NLP4Gov: A Comprehensive Library for Computational Policy Analysis

Mahasweta Chakraborti
University of California Davis
Department of Communication
Davis, CA, USA
mchakraborti@ucdavis.edu

Sailendra Akash Bonagiri
University of California Davis
Department of Computer Science
Davis, CA, USA
sbonagiri@ucdavis.edu

Santiago Virgüez-Ruiz
University of Massachusetts Amherst
Department of Political Science
Amherst, MA, USA
svirguezruiz@umass.edu

Seth Frey
University of California Davis
Department of Communication
Davis, CA, USA
sethfrey@ucdavis.edu

ABSTRACT

Formal rules and policies are fundamental in formally specifying a social system: its operation, boundaries, processes, and even ontology. Recent scholarship has highlighted the role of formal policy in collective knowledge creation, game communities, the production of digital public goods, and national social media governance. Researchers have shown interest in how online communities convene tenable self-governance mechanisms to regulate member activities and distribute rights and privileges by designating responsibilities, roles, and hierarchies. We present NLP4Gov, an interactive kit to train and aid scholars and practitioners alike in computational policy analysis. The library explores and integrates methods and capabilities from computational linguistics and NLP to generate semantic and symbolic representations of community policies from text records. Versatile, documented, and accessible, NLP4Gov provides granular and comparative views into institutional structures and interactions, along with other information extraction capabilities for downstream analysis.

CCS CONCEPTS

• **Human-centered computing** → **Computer supported cooperative work**; **Empirical studies in collaborative and social computing**; *Social engineering (social sciences)*; **Open source software**; *Empirical studies in HCI*.

KEYWORDS

Open Source Software, Peer Production, Online Communities, Policy Analysis, Collective Action, OSS Governance

ACM Reference Format:
Mahasweta Chakraborti, Sailendra Akash Bonagiri, Santiago Virgüez-Ruiz, and Seth Frey. 2024. NLP4Gov: A Comprehensive Library for Computational Policy Analysis. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '24), May 11–16, 2024, Honolulu, HI, USA.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3613905.3650810

CHI EA '24, May 11–16, 2024, Honolulu, HI, USA

ACM ISBN 979-8-4007-0331-7/24/05
https://doi.org/10.1145/3613905.3650810

1 INTRODUCTION

Online governance has lately drawn significant research interest in HCI and been extensively studied through their instantiations in collective knowledge generation [25, 31, 32], content moderation [10, 11, 40], crowd-sourcing and production of digital public infrastructure [27], including software critical to infrastructure at large [14, 34, 45, 56–58]. Consequently, formal policy analysis is gaining traction in socio-technical systems research, particularly in how these collaborative initiatives regulate roles, rights, and responsibilities across the communities and beneficiaries involved.

Online communities are central in defining the virtual sphere and generating these goods and services of considerable economic value [4]. Moreover, they are also uniquely situated in their research significance, as circumstances motivating their emergent self-governance and mature policy systems are a microcosm of several other real-world challenges, particularly in public administration and natural resources management. Described as the 'Tragedy of the Commons" [29], natural resource conservation is especially complicated by 'free-riders' who harvest resources without reciprocal investments in production and upkeep. Policy functions to define resource and user boundaries, articulate sanctions, assign user rights and responsibilities, and overall implement sustainable management of finite resources [43, 54]. While digital goods are non-expendable, without adequate incentive structures and monitoring, collective interests, and benefits are similarly endangered by opportunistic users and limited contributions. Therefore, community policies are formal mechanisms formulated to mobilize participation and pool contributions, oversee maintenance and distribution of resources and goods generated, and also to monitor violations and other harmful behaviors [21, 22, 55]. Lessons from self-sustaining communities may indeed carry implications for critical questions in public life and have intrigued researchers from fields beyond HCI, such as social science, anthropology, economics, and public policy [17, 44, 49].

Significant contributions have been made lately in understanding the dynamics of governance activity, evolutionary trends, overlaps,

and differences in policy content as well as their scope and themes, whether in online policy systems [15, 42] or in policy analysis generally [47, 64, 66, 67, 72]. This has unlocked rich collections of governance documents, conversational corpus, and other archival footprints from communities. In order to derive key insights and patterns, empirical and data-driven researchers have explored computational methods to systematize and automate powerful and costly analytic approaches of policy analysis [19, 26, 52]. Yet relatively few scholars have pursued algorithmic approaches to semantically comprehend information from policy texts [9, 11], or derive the underlying cognitive structures [2, 13] and symbolic abstractions that bind institutional transactions [1, 20]. This is especially crucial as online collectivism continues to grow at an incredible pace, thus necessitating systematic, computationally scalable approaches to delve deeper into their governance behavior. NLP4Gov is a step to bridge this gap, as we extensively survey and string together concepts and approaches from computational linguistics and NLP to make computational policy analysis accessible to researchers across diverse backgrounds and questions. Built in a modular fashion while also supporting end-to-end data processing, NLP4Gov is intended to open up and demonstrate a range of possibilities for natural language understanding in the study of human behavior. It is built to assist data retrieval and provide detailed insights and visualizations into institutions, along with other information and measurements for downstream analysis.

2 RELATED WORK

Considerable HCI and social media scholarship have explored regulatory patterns across online platforms [16, 27]. Understanding the need for democratic governance and approaches for adaptive platform design has garnered focal interest with burgeoning user participation [36, 38, 55, 59, 69, 76]. Researchers have utilized advances in policy analysis, and increasingly available policy records and other governance instruments to characterize regulation approaches and their implications for community health and sustenance [47, 61]. Multiple studies have discovered associations between governance complexity and growth among online communities [23, 24, 33]. Several studies centered their design around policy records or related corpus. Temporal analysis of major vernacular Wikipedias revealed similar patterns of governance activity yet variation in rule composition over time [35]. Pater et al. found noticeable differences across multiple social media sites in their approach to harassment-related behavior [46]. In order to understand policy configurations across communities, Fiesler et al. codified rules written by Reddit moderators along their thematic and regulatory categories [15]. Policy typologies [48] have been applied to constitutions from private trackers to study how pirate communities operate [30]. Discourse from community threads has informed understanding of intellectual property protection in crowd-sourcing platforms [3]. Chandrasekhar et al. examined posts and threads to understand the impact of platform moderation tactics across subreddits [10]. In order to identify and mitigate abusive behavior online, content similarity-based approaches have been proposed that conserve human annotation and intuitively leverage existing big data traces for real-time governance [11]. Recent work in open source management has studied interrelationships between rules invoked by OSS developers in their daily operations and the sociotechnical evolution of their communities [74]. More recently, Chakraborti et al. studied formalization among OSS developers through email records by measuring routine activities described and their semantic internalization of written, established policies [9]. These scholastic developments open promising directions for big-data corpus in governance analysis and consolidates the motivation for a comprehensive kit to support cross-platform research and interdisciplinary convergence supported by state of art natural language processing.

3 SETUP AND USAGE

The NLP4Gov repository comprises a collection of interactive applications developed on Jupyter notebooks hosted through Google Colaboratory. We opted for Colaboratory for its dedicated, interactive environments, usability, and legibility for both proficient programmers and researchers with limited computational training. Importantly, it supports a range of self-contained platform-agnostic demonstrations, permits runtime requests to personal or external institutional systems, poses minimal setup requirements, and allows upgrades to storage and compute capabilities (through Google's own infrastructure) specific to research needs. The basic free version itself comes with CPU/GPU sufficient to execute the models (also the library defaults) and produce the results we describe in the current work.

The back-ends use transformer-based language models that were fine-tuned on standardized NLP benchmarks and further adapted by us for end-to-end policy analysis. The repository is designed to be amenable to researchers across disciplines and does not require extensive development experience for use. The notebooks simply require users to execute a small set of inline commands to load requisite scripts and obtain inferences on their data. Additionally, we provide extensive documentation on how to format data for specific applications and navigate the Colaboratory interface. We include ample datasets from different online communities such as Reddit [24], and OSS foundations like Apache [60, 75], and also data from public policy case studies [63, 65, 66] to extend general application, experimentation, and evaluation for researchers to assess how the applications may serve their goals.

4 APPLICATIONS AND PIPELINES

Researchers have analyzed policy content to understand the structural features of governance systems or discern similarities and uniqueness of regulatory patterns. Table 1. provides an overview of the six major applications we have currently developed. Each of these can be used independently or cascaded into pipelines. Subsequent sections further detail their development and potential analytical approaches they can support. Researchers interested in adapting the applications to more specific constructs of interest or wishing to incorporate domain-specific/newer language models may additionally modify the source to their requirements.

Comparative Policy Analysis

Policy_comparison explores semantic approaches to assess the similarity of policy texts at the conceptual level [37], and facilitates comparative analysis of governance. Classical methods such as regular expressions or lexical matching, while extremely powerful, may sometimes be limited in rating equivalence (or lack thereof)

Application	Description
ABDICO_coreferences	Substitutes determiners and other coreferences [37, 39] with the actual named entity over document sections
ABDICO_parsing	Institutional Grammar [12] parsing from individual policy statements
ABDICO_clustering	Clustering / topic modeling of policies or their components using Bertopic [28] and semantic embeddings [51]
Policy_comparison	Semantic comparison of policies across community databases
Policy_explore	Data retrieval with policies / governance topics from trace data / community records
Reddit_governance	Interactive policy comparison across popular subreddits [24]

Table 1: Summary of current applications and usage

between complex concepts articulated by policies. This is particularly true when related ideas are expressed through synonyms or statements with high word overlap yet bearing different implications.

The *Policy_Comparison* application builds on language models that interpret and encode text based on the context jointly conveyed by words within the sentences. They were trained in a manner such that conceptually similar statements are also encoded into representations closer to each other. For any given pair of policy databases, we generate a score for the mutual similarity between all pairs of policies through the cosine similarity between their encoded text embeddings. For encoding policies, we use a bi-encoder model [51] built on the MPNET architecture [68], which was trained on multiple datasets for broad, multi-domain usage[1]. The final result presents a list of policy pairs ranked in descending order of similarity.

Table 2. shows examples of rules and their mutual similarity produced using *Reddit_governance*. This demo app demonstrates Policy_comparison over the top 100 most popular of all the subreddits studied by Frey et al. [24]. Users may choose any pair of communities from a webform in the notebook to compare their policy databases. Having computed these similarities, a researcher could potentially use them to link related institutions or study community attributes explaining such similarities/dissimilarities.

Tracking Governance in Action

Practical incidents reported and deliberated by members are invaluable for tracking the impact and evolution of governance systems. While policies are generally concise statements, narratives from community threads and discussions are often longer, more expressive, and uniquely informative of lived realities. *Policy_explore* enables one to host their own search engine and discover interactions relevant to a policy or concept of interest. Also based on a semantic approach, this application pursues principles similar to *Policy_comparison*. However, it employs asymmetric search and performs a slightly elevated task of efficiently comparing texts of dissimilar lengths, i.e., policies and conversations. Therefore, the application incorporates models designed to encapsulate wider contexts for comparison with the search query. In Policy_explore, we use a Sentenceformer [51] model built on Distilbert [53] and trained on MSMARCO [41] for document and passage retrieval capabilities. The search query can either be a phrase (e.g., *"Content Moderation"*) or a full policy statement, and the application retrieves all posts/exchanges ranked in decreasing order of semantic relevance to the query. Example demonstrations walk users through retrieval of full-text emails related to policies from OSS community lists [9, 75] in the Apache Software Foundation.

Policy_explore is expected to supplement multiple research directions. Semantic data retrieval can help researchers focus on relevant data subsets related to governance, thus streamlining coding efforts. Case studies of community behavior across their relatedness to policy may provide rich insight into pertinent questions, such as their changing scope or interpretation, conditions that perpetrate exceptions/violations, and how communities address such challenges.

Institutional Analysis

4.0.1 Institutional Grammar Framework. Researchers have explored systematic approaches to represent the interdependencies spanning organizations and communities through the granular decomposition of the policies binding them. Notable among them is the Institutional Grammar (IG) [12], a comprehensive framework that takes a syntactic approach to decompose policy texts into granular units. It defines an institutional statement or a self-contained policy sentence as the most fundamental unit of policy analysis. The updated IG 2.0 specification [18, 67] lays down the taxonomy of policy constituents, with four core components being the attribute, object, aim, deontic, context, and or else. The aim (I) or the main verb of the statement specifies the central actions or goals of the institutional statement. The attribute (A) or agent is usually the grammatical subject and represents the individuals or organizations who are required to execute the aims outlined by a policy. The object (B) of an institutional statement (generally the grammatical object) spans other entities/organizational devices that are the target recipients of the policy aim. The deontic (D) conveys the strength of an institutional statement. It's the prescriptive component that states the extent to which the particular regulation is binding, typically through modals such as may, can, should, must, and shall, etc. Our library currently supports the extraction of these four components from text data, as these are most fundamental to discovering existing interaction networks and power structures within an institution.

4.0.2 Parsing Policy Constituents. There has been significant attention in recent years on automating IG parsing to draw insights from extensive collections of policy data. Vannoni [70] performed IG analysis through dependency parsing. Rice et al. [52] identified ABDICO components using neural network classifiers over dependency features. We hereby introduce our approach, which meticulously combines dependency structures and semantic role labeling (SRL). While dependency parsing identifies grammatical

[1]https://www.sbert.net/docs/pretrained_models.html, accessed 03/20/2024

Rule	Community	Rule	Community	Similarity Score
Medical advice is strictly prohibited on AskScience. Asking for or soliciting medical advice are both against the rules.	r/Ask Science	Offering or seeking medical advice is strictly prohibited and offending comments will be removed. Discussions regarding the advantages and/or disadvantages of certain treatments, diets, or supplements are allowed as long as relevant and reputable evidence is provided.	r/Science	0.67
AskScience has a strict policy against abusive and offensive language. Unless that language is in the context of research, it has no place here. We hold comments and posts to a high level of professionalism. We require our users and volunteers to always maintain a level of professionalism in order to participate.	r/Ask Science	* Threats, suggestions of harm, personal insults and personal attacks are prohibited	r/sports	0.53

Table 2: Examples of most related policies detected by relative cosine similarity between subreddits with similar and dissimilar interests

ABDICO Constituent	SRL-IG mapping
Aim (I)	ROOT of dependency tree (if verb) or predicate of the longest SRL graph
Attribute (A)	ARG0 of Aim if present else ARG1 if the statement has a nominal subject. None otherwise
Object (O)	Second highest core argument after the Attribute, in order of ARG1-ARG5
Deontic (D)	ARGM-MOD + ARGM-NEG

Table 3: Rule-based framework for combining Semantic Roles and Dependency Parsing for Institutional Grammar

Dataset	Attribute (F1)	Object (F1)	Aim (F1)	Deontic (F1)
Food Policy Data (N = 398)	0.71	0.57	0.82	0.94
Aquaculture Policy (N = 153)	0.76	0.57	0.81	0.93
National Organic Policy (N = 835)	0.86	0.49	0.84	0.94

Table 4: Performance of Semantic Role Labeling for parsing Institutional Grammar Constituents

tags of words, semantic role labeling/SRL is a computational linguistics approach that selects words/text spans within sentences that describe actor-objects, as well as conditions associated with a predicate (verb) in a given sentence. SRL comprises a set of core arguments, numbered ARG0-ARG5 and ranked in order of agentative precedence. ARG0 is generally the main agentative argument except for intransitive verbs called unaccusatives [5], in which case the verb's agent is annotated as ARG1. The remaining core arguments (after the agent) represent direct/indirect objects. The modals describing the verb and negation (e.g., "shall *not* submit a proposal") are classified as ARGM-MOD and ARGM-NEG, respectively.

We use pre-trained SRL models developed and validated on standardized datasets [73] to facilitate off-the-shelf use for IG parsing. Shi et al.[62] developed a BERT-based approach to parse semantic roles for every verb in a sentence. Since the aim concerns the central activity being regulated by an institutional statement, we assign the ROOT verb of a statement's dependency tree to the Aim using Stanza [50]. In cases where the ROOT of the statement is not a verb (e.g., certain subordinating clauses), we treat the verb with the most extensive SRL parsing (spanning most semantic roles and dependencies) as the aim. Next, we treat the Aim as the anchor and map other IG constituents to their semantic roles. Table 3. summarizes our approach. In cases where ARG0 is not present but the aim otherwise points to a direct subject, we treat ARG1 as the agent.

Table 4. reports the performance of our framework in terms of the word-constituent match on multiple datasets that were manually coded for IG constituents. Statements may often have the same modals governing multiple verbs, as a result the detection of Deontics is often unaffected by the exact predicate we chose as Aim or anchor for our SRL-IG mapping. Rice et al. reported performances higher for objects (0.75 F1), lower on attributes (0.62 F1), and comparable for aims and deontics on the Food Policy (FPC) dataset [52], a frequent benchmark used in IG-based studies. Despite some differences in our respective validation approaches[2] This reference highlights the need to improve our performance for extracting Objects. In order to assess methodological generalizability across different data distributions, we performed additional validation with two other datasets. We find similar trends in performances across components for all three cases, with room for further improvement in Objects. Further details on the datasets and processing are available in Appendix A.

[2]While we report performance on the full dataset without further training, Rice et al. developed and validated their approach specifically on the FPC dataset.

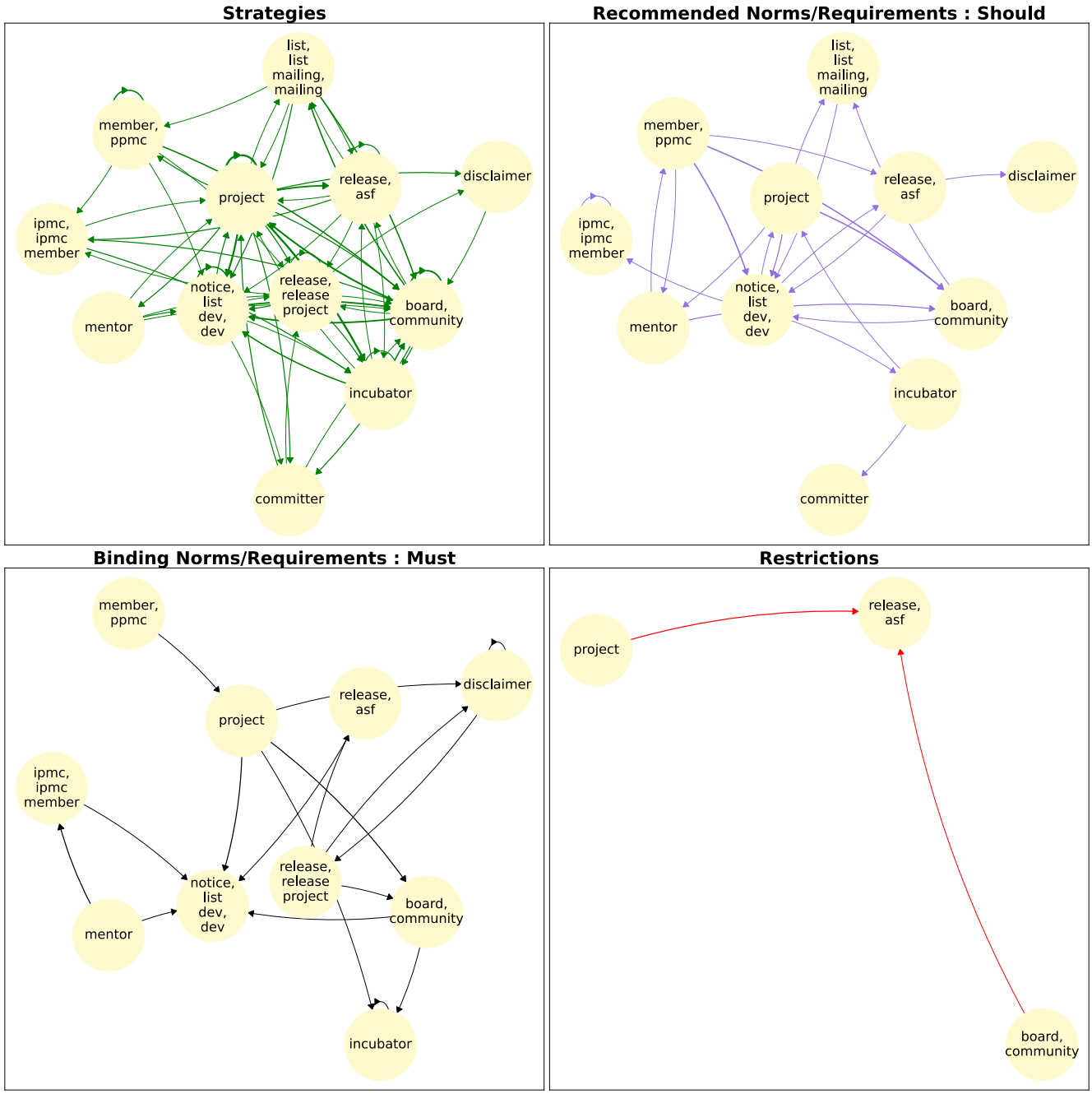

Figure 1: Projects in the Apache Software Foundation Incubator often comprise volunteer developers and are mostly directed by strategies, with fewer strong regulations or restrictions. The network was generated by parsing policy constituents and aggregating similar actors/objects into nodes. Edges are directed from actors to objects and are logarithmically weighted by the number of policies between the pair. Node labels are the top representative words from each cluster. Projects are nodal actors and objects and are subject to certain recommended and binding practices towards their communities, software releases, the Incubator, Foundation board, Management Committees (PPMC/IPMC), and mentors. Notably, there are very few restrictions, and they are only applicable with regard to project releases.

4.0.3 Modeling and Visualizing Institutions. The Institutional Analysis pipeline comprises three major NLP tasks. The first task involves coreference resolution (*ABDICO_coreferences*) using the implementation developed by Lee et al. [39]. Coreferences are pairs of words referring to the same entity or object, e.g., '*Paula* got a new sweater. *She* loves it'. This module reads sections from policy documents and substitutes pronouns, articles and other determiners across sentences with the actual named entity they are referring. This preprocessing preserves valuable information and continuity of context between sentences and policy components when documents are tokenized into institutional statements and parsed. Policy constituents can be then parsed using *ABDICO_parsing*, which has been described in Section 4.0.2.

The final module groups together the extracted components through their semantic similarity (*ABDICO_clustering*). This is particularly useful to discover patterns in the frequency and directionality of regulation between related institutional entities and instruments. It uses BERTopic [28], a versatile topic modeling library that also uses sentence embeddings [51] for text clustering, and can be extended to thematic categorization of policies or their constituents. IG components are used to derive qualitative inferences about institutions and even further inform downstream statistical and network analysis. Fig. 1 visualizes institutional interactions among OSS communities from the Apache Software Foundation Incubator [60]. We only consider institutional statements where both objects and attributes were explicitly specified, and for the purposes of interpretable aggregation set the minimum topic size at 10 components. We adopt the SNR (Strategy-norm-requirements) taxonomy from Frey et al. [24]. Strategies are day-to-day operations or optional processes (can/may) that community members adopt to their needs, while norms and requirements are denoted by stronger deontics, such as should (recommended practices) and must (binding). Norms and requirements are similar, except the latter often explicitly lay out penalties and consequences for non-compliance. For the purposes of the current illustration, we consider policies broadly under this deontic-based scheme. Finally, we separately consider restrictions (cannot/may not, etc.), which are policies that forbid/discourage certain practices and are characterized by negation of the aim.

5 FURTHER WORK

NLP4Gov is under active development as we continue to expand its capabilities along emerging directions in computational policy analysis, online collective action research, and language modeling. We are actively refining the existing features through domain adaptation of the pre-trained models with data from online communities and curation of validation benchmarks. Developments along policy decomposition could help realize granularity over a larger subset of the IG 2.0 [18] framework. Other pertinent components we intend to explore in the future are the "context" and "or else." The context (C) spans the conditions and constraints upon which the policy is carried out, while Or-else (O) administers 'monitoring' and lays out consequences or penalties in the event the policy non-compliance.

The current implementation of our Institutional Grammar parser is designed to work for simple institutional statements, while interpretation and representation of policy concepts and their interdependencies is a cognitively involved process. Lately, large language models, through their emergent capabilities, intrinsic knowledge, and superior semantic understanding [7], hold immense promise for our intended goals. We intend to devise and demonstrate pedagogical prompting [71] methods that can leverage large language models. These can further help comprehend the nested structures and elucidate the underlying relationships between institutional instruments and entities. Low resource learning approaches, including LLM-based, [6] can utilize limited expert annotations for confident and accurate predictions. Such methodologies, when adapted and demonstrated appropriately, are expected to be of particular interest to scholars who wish to operationalize different behavioral constructs over big data.

6 CONCLUSION

In the design of sociotechnical systems, it is often the case the advances in analyzing the specifying the "socio-" side lag behind "-technical" developments. In contrast to code, plain language articulations of formal governance structure suffer from all the ambiguities of human language, making written policy much more difficult to analyze and even design than the "code" part of sociotechnical law. Even where rigorous frameworks are available for policy analysis, they rely overwhelmingly on tedious hand-coding of subtle technical concepts, endangering inter-rater reliability, replicability, and generality, as quantitative studies of policies confine themselves to the analysis of single cases.

With NLP4Gov, we provide computational, quantitative representations of written policies, increasing the rigor, scale, replicability, and accessibility of policy analysis advances. As we have designed the toolkit, researchers with any level of programming proficiency can automatically perform a half dozen difficult policy analysis tasks, enabling semantic-level analysis and comparison of rules within and across institutions.

By increasing access to tools for formally representing policy systems, we hope to empower information scientists to offer users more powerful policy design tools while empowering policy analysts and governance scholars with more powerful insights into what policies work and why. The need for such tools is especially pressing as technological advances reveal new threats and opportunities in the online social systems that structure our lives.

ACKNOWLEDGMENTS

We would like to thank Saba Siddiki, Associate Professor of Public Administration and International Affairs, Syracuse University, Anamika Sen, Economics Ph.D. candidate at the University of Amherst, Massachusetts; and Likang Yin, Computer Science Ph.D. candidate at the University of California, Davis for data access. This work was supported by National Science Foundation grants #2020751 and #1917908.

REFERENCES

[1] Roba Abbas, Jeremy Pitt, and Katina Michael. 2021. Socio-Technical Design for Public Interest Technology. *IEEE Transactions on Technology and Society* 2, 2 (2021), 55–61. https://doi.org/10.1109/TTS.2021.3086260

[2] Alexander Artikis, Marek Sergot, and Jeremy Pitt. 2009. Specifying norm-governed computational societies. *ACM Transactions on Computational Logic (TOCL)* 10, 1 (2009), 1–42.

[3] Julia Bauer, Nikolaus Franke, and Philipp Tuertscher. 2016. Intellectual Property Norms in Online Communities: How User-Organized Intellectual Property

Regulation Supports Innovation. *Information Systems Research* 27, 4 (Dec. 2016), 724–750. https://doi.org/10.1287/isre.2016.0649

[4] Yochai Benkler. 2006. The Wealth of Networks: How Social Production Transforms Markets and Freedom.

[5] Claire Bonial, Olga Babko-Malaya, Jinho D Choi, Jena Hwang, and Martha Palmer. 2010. Propbank annotation guidelines. *Center for Computational Language and Education Research, CU-Boulder* 9 (2010), 90 pages.

[6] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language models are few-shot learners. *Advances in neural information processing systems* 33 (2020), 1877–1901.

[7] Sébastien Bubeck, Varun Chandrasekaran, Ronen Eldan, Johannes Gehrke, Eric Horvitz, Ece Kamar, Peter Lee, Yin Tat Lee, Yuanzhi Li, Scott Lundberg, et al. 2023. Sparks of artificial general intelligence: Early experiments with gpt-4. arXiv:2303.12712

[8] David P Carter, Christopher M Weible, Saba N Siddiki, and Xavier Basurto. 2016. Integrating core concepts from the institutional analysis and development framework for the systematic analysis of policy designs: An illustration from the US National Organic Program regulation. *Journal of Theoretical Politics* 28, 1 (2016), 159–185.

[9] Mahasweta Chakraborti, Curtis Atkisson, Stefan Stanciulescu, Vladimir Filkov, and Seth Frey. 2023. Do We Run How We Say We Run? Formalization and Practice of Governance in OSS Communities. arXiv:2309.14245

[10] Eshwar Chandrasekharan, Shagun Jhaver, Amy Bruckman, and Eric Gilbert. 2022. Quarantined! Examining the Effects of a Community-Wide Moderation Intervention on Reddit. *ACM Transactions on Computer-Human Interaction* 29, 4 (Aug. 2022), 1–26. https://doi.org/10.1145/3490499

[11] Eshwar Chandrasekharan, Mattia Samory, Anirudh Srinivasan, and Eric Gilbert. 2017. The Bag of Communities: Identifying Abusive Behavior Online with Preexisting Internet Data. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, Denver Colorado USA, 3175–3187. https://doi.org/10.1145/3025453.3026018

[12] Sue ES Crawford and Elinor Ostrom. 1995. A grammar of institutions. *American political science review* 89, 3 (1995), 582–600. Publisher: Cambridge University Press.

[13] Daniel A DeCaro, Marco A Janssen, and Allen Lee. 2021. Motivational foundations of communication, voluntary cooperation, and self-governance in a common-pool resource dilemma. *Current Research in Ecological and Social Psychology* 2 (2021), 100016.

[14] Nadia Eghbal. 2020. Working in public: the making and maintenance of open source software.

[15] Casey Fiesler, Jialun Jiang, Joshua McCann, Kyle Frye, and Jed Brubaker. 2018. Reddit rules! characterizing an ecosystem of governance. https://doi.org/10.1609/icwsm.v12i1.15033

[16] Adam Fish, Luis FR Murillo, Lilly Nguyen, Aaron Panofsky, and Christopher M. Kelty. 2011. Birds of the Internet: Towards a field guide to the organization and governance of participation. *Journal of Cultural Economy* 4, 2 (2011), 157–187. ISBN: 1753-0350 Publisher: Taylor & Francis.

[17] Anders Forsman, Tine De Moor, René Van Weeren, Mike Farjam, Molood Ale Ebrahim Dehkordi, Amineh Ghorbani, and Giangiacomo Bravo. 2021. Comparisons of historical Dutch commons inform about the long-term dynamics of social-ecological systems. *Plos one* 16, 8 (2021), e0256803.

[18] Christopher K Frantz and Saba Siddiki. 2021. Institutional Grammar 2.0: A specification for encoding and analyzing institutional design. *Public Administration* 99, 2 (2021), 222–247. Publisher: Wiley Online Library.

[19] Christopher K Frantz and Saba Siddiki. 2022. *Institutional Grammar*. Springer, New York, USA.

[20] Seth Frey, Jules Hedges, Joshua Tan, and Philipp Zahn. 2023. Composing games into complex institutions. *Plos one* 18, 3 (2023), e0283361.

[21] Seth Frey, PM Krafft, and Brian C Keegan. 2019. " This Place Does What It Was Built For" Designing Digital Institutions for Participatory Change. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–31.

[22] Seth Frey and Robert W Sumner. 2019. Emergence of integrated institutions in a large population of self-governing communities. *PloS one* 14, 7 (2019), e0216335.

[23] Seth Frey and Robert W. Sumner. 2019. Emergence of integrated institutions in a large population of self-governing communities. *PLOS ONE* 14, 7 (July 2019), e0216335. https://doi.org/10.1371/journal.pone.0216335

[24] Seth Frey, Qiankun Zhong, Beril Bulat, William D. Weisman, Caitlyn Liu, Stephen Fujimoto, Hannah Wang, and Charles M. Schweik. 2022. Governing Online Goods: Maturity and Formalization in Minecraft, Reddit, and World of Warcraft Communities. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (Nov. 2022), 1–23. https://doi.org/10.1145/3555191

[25] Brett M Frischmann, Michael J Madison, and Katherine Jo Strandburg. 2014. *Governing knowledge commons*. Oxford University Press, Oxford, UK.

[26] Amineh Ghorbani, Francien Dechesne, Virginia Dignum, and Catholijn Jonker. 2014. Enhancing ABM into an inevitable tool for policy analysis. *Policy and Complex Systems* 1, 1 (2014), 61–76.

[27] Tarleton Gillespie. 2010. The politics of 'platforms'. *New media & society* 12, 3 (2010), 347–364.

[28] Maarten Grootendorst. 2022. BERTopic: Neural topic modeling with a class-based TF-IDF procedure. arXiv:2203.05794

[29] Garrett Hardin. 1968. The tragedy of the commons: the population problem has no technical solution; it requires a fundamental extension in morality. *science* 162, 3859 (1968), 1243–1248.

[30] Colin Harris. 2018. Institutional solutions to free-riding in peer-to-peer networks: a case study of online pirate communities. *Journal of Institutional Economics* 14, 5 (Oct. 2018), 901–924. https://doi.org/10.1017/S1744137417000650

[31] Charlotte Hess and Elinor Ostrom. 2007. Understanding knowledge as a commons: From theory to practice.

[32] David J. Hess. 2005. Technology-and product-oriented movements: Approximating social movement studies and science and technology studies. *Science, Technology, & Human Values* 30, 4 (2005), 515–535.

[33] Benjamin Mako Hill and Aaron Shaw. 2021. The Hidden Costs of Requiring Accounts: Quasi-Experimental Evidence From Peer Production. *Communication Research* 48, 6 (Aug. 2021), 771–795. https://doi.org/10.1177/0093650220910345

[34] Eric von Hippel and Georg von Krogh. 2003. Open source software and the "private-collective" innovation model: Issues for organization science. *Organization science* 14, 2 (2003), 209–223.

[35] Sohyeon Hwang and Aaron Shaw. 2022. Rules and Rule-Making in the Five Largest Wikipedias. *Proceedings of the International AAAI Conference on Web and Social Media* 16 (May 2022), 347–357. https://doi.org/10.1609/icwsm.v16i1.19297

[36] Shagun Jhaver, Seth Frey, and Amy X Zhang. 2023. Decentralizing platform power: A design space of multi-level governance in online social platforms. *Social Media+ Society* 9, 4 (2023), 20563051231207857.

[37] Daniel Jurafsky and James H Martin. 2000. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition.

[38] Aniket Kittur, Jeffrey V. Nickerson, Michael Bernstein, Elizabeth Gerber, Aaron Shaw, John Zimmerman, Matt Lease, and John Horton. 2013. The future of crowd work. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work* (San Antonio, Texas, USA) *(CSCW '13)*. Association for Computing Machinery, New York, NY, USA, 1301–1318. https://doi.org/10.1145/2441776.2441923

[39] Kenton Lee, Luheng He, and Luke Zettlemoyer. 2018. Higher-Order Coreference Resolution with Coarse-to-Fine Inference. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, Marilyn Walker, Heng Ji, and Amanda Stent (Eds.). Association for Computational Linguistics, New Orleans, Louisiana, 687–692. https://doi.org/10.18653/v1/N18-2108

[40] J Nathan Matias. 2019. The civic labor of volunteer moderators online. *Social Media+ Society* 5, 2 (2019), 2056305119836778.

[41] Tri Nguyen, Mir Rosenberg, Xia Song, Jianfeng Gao, Saurabh Tiwary, Rangan Majumder, and Li Deng. 2016. MS MARCO: A human generated machine reading comprehension dataset. *choice* 2640 (2016), 660.

[42] Matthew N. Nicholson, Brian C Keegan, and Casey Fiesler. 2023. Mastodon Rules: Characterizing Formal Rules on Popular Mastodon Instances. In *Companion Publication of the 2023 Conference on Computer Supported Cooperative Work and Social Computing* (Minneapolis, MN, USA) *(CSCW '23 Companion)*. Association for Computing Machinery, New York, NY, USA, 86–90. https://doi.org/10.1145/3584931.3606970

[43] Elinor Ostrom. 1990. *Governing the Commons: The evolution of institutions for collective action*. Cambridge University Press, Cambridge, MA.

[44] Elinor Ostrom. 2009. *Understanding institutional diversity*. Princeton University Press, Princeton, NJ.

[45] Siobhán O'Mahony. 2003. Guarding the commons: how community managed software projects protect their work. *Research policy* 32, 7 (2003), 1179–1198.

[46] Jessica A. Pater, Moon K. Kim, Elizabeth D. Mynatt, and Casey Fiesler. 2016. Characterizations of Online Harassment: Comparing Policies Across Social Media Platforms. In *Proceedings of the 19th International Conference on Supporting Group Work*. ACM, Sanibel Island Florida USA, 369–374. https://doi.org/10.1145/2957276.2957297

[47] LEAH PIEPER, SANTIAGO VIRGÜEZ, EDELLA SCHLAGER, and CHARLIE SCHWEIK. 2023. The Use of the Institutional Grammar 1.0 for Institutional Analysis: A Literature Review. *International Journal of the Commons* 17, 1 (2023), pp. 256–270. https://www.jstor.org/stable/48756450

[48] Margaret M. Polski and Elinor Ostrom. 1999. An institutional framework for policy analysis and design.

[49] Amy R Poteete, Marco A Janssen, and Elinor Ostrom. 2010. *Working together: collective action, the commons, and multiple methods in practice*. Princeton University Press, Princeton, NJ.

[50] Peng Qi, Yuhao Zhang, Yuhui Zhang, Jason Bolton, and Christopher D Manning. 2020. Stanza: A Python Natural Language Processing Toolkit for Many Human Languages. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*. Association for Computational Linguistics, Online, 101–108.

[51] Nils Reimers and Iryna Gurevych. 2019. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. In *Proceedings of the 2019 Conference on Empirical*

Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). Association for Computational Linguistics, Hong Kong, China, 3980–3990. https://doi.org/10.18653/v1/D19-1410

[52] Douglas Rice, Saba Siddiki, Seth Frey, Jay H Kwon, and Adam Sawyer. 2021. Machine coding of policy texts with the Institutional Grammar. *Public Administration* 99, 2 (2021), 248–262. Publisher: Wiley Online Library.

[53] Victor Sanh, Lysandre Debut, Julien Chaumond, and Thomas Wolf. 2019. DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. arXiv:1910.01108

[54] Edella Schlager and E. Ostrom. 1992. Property-Rights Regimes and Natural Resources: A Conceptual Analysis. *Land Economics* 68 (1992), 249. https://api.semanticscholar.org/CorpusID:2908275

[55] Nathan Schneider, Primavera De Filippi, Seth Frey, Joshua Z. Tan, and Amy X. Zhang. 2021. Modular Politics: Toward a Governance Layer for Online Communities. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (April 2021), 1–26. https://doi.org/10.1145/3449090

[56] Charles M Schweik and Robert English. 2007. Tragedy of the FOSS commons? Investigating the institutional designs of free/libre and open source software projects. *First Monday* 12 (2007), 35 pages. https://doi.org/10.5210/fm.v12i2.1619

[57] Charles M. Schweik and Robert C. English. 2012. *Internet success: a study of open-source software commons*. MIT Press, Cambridge, MA.

[58] Charles M. Schweik and Meelis Kitsing. 2010. Applying Elinor Ostrom's Rule Classification Framework to the Analysis of Open Source Software Commons. *Transnational Corporations Review* 2, 1 (Jan 2010), 13–26. https://doi.org/10.1080/19186444.2010.11658219

[59] Joseph Seering. 2020. Reconsidering Self-Moderation: the Role of Research in Supporting Community-Based Models for Online Content Moderation. *Proc. ACM Hum.-Comput. Interact.* 4, CSCW2, Article 107 (oct 2020), 28 pages. https://doi.org/10.1145/3415178

[60] Anamika Sen, Curtis Atkisson, and Charlie Schweik. 2022. Cui Bono: Do open source software incubator policies and procedures benefit the projects or the incubator? *International Journal of the Commons* 16, 1 (2022), 64–77.

[61] Aaron Shaw and Benjamin M. Hill. 2014. Laboratories of Oligarchy? How the Iron Law Extends to Peer Production: Laboratories of Oligarchy. *Journal of Communication* 64, 2 (Apr 2014), 215–238. https://doi.org/10.1111/jcom.12082

[62] Peng Shi and Jimmy Lin. 2019. Simple bert models for relation extraction and semantic role labeling. arXiv:1904.05255

[63] Saba Siddiki. 2014. Assessing Policy Design and Interpretation: An Institutions-Based Analysis in the Context of Aquaculture in F lorida and V irginia, U nited S tates. *Review of Policy Research* 31, 4 (2014), 281–303.

[64] Saba Siddiki and Graham Ambrose. 2023. Evaluating Change in Representation and Coordination in Collaborative Governance Over Time: A Study of Environmental Justice Councils. *Environmental Management* 71, 3 (2023), 620–640.

[65] Saba Siddiki, Xavier Basurto, and Christopher M Weible. 2012. Using the institutional grammar tool to understand regulatory compliance: The case of Colorado aquaculture. *Regulation & Governance* 6, 2 (2012), 167–188.

[66] Saba Siddiki and Christopher Frantz. 2023. Understanding the Effects of Social Value Orientations in Shaping Regulatory Outcomes through Agent-Based Modeling: An Application in Organic Farming. https://doi.org/10.4000/irpp.3398

[67] Saba Siddiki, Tanya Heikkila, Christopher M. Weible, Raul Pacheco-Vega, David Carter, Cali Curley, Aaron Deslatte, and Abby Bennett. 2022. Institutional analysis with the institutional grammar. *Policy Studies Journal* 50, 2 (2022), 315–339.

[68] Kaitao Song, Xu Tan, Tao Qin, Jianfeng Lu, and Tie-Yan Liu. 2020. Mpnet: Masked and permuted pre-training for language understanding. *Advances in Neural Information Processing Systems* 33 (2020), 16857–16867.

[69] Ofer Tchernichovski, Seth Frey, Nori Jacoby, and Dalton Conley. 2021. Experimenting With Online Governance. *Frontiers in Human Dynamics* 3 (April 2021), 629285. https://doi.org/10.3389/fhumd.2021.629285

[70] Matia Vannoni. 2022. A political economy approach to the grammar of institutions: Theory and methods. *Policy Studies Journal* 50, 2 (2022), 453–471.

[71] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. 2022. Chain-of-thought prompting elicits reasoning in large language models. *Advances in Neural Information Processing Systems* 35 (2022), 24824–24837.

[72] Christopher M Weible, Saba N Siddiki, and Jonathan J Pierce. 2011. Foes to friends: Changing contexts and changing intergroup perceptions. *Journal of Comparative Policy Analysis: Research and Practice* 13, 5 (2011), 499–525.

[73] Ralph Weischedel, Martha Palmer, Mitchell Marcus, Eduard Hovy, Sameer Pradhan, Lance Ramshaw, Nianwen Xue, Ann Taylor, Jeff Kaufman, Michelle Franchini, et al. 2013. Ontonotes release 5.0 ldc2013t19. *Linguistic Data Consortium, Philadelphia, PA* 23 (2013), 170.

[74] Likang Yin, Mahasweta Chakraborti, Yibo Yan, Charles Schweik, Seth Frey, and Vladimir Filkov. 2022. Open Source Software Sustainability: Combining Institutional Analysis and Socio-Technical Networks. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (2022), 1–23.

[75] Likang Yin, Zhiyuan Zhang, Qi Xuan, and Vladimir Filkov. 2021. Apache software foundation incubator project sustainability dataset. In *2021 IEEE/ACM 18th International Conference on Mining Software Repositories (MSR)*. IEEE, Madrid, Spain, 595–599. https://doi.org/10.1109/MSR52588.2021.00081

[76] Amy X. Zhang, Grant Hugh, and Michael S. Bernstein. 2020. PolicyKit: Building Governance in Online Communities. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*. ACM, Virtual Event USA, 365–378. https://doi.org/10.1145/3379337.3415858

A VALIDATION DATA: INSTITUTIONAL GRAMMAR

The Food Policy data [52, 63] comprises 19 documents coded according to the Institutional Grammar (IG). We reconstructed polices from the word-level annotations used by Rice et al. (excluding punctuations). We also evaluated our approach on two other datasets. The Colorado aquaculture rules [65] encompass 153 institutional statements. These, too, were developed by applying IG (ABDICO syntax [18]) to three regulatory documents governing aquaculture practices in the State of Colorado. These include the Colorado Aquaculture Act Statute (59 statements), the Rules Pertaining to the Administration and Enforcement of the Colorado Aquaculture Act (54 statements), and the section on fish health in the Colorado Division of Wildlife regulations (40 statements). Lastly, the National Organic Program regulations dataset [8, 66] comprises approximately 1078 IG-coded institutional statements.

Despite the increasing popularity of systematic corpus-based policy research, the availability of annotated datasets for ML-based development/validation is still limited. These datasets were collected for diverse case studies, and each was annotated specific to particular research questions and analytical approaches. Moreover, their usage for automated evaluation of language technologies was particularly challenging due to subtle differences in styles and subjective discretion of the annotators. We hereby describe some additional pre-processing steps that were necessary to ensure uniform evaluation.

For the National Organic Program and Aquaculture datasets, several constituents that were not explicitly specified in the statement itself were found to have been filled in by annotators through external sources of information. These entries were unsuitable for validating algorithms meant to extract information only from available text inputs. We performed a preliminary cleaning from all three datasets to exclude policies that were not coded (e.g., bulleted items) or only carried abstractive/implicit annotations for all ABDI constituents. We report the number of statements retained for evaluation in Table. 4. For statements carrying one or more implicit ADBI labels, these specific constituents were further excluded at the time of evaluation. We also noticed and resolved some coding consistencies across the datasets prior to evaluation. We reduced verbs to their lemma root for matching (i.e., "Driving" was represented as "Drive") to account for differences in tense for annotations of 'Aim'. We further exclude non-informative words (stopwords) from word-matching, i.e., "The Committee" and "Committee" are treated as one and the same.