aggregating these ratings using a sophisticated statistical model, such as the Bayesian Item Response Theory [37]. To compare the human values in blockchain and the democracies around the world, we score different countries using the V-Dem project based indicators as summarized below. For detailed explanation of each index, see V-Dem methodology [77]:

- The *electoral* index embodies the value of making rulers responsive to citizens through occurrences of periodic elections.
- The *liberal* index embodies the intrinsic value of protecting individual and minority rights against a potential repression from the majority and state.
- The *egalitarian* index captures how much every electoral group enjoys equal de jure and de facto capabilities to participate, to serve in positions of political power, to put issues on the agenda, and to influence policymaking.
- The *participatory* index scores the active participation by citizens in all political processes, including elections, civil society organizations and other forms of both non-electoral and electoral mechanisms of direct democracy.

We used the V-Dem 2024 dataset provided by Our World in Data [7] project. Figure 7 illustrates V-Dem datasets for the above mentioned indices plotted over world maps. Setting aside the ground realities, it is reasonable to expect that countries with higher V-Dem rankings are more democratic



(a) Proportion of human values referenced in each country constitution to the human values referenced in constitutions corpus.



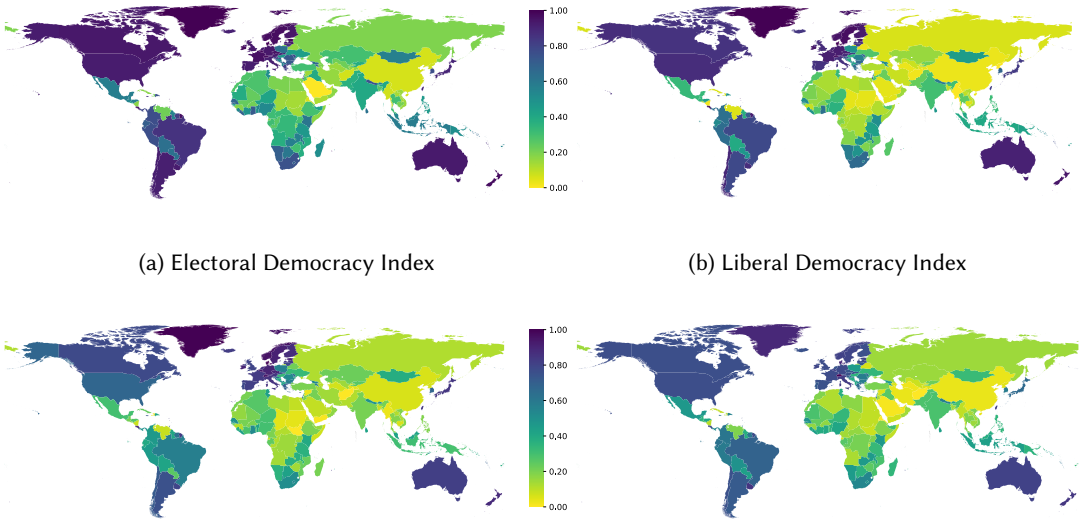
(b) Proportion of human values referenced in each country constitution to the human values referenced in blockchain corpus.

Fig. 6. **Worldwide distribution of proportional overlap of human values.** Figures (a) and (b) above illustrate the overlap of human values referenced in our constitutions and blockchain corpora with the human values represented across different country constitutions. Lighter colors indicate lower proportional overlap, while darker colors indicate higher overlap.

and incorporate greater human value-related provisions in their constitutions. To analyze quantitatively, let N_B denote the total number of unique human values mentioned in blockchain dataset, and N_x denote the number of unique human values mentioned in the x-th country's constitution, where $x = 1, 2, ..., T$ and T denotes the total number of countries. We compute the proportion of human values in blockchain that overlap with the human values in the x-th country's constitution as follows:

$$U(x) = \frac{N_B \cap N_x}{N_x} \qquad i = 1, 2, ..., T \tag{1}$$

Let $V(x)$ represent a V-Dem index values of the x-th country. The distributions of $U(x)$ and $V(x)$ can be compared using the Kolmogorov–Smirnov (KS) statistic to show that the two distributions



(a) Electoral Democracy Index

(b) Liberal Democracy Index

(c) Egalitarian Democracy Index

(d) Participatory Democracy Index

Fig. 7. **Measures of democracy using V-Dem [19] indices for different countries.** Each V-Dem index depicted in the above figures is described in Section 6.2. The values of V-Dem indices range from 0 to 1 (most democratic). Lighter colors in the figures indicate lower values of the V-Dem indices, while darker colors indicate higher values of the V-Dem indices. In Section 6.2, we analyze how democratic are the values in blockchain on the basis of these V-Dem measures of democracy.

are different. The results are summarized in Table 3.

$$\text{K-S stat} = sup_x |U(x) - V(x)| \tag{2}$$

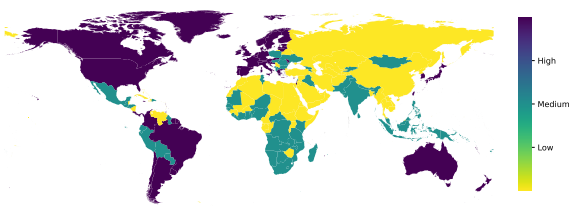
As $U(x)$ and $V(x)$ were computed differently, for a link-for-like comparison, we discretize their values as follows.

$$\begin{aligned}\Delta(x) &= \begin{cases} 0.75 & \text{if} \quad x \geq 0.70 \\ 0.50 & \text{if} \quad 0.30 \leq x < 0.70 \\ 0.25 & \text{if} \quad 0 \leq x < 0.30 \end{cases} \\ &= 0.75 \cdot \mathbb{1}_{[0.7,1]}(x) + 0.50 \cdot \mathbb{1}_{[0.3,0.7)}(x) + 0.25 \cdot \mathbb{1}_{[0,0.3)}(x) \end{aligned} \tag{3}$$

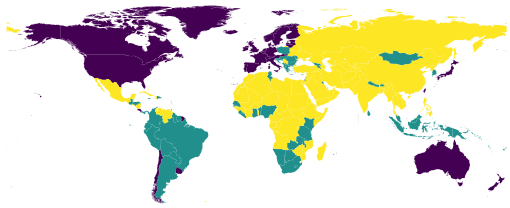
To verify that discretization does not alter the distributional properties and invalidate the analysis, we calculated chi-square statistics for the new samples. The chi-square statistics shown in Table 3 fail to reject the hypothesis that the distributions of the overlap scores and V-Dem indices are independent. Thus, the two distributions remain distinct after discretization.

The outputs of the discretization step are plotted in Figure 8 to compare the democracy score of each country with the proportion of human values in blockchain represented in that country's constitution. Overall, blockchain shares values with a variety of regime types, and the values overlap show central tendency. In other words, blockchain seems to represent diversity of human values across different political regimes.

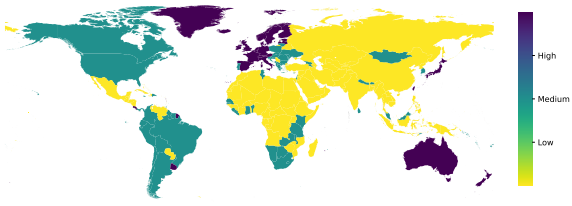
RQ5: *Do the human values in blockchains align with certain geopolitical ideologies?*



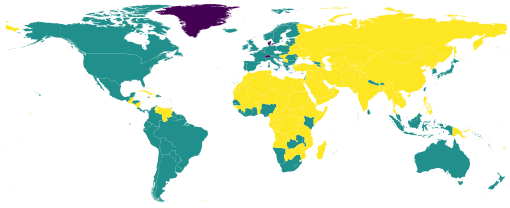
(a) Categorized Electoral Democracy Index



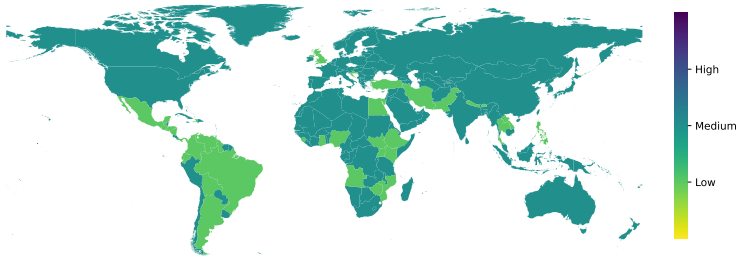
(b) Categorized Liberal Democracy Index



(c) Categorized Egalitarian Democracy Index



(d) Categorized Participatory Democracy Index



(e) Categorized Blockchain and Country Constitution Human Values Overlap

Fig. 8. **How democratic are the human values in blockchains?** To analyze this question, we compute the proportional overlap of human values in blockchain to the human values in constitutions of various regimes (countries). For a like-for-like comparison, we transform the measurements to a common scale. First, we categorize the democratic ratings of different countries published by the V-Dem project [19]. Our categorization of various V-Dem ratings are illustrated in (a), (b), (c) and (d). The original V-Dem ratings are categorized as Low (<0.3), Medium (>=0.3 and <0.7) and High (>=0.7) scores. Lighter colors indicate lower ratings, while darker colors indicate higher ratings. Next, we follow the same methodology to categorize the overlap between human values in blockchain corpus and the human values in constitutions corpus for each regime, as depicted in (e). Hence, we can draw observations for different countries. For example, the human values in the Australian constitution align closer with the participatory democracy index compared to other indices. Overall, human values in blockchains exhibit *centrism*, i.e. mostly medium (mid-range) overlap with the human values in constitutions from diverse regime types.

The analysis of political regimes can be nuanced. Aristotle classified political regimes into six categories, namely, monarchy, aristocracy, polity, tyranny, oligarchy and democracy [6]. The boundaries between different regime types have blurred in the modern era. Since the emergence of the information age in 1990s, technology and social media have become increasingly pervasive in

human lives. The control over information and intellectual property (IP) systems leads to power imbalances that parallel historical feudal structures, where a small elite wielded control over critical resources, now replaced by information and knowledge [25].

Figure 8(bottom) indicates that the human values found in blockchain overlap more with democratic regimes, with noteworthy exceptions such as Brazil. However, the overlap is also in the mid-range for countries with lower democratic index scores, such as Sudan and Saudi Arabia. These overlaps may suggest greater acceptance of more feudalistic or oligarchic values in technologies such as blockchain.

What are the implications of this for policy design? One must carefully consider traditional and ground realities, that is, how much a society cares about the values they have embraced since a long time. Simply because blockchain shares certain values prevalent in feudalistic regimes does not necessarily imply that those values will be easily accepted in democratic regimes where people can exercise their voices. However, the advancement of blockchain technology may make such policy discussions more pertinent. For policy development, it is also important to highlight that constitutions may represent specific legal traditions rather than political ideologies. For example, the constitutions of Islamic Law States embed a significant amount of Islam-based legal language as they try to navigate the balance between secular and religious laws in their legal systems [58–60]. We expect that these hypotheses and questions can be explored in future work using our datasets.

7 Conclusion and Future Work

In this study, we introduced a novel dataset and a systematic methodology for analyzing human values in blockchain documents and the constitutions of 194 sovereign states. By bridging socio-technical systems with legal frameworks, our work enables a comparative exploration of how human values are expressed and prioritized in decentralized technologies versus constitutional texts. Among the primary contributions of this research are the curated datasets on human values, which include annotated references from both blockchain-related documents and constitutions. These datasets provide a valuable resource for training machine learning models and conducting further research on the intersection of technology, ethics, and governance.

Our analysis revealed shared and distinct human values across blockchains and constitutions, shedding light on their contextual nuances and alignment with geopolitical ideologies and democratic indices. Notably, blockchains appear to reflect a diverse range of human values while demonstrating a centrist tendency, making them broadly applicable across various political regimes.

Building on these contributions, future research can explore several promising directions. First, researchers can investigate the evolution of human values in blockchains over time as these technologies adapt to societal needs and challenges. Second, enhancing word sense disambiguation models specific to human values can improve the accuracy of value extraction, addressing issues such as polysemy and contextual ambiguity. A limitation of our model was that it used only a smaller proportion of the prioritized human values. Hence, it missed to identify certain human values, for example forgiveness and gratitude, which had low counts in the dataset. We recognize that more complex models that utilize word embeddings (e.g. BERT encoder [22]) and transfer learning may perform better to identify less frequent human values. These can be pursued in a future work. Finally, the datasets introduced in this work create opportunities for comparative studies of human values in other socio-technical systems, enabling cross-domain insights into their ethical implications and global relevance.

By providing these datasets and insights, this work establishes a foundation for interdisciplinary research that can guide the design of value-sensitive technologies and inform discussions on the societal impact and ethical alignment of emerging systems like blockchain.

8 Acknowledgments

This research was partially funded by an NSF Award Number: 2131515, DASS: Principled Software Design and Accountability.

References

[1] Hayden Adams, Noah Zinsmeister, Moody Salem, River Keefer, and Dan Robinson. 2021. Uniswap v3 core. *Tech. rep., Uniswap, Tech. Rep.* (2021).
[2] Signe Agerskov, Asger Balle Pedersen, and Roman Beck. 2023. Ethical Guidelines for Blockchain Systems.. In *ECIS*.
[3] Eneko Agirre, Oier López de Lacalle, and Aitor Soroa. 2018. The risk of sub-optimal use of open source NLP software: UKB is inadvertently state-of-the-art in knowledge-based WSD. *arXiv preprint arXiv:1805.04277* (2018).
[4] Phyllis Ang, Bhuwan Dhingra, and Lisa Wu Wills. 2022. Characterizing the efficiency vs. accuracy trade-off for long-context NLP models. *arXiv preprint arXiv:2204.07288* (2022).
[5] Andreas M Antonopoulos and Gavin Wood. 2018. *Mastering ethereum: building smart contracts and dapps*. O'reilly Media.
[6] Derek WM Barker. 2013. Oligarchy or elite democracy? Aristotle and modern representative government. *New Political Science* 35, 4 (2013), 547–566.
[7] Bastian Herre and Lucas Rodés-Guirao and Esteban Ortiz-Ospina. 2013. Democracy. *Our World in Data* (2013). https://ourworldindata.org/democracy.
[8] Aditya Bhattacharya, Simone Stumpf, and Katrien Verbert. 2024. Representation Debiasing of Generated Data Involving Domain Experts. In *Adjunct Proceedings of the 32nd ACM Conference on User Modeling, Adaptation and Personalization*. 516–522.
[9] Reuben Binns. 2018. Fairness in machine learning: Lessons from political philosophy. In *Conference on fairness, accountability and transparency*. PMLR, 149–159.
[10] Steven Bird, Ewan Klein, and Edward Loper. 2009. *Natural language processing with Python: analyzing text with the natural language toolkit*. " O'Reilly Media, Inc.".
[11] Daniel M Brinks, Varun Gauri, and Kyle Shen. 2015. Social rights constitutionalism: negotiating the tension between the universal and the particular. *Annual Review of Law and Social Science* 11, 1 (2015), 289–308.
[12] Ann Brody and Stéphane Couture. 2021. Ideologies and imaginaries in blockchain communities: The case of Ethereum. *Canadian journal of communication* 46, 3 (2021), 543–561.
[13] Vitalik Buterin et al. 2013. Ethereum white paper. *GitHub repository* 1 (2013), 22–23.
[14] Toni Caradonna. 2020. Blockchain and society. *Informatik Spektrum* 43, 1 (2020), 40–52.
[15] Devendra Singh Chaplot and Ruslan Salakhutdinov. 2018. Knowledge-based word sense disambiguation using topic models. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 32.
[16] An-Shou Cheng and Kenneth R Fleischmann. 2010. Developing a meta-inventory of human values. *Proceedings of the American Society for Information Science and Technology* 47, 1 (2010), 1–10.
[17] Adam Chilton and Mila Versteeg. 2020. *How constitutional rights matter*. Oxford University Press.
[18] Constitute. 2025. Constitutions. https://constituteproject.org/constitutions
[19] Michael Coppedge, John Gerring, Carl Henrik Knutsen, Staffan I Lindberg, Jan Teorell, Nazifa Alizada, David Altman, Michael Bernhard, Agnes Cornell, M Steven Fish, et al. 2024. V-Dem Country-Year Dataset v14. https://doi.org/10.23696/mcwt-fr58
[20] Niladri Sekhar Dash. 2008. Context and contextual word meaning. *SKASE Journal of Theoretical Linguistics* (2008).
[21] Primavera De Filippi. 2018. *Blockchain and the Law: The Rule of Code*. Vol. 84. Harvard University Press.
[22] 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 American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)*. 4171–4186.
[23] Vikram Dhillon, David Metcalf, Max Hooper, Vikram Dhillon, David Metcalf, and Max Hooper. 2017. The hyperledger project. *Blockchain enabled applications: Understand the Blockchain ecosystem and how to make it work for you* (2017), 139–149.
[24] Cambridge Dictionary. 2024. Electronic resource. *URL: https://dictionary. cambridge. org/us/dictionary/english/value (date of access: 25.08. 2024)* (2024).
[25] Peter Drahos. 1995. Information feudalism in the information society. *The Information Society* 11, 3 (1995), 209–222.
[26] Electic Capital. 2023. Blockchain Developer Geography Analysis. https://github.com/electric-capital/developer-reports
[27] Zachary Elkins and Tom Ginsburg. 2021. What can we learn from written constitutions? *Annual Review of Political Science* 24, 1 (2021), 321–343.
[28] Zachary Elkins, Tom Ginsburg, and James Melton. 2009. *The endurance of national constitutions*. Cambridge University Press.

[29] Zachary Elkins, Tom Ginsburg, James Melton, Robert Shaffer, Juan F Sequeda, and Daniel P Miranker. 2014. Constitute: The world's constitutions to read, search, and compare. *Journal of web semantics* 27 (2014), 10–18.

[30] Zachary Elkins, Tom Ginsburg, and Beth Simmons. 2013. Getting to rights: Treaty ratification, constitutional convergence, and human rights practice. *Harv. Int'l LJ* 54 (2013), 61.

[31] Chris Elsden, Arthi Manohar, Jo Briggs, Mike Harding, Chris Speed, and John Vines. 2018. Making sense of blockchain applications: A typology for HCI. In *Proceedings of the 2018 chi conference on human factors in computing systems*. 1–14.

[32] Norman T Feather. 1995. Values, valences, and choice: The influences of values on the perceived attractiveness and choice of alternatives. *Journal of personality and social psychology* 68, 6 (1995), 1135.

[33] Lars P Feld and Stefan Voigt. 2003. Economic growth and judicial independence: cross-country evidence using a new set of indicators. *European Journal of Political Economy* 19, 3 (2003), 497–527.

[34] Filecoin Docs. 2025. Welcome to Filecoin Docs. https://docs.filecoin.io/

[35] Maxwell Forbes, Jena D Hwang, Vered Shwartz, Maarten Sap, and Yejin Choi. 2020. Social chemistry 101: Learning to reason about social and moral norms. *arXiv preprint arXiv:2011.00620* (2020).

[36] Fortune Business Insight. 2024. Blockchain Technology Market Size, Share & Industry Analysis. https://www.fortunebusinessinsights.com/industry-reports/blockchain-market-100072.

[37] JP Fox. 2010. Bayesian Item Response Modeling: Theory and Applications.

[38] Emilio Frangella and Lasse Herskind. 2022. Aave v3 technical paper. (2022).

[39] Batya Friedman, Peter H Kahn, Alan Borning, and Alina Huldtgren. 2013. Value sensitive design and information systems. *Early engagement and new technologies: Opening up the laboratory* (2013), 55–95.

[40] Iason Gabriel. 2020. Artificial intelligence, values, and alignment. *Minds and machines* 30, 3 (2020), 411–437.

[41] Dan Hendrycks, Collin Burns, Steven Basart, Andrew Critch, Jerry Li, Dawn Song, and Jacob Steinhardt. 2021. Aligning AI With Shared Human Values. *Proceedings of the International Conference on Learning Representations (ICLR)* (2021).

[42] M Honnibal and I SpaCy Montani. 2023. 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing. 2017. *Unpublished software application. https://spacy. io* (2023).

[43] Olivia Inwood and Michele Zappavigna. 2023. Ideology, attitudinal positioning, and the blockchain: A social semiotic approach to understanding the values construed in the whitepapers of blockchain start-ups. *Social Semiotics* 33, 3 (2023), 451–469.

[44] Sheila Jasanoff. 2016. *The ethics of invention: Technology and the human future*. WW Norton & Company.

[45] Clyde Kluckhohn. 1951. Values and value-orientations in the theory of action: An exploration in definition and classification. In *Toward a general theory of action*. Harvard university press, 388–433.

[46] L Kohlberg. 1976. Moral stages and moralization: The cognitive development approach. *Moral development and behavior: Theory, research and social issues/Holt, Rinehart and Winston* (1976).

[47] Michael Lesk. 1986. Automatic sense disambiguation using machine readable dictionaries: how to tell a pine cone from an ice cream cone. In *Proceedings of the 5th annual international conference on Systems documentation*. 24–26.

[48] Hongfang Liu, Virginia Teller, and Carol Friedman. 2004. A multi-aspect comparison study of supervised word sense disambiguation. *Journal of the American Medical Informatics Association* 11, 4 (2004), 320–331.

[49] Nicholas Lourie, Ronan Le Bras, and Yejin Choi. 2021. Scruples: A corpus of community ethical judgments on 32,000 real-life anecdotes. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 35. 13470–13479.

[50] George A Miller. 1995. WordNet: a lexical database for English. *Commun. ACM* 38, 11 (1995), 39–41.

[51] Eric Lease Morgan, Eroma Abeysinghe, Sudhkar Pamidighantam, Eric Coulter, Suresh Marru, and Marlon Pierce. 2019. The distant reader: Tool for reading. In *Proceedings of the Practice and Experience in Advanced Research Computing on Rise of the Machines (learning)*. 1–4.

[52] Satoshi Nakamoto. 2008. Bitcoin: A Peer-to-Peer Electronic Cash System. *Bitcoin.–URL: https://bitcoin. org/bitcoin. pdf* 4, 2 (2008), 15.

[53] Roberto Navigli. 2009. Word sense disambiguation: A survey. *ACM computing surveys (CSUR)* 41, 2 (2009), 1–69.

[54] Puja Ohlhaver, E Glen Weyl, and Vitalik Buterin. 2022. Decentralized society: Finding web3's soul. *Available at SSRN 4105763* (2022).

[55] Patrick Pantel and Dekang Lin. 2002. Discovering word senses from text. In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*. 613–619.

[56] Alex Pazaitis, Primavera De Filippi, and Vasilis Kostakis. 2017. Blockchain and value systems in the sharing economy: The illustrative case of Backfeed. *Technological Forecasting and Social Change* 125 (2017), 105–115.

[57] Simone Paolo Ponzetto and Roberto Navigli. 2010. Knowledge-rich word sense disambiguation rivaling supervised systems. In *Proceedings of the 48th annual meeting of the association for computational linguistics*. 1522–1531.

[58] Emilia Justyna Powell. 2019. *Islamic law and international law: Peaceful resolution of disputes*. Oxford University Press.

[59] Emilia Justyna Powell, Steven Christian McDowell, Robert O'Brien, and Julia Oksasoglu. 2021. Islam-based legal language and state governance: democracy, strength of the judiciary and human rights. *Constitutional Political Economy* 32 (2021), 376–412.

[60] Emilia Justyna Powell and Joshua Paldino. 2025. The Taliban, Afghan constitutionalism and modern Islamic law states: renegotiating the balance of religious and secular law. *Global Constitutionalism* (2025), 1–28.
[61] Milton Rokeach. 1973. The nature of human values.
[62] Milton Rokeach. 2008. *Understanding human values.* Simon and Schuster.
[63] Lilach Sagiv, Sonia Roccas, Jan Cieciuch, and Shalom H Schwartz. 2017. Personal values in human life. *Nature human behaviour* 1, 9 (2017), 630–639.
[64] Eduardo Salcedo and Manjul Gupta. 2021. The effects of individual-level espoused national cultural values on the willingness to use Bitcoin-like blockchain currencies. *International Journal of Information Management* 60 (2021), 102388.
[65] Simanta Shekhar Sarmah. 2018. Understanding blockchain technology. *Computer Science and Engineering* 8, 2 (2018), 23–29.
[66] Bianca Scarlini, Tommaso Pasini, Roberto Navigli, et al. 2020. With more contexts comes better performance: Contextualized sense embeddings for all-round word sense disambiguation. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. The Association for Computational Linguistics, 3528–3539.
[67] Shalom H Schwartz. 2009. Basic human values. *sociologie* 42 (2009), 249–288.
[68] Shalom H Schwartz and Wolfgang Bilsky. 1987. Toward a universal psychological structure of human values. *Journal of personality and social psychology* 53, 3 (1987), 550.
[69] Nima Shahbazi, Yin Lin, Abolfazl Asudeh, and HV Jagadish. 2023. Representation bias in data: A survey on identification and resolution techniques. *Comput. Surveys* 55, 13s (2023), 1–39.
[70] Monica M Sharif and Farshad Ghodoosi. 2022. The ethics of blockchain in organizations. *Journal of Business Ethics* 178, 4 (2022), 1009–1025.
[71] Doron Shulztiner and Guy E Carmi. 2014. Human dignity in national constitutions: functions, promises and dangers. *The American Journal of Comparative Law* 62, 2 (2014), 461–490.
[72] Nicholas Stifter, Aljosha Judmayer, Philipp Schindler, Alexei Zamyatin, and Edgar Weippl. 2018. Agreement with satoshi–on the formalization of nakamoto consensus. *Cryptology ePrint Archive* (2018).
[73] Whitney K Taylor. 2023. *The Social Constitution: Embedding Social Rights Through Legal Mobilization.* Cambridge University Press.
[74] The Comparative Constitutions Project. 2013. Constitute. https://comparativeconstitutionsproject.org/about-constitute. Accessed: 2024-12-08.
[75] Dimitrios Theofanidis and Antigoni Fountouki. 2018. Limitations and delimitations in the research process. *Perioperative Nursing-Quarterly scientific, online official journal of GORNA* 7, 3 September-December 2018 (2018), 155–163.
[76] Laurence H Tribe. 2008. *The invisible constitution.* Oxford University Press.
[77] University of Gothenburg, V-Dem Institute. 2021. Varieties of Democracy Methodology. https://www.v-dem.net/static/website/img/refs/methodologyv111.pdf
[78] Angela Walch. 2016. The path of the blockchain lexicon (and the law). *Rev. Banking & Fin. L.* 36 (2016), 713.
[79] Yinglin Wang, Ming Wang, and Hamido Fujita. 2020. Word sense disambiguation: A comprehensive knowledge exploitation framework. *Knowledge-Based Systems* 190 (2020), 105030.
[80] Sam Williams, Viktor Diordiiev, Lev Berman, and Ivan Uemlianin. 2019. Arweave: A protocol for economically sustainable information permanence. *Arweave Yellow Paper* (2019).
[81] Gavin Wood et al. 2014. Ethereum: A secure decentralised generalised transaction ledger. *Ethereum project yellow paper* 151, 2014 (2014), 1–32.
[82] Luyao Zhang, Xinshi Ma, and Yulin Liu. 2022. Sok: blockchain decentralization. *arXiv preprint arXiv:2205.04256* (2022).

Examples of values associated with WordNet synsets


```
{
  "power" : [
    {"power.n.01" : "possession of controlling influence"},
    {"ability.n.02" : "possession of the qualities (especially mental qualities) required to do something or get
something done"}
  ],
  "freedom" : [
    { "freedom.n.01" : "the condition of being free; the power to act or speak or think without externally imposed
restraints" },
    { "exemption.n.01" : "immunity from an obligation or duty" }
  ]
}
```


Examples of sentences with offsets denoting the location of values


```
[
  ["i never knew people could enjoy writing tests like you do, you've got my uttermost most respect.",
    {"entities": [[88, 95, "HUMANVALUE"]]}
  ],
  ["since the start of the project, one of our primary dreams has been to not just deliver a world-class product, but
also build a world-class organization, with quality of governance and transparency suitable for a foundation that
would be tasked with helping to organize the maintenance of the ethereum code for potentially decades to
come.",
    {"entities": [[184, 196, "HUMANVALUE"]]}
  ]
]
```


Fig. 9. **Intermediate outputs of the NER modeling process** (a) Examples of values associated with WordNet synsets: Words are associated different WordNet synsets that connote human values (b) Examples of sentences with offsets denoting the location of values: The first posits a human value *respect* starting at character 88 and ending at character 95. The second posits another human value *transparency* starting at character 184 and ends at character 196.

A Supplementary Analysis

A.1 Intermediate Outputs of the Model for Identifying Human Values

Figure 9 contains examples of the intermediate steps while developing the NER model. These steps highlight how the NER model was trained on the objective of identifying the human values.

A.2 Frequency characteristics

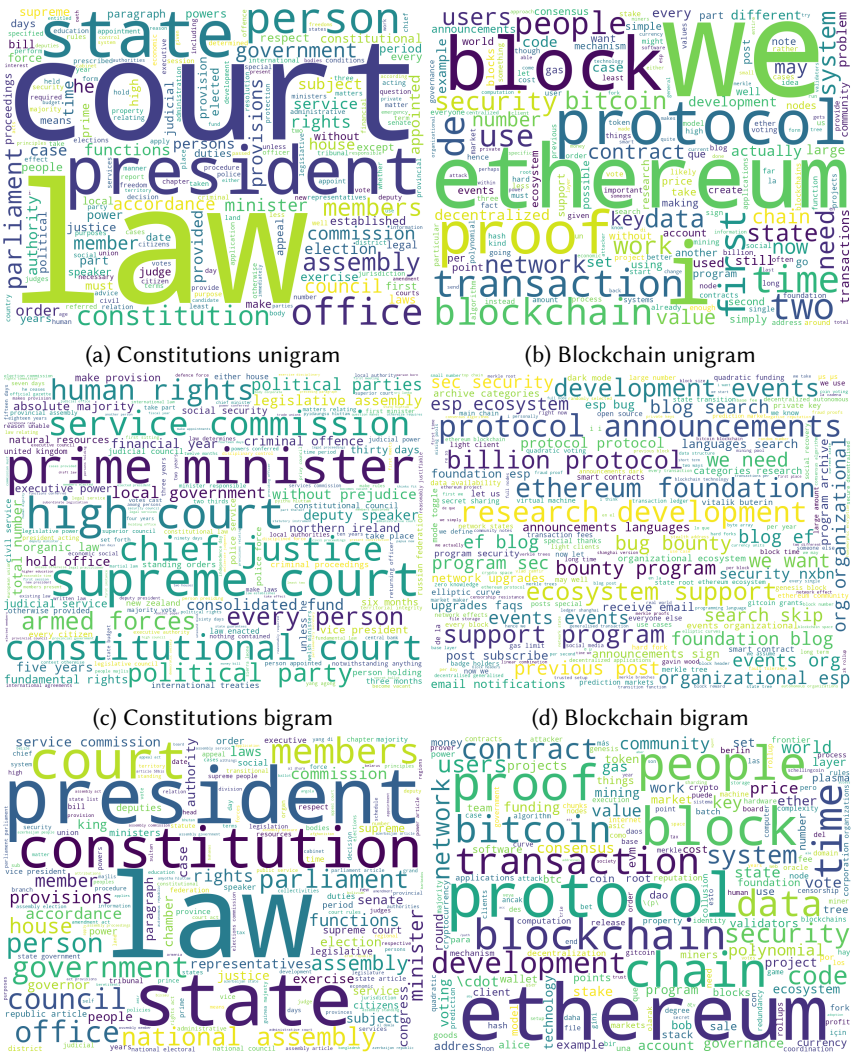
In this dataset, unigrams and bigrams exclude stop words, and represent the most frequent individual words and two-word combinations in the corpus. The computed keywords are akin to subject terms, and they help address the questions of, "What are the items in this carrel about and to what degree?"

Figure 10 illustrates the frequencies of unigrams, bigrams, and computed keywords from the corpus of constitutions, blockchains and combined datasets.

Table 4 contains the Hapax Richness Score, a statistic related to hapax legomena (words that occur only once), calculated by dividing the total number of unigrams by the total number of words (size of corpus). The scores indicate that the blockchain data set is the "richest", meaning, its vocabulary is the most diverse (or difficult to read).

Table 4. **Hapax Richness Scores**

Datasets	Hapaxes	Size	Score
curated-blockchain_discussions-2025	18,339	4,979,903	0.0037
curated-world_constitutions-2025	10,010	4,702,380	0.0021
curated-human_values-2025	25,734	9,450,145	0.0027



(a) Constitutions unigram (b) Blockchain unigram

(c) Constitutions bigram (d) Blockchain bigram

(e) Constitutions keywords (f) Blockchain keyword

Fig. 10. **Illustration of unigram, bigram and keyword frequencies in constitutions and blockchain datasets.**

A.3 Model Error Analysis

The precision scores of the final trained model with the validation dataset are shown in Table 5a and 5b, stratified by constitutions and blockchain datasets. For each word in the top 10 sets derived in Section 6.1, we analyzed whether the model's output with the validation dataset actually connoted human values. Ambiguous words are italicized in the table.

We analyzed the results with low precision scores (< 0.6). In the constitutions dataset, the word *justice* occurs several times in titles, for example, "If the office of Chief Justice or Deputy Chief Justice is vacant..." The word *respect* is used in phrases such as "in respect to", "with respect to", and "in respect of." The model misidentified *justice* and *respect* as human values in such cases.

Table 5. **Model precision with the validation dataset for the top 10 values**

(a) **Constitutions dataset**

Word	Precision
Justice	0.71
Respect	0.52
Power	0.96
Freedom	1.0
Individual	0.92
Integrity	1.0
Equality	1.0
Unity	1.0
Religion	1.0
God	1.0

(b) **Blockchain dataset**

Word	Precision
Consensus	0.99
Privacy	1.0
Trust	0.98
Individual	0.25
Transparency	1.0
Integrity	1.0
Power	0.81
Decentralization	1.0
Respect	0.05
Freedom	1.0

In the blockchain dataset, the word *individual* often describes inanimate objects, for example, "individual tokens" and "individual clocks" instead of a person. Also, similar to constitutions dataset, the word *respect* is mostly found to occur in phrases such as "in respect to." Moreover, the word *power* represents abstract quantities in several cases, for example, "power payments", "computer power", and "power law." The model misidentified *individual*, *respect* and *power* as human values in such cases.

The model also scored high on precision (>0.9) for several words. Are such high values implausible? It depends on data distribution. Here, high values indicate the fact that our datasets have almost no negative examples for such words, that is, these words almost always unambiguously connote human values in the corpus we used to create the datasets.

A.4 Topic Modeling

We apply topic modeling to identify twelve arbitrary topics in the combined constitutions and blockchain datasets. Table 6 summarizes the themes for each of these topics.

Table 6. **Topic modeling of the combined datasets**

Topics	Weights	Features
people	0.14722	people users system two time world price blockchain use community large market
law	0.11089	law state constitution rights president social congress established office right exercise accordance
ethereum	0.10174	ethereum protocol development announcements security bitcoin program events ethereum.org network research contract
block	0.09483	block transaction proof data state chain time blocks two transactions first need
president	0.07803	president constitution parliament commission office court person government assembly law state members
council	0.07616	law president council members assembly government state court constitutional constitution chamber right
state	0.06612	law state president right assembly government constitution court rights constitutional members council
person	0.06398	person court commission party election part proceedings district candidate judge paragraph provisions
office	0.04545	person office court law constitution member parliament commission minister functions provisions service
court	0.01982	state law court council constitution parliament president house provisions government part federation
provision	0.01412	court provision assembly order person parliament ireland minister proceedings relation northern part
hluttaw	0.00611	hluttaw union region que pyidaungsu pyithu para self-administered una los accord die