Supporting Participation Processes Using NLP in Constrained Resources Settings

Jose A. Guridi[1,*], Julio A. Pertuze[2,*] and Catalina Zamora[2]

[1]*Cornell University, Ithaca NY, USA*

[2]*Pontificia Universidad Católica de Chile, Santiago, Chile*

Abstract

For policymakers, making sense of stakeholder participatory data is a complex task. Natural Language Processing (NLP) can aid in processing this data, reducing policymakers' cognitive overload and supporting multi-stakeholder engagement. However, implementing NLP can be challenging in settings with limited resources, knowledge, or infrastructure. This study analyzes the feasibility and limitations of using Latent Dirichlet Allocation (LDA) to examine data from Chile's AI policy in which more than 1,700 people participated in a public deliberation process yielding data containing citizen reflections that varied in format, quality, depth, and length. We matched LDA topics from the public deliberation data to the objectives of Chile's AI policy draft, written by five experts over 4-months. LDA effectively detected 87% of the topics in the draft, requiring the researchers only to manually inspect 26% of the participation data to deliver this result. We discuss the potential and limitations of using LDA in participatory processes and contribute by showing how it can aid in the strategic management of stakeholders in a real-world resource-constrained setting.

Keywords

Participation, Natural Language Processing, Artificial Intelligence, Stakeholders

1. Introduction

Policymakers struggle when planning and managing multi-stakeholder participation processes. Participatory data often lack structure, consist of atomic units, and exhibit considerable divergence, resulting in heterogeneity in content and format [1, 2, 3]. When information overloads policymakers, they start filtering information, ignoring some or all inputs [4], which can make policies fail in the short term and erode legitimacy and trust.

Natural Language Processing (NLP) can support policymakers in engaging multiple stakeholders by augmenting their information processing capacities[e.g., 5, 6, 7]. However, literature has raised concerns and challenges about using artificial intelligence (AI) systems in the public sector [8]. When using NLP, in particular, many approaches require high-quality datasets or sophisticated algorithms [9], which are not always present in real-world policy settings. To address this issue, we explore the following research question: **Can topic modeling identify**

the main topics that policymakers do when analyzing data from a deliberative process unprepared for automated analysis?

To answer this question, we explore how a simple-to-implement topic modeling algorithm can be applied to reconcile different stakeholders' needs in data from a deliberative process not curated for automated analysis. Topic modeling is an NLP approach that can identify topics through a statistical analysis of the words and their relations in documents. These techniques can extract valuable insights, identify patterns, and discern emerging themes from text inputs provided by citizens [7, 10, 5, 6]. Among diverse topic modeling techniques, Latent Dirichlet Allocation (LDA) is considered cost-effective and achieves high-quality topics when rigorously trained despite requiring iteration and manual inspection [7].

We used LDA to analyze stakeholders' deliberation data generated during the drafting phase of the Chilean AI Policy. The Chilean case is unique since around 1,700 people participated in a public deliberation process via self-organized unstructured processes and regional structured discussions. Both participatory formats yielded documents containing citizen reflections, which varied in format, quality, depth, and length.

We compared the results of the LDA analysis to Chile's AI policy draft, manually written from the analysis of five experts over 4-months. We found that LDA could extract most of the topics that policymakers identified and extracted from the deliberation transcripts. Specifically, LDA found 87% of the topics in the AI policy draft. LDA was also efficient as it only required manually assessing 26% of the documents to assign meaning to the topics. LDA was also robust to structured and unstructured participation formats. The topics obtained from the structured regional roundtables covered 83% (19/23) of the topics in the AI policy draft, and the topics from unstructured self-organized roundtables covered 78% (18/23).

This study empirically assesses the effectiveness of using LDA to analyze real-world public deliberation data from multiple stakeholders. Rather than refining the algorithm to enhance output precision and quality, we focused on how policymakers can use straightforward and accessible NLP techniques to scrutinize heterogeneous and unstructured participatory data. Our results underscore that policymakers can benefit from using LDA in large stakeholder datasets even with unprepared data and limited resources. Policymakers can decrease their cognitive load, freeing capacities to strategically design stakeholders' journeys and plan for value co-creation in a complex multi-stakeholder ecosystem.

2. Related Work

2.1. Multi-stakeholder processes in technology policy

Governments must manage internal and external stakeholders when designing and implementing policies [11, 12]. Stakeholders, however, differ in terms of size and interests both within and outside government organizations [13, 14, 15]. Given stakeholders' heterogeneity, there are differences in managing them regarding resources, conditions for participation, and the extent of their involvement in the policy process [16].

Multi-stakeholder deliberation generates epistemic, democratic, and economic value that can benefit policymaking [17]. By engaging a broader set of stakeholders, policymakers can collect more information and access experience-based knowledge and expertise [9, 17]. Participation

can also generate collective learning and increase the likelihood of delivering more legitimate and effective policies [18, 9]. The latter improves democratic institutions over the long term by enhancing the process's inclusiveness, transparency, and accountability [17, 19].

If policymakers fail to listen to stakeholders and respond timely, participation processes and institutions can lose legitimacy [20, 4, 21, 22], which can hurt democracy in the long term [23, 24]. Engaging stakeholders requires policymakers to be transparent about the process's purpose and scope and recognize the different needs and knowledge levels of the different groups [25, 15]. Thus, policymakers must manage stakeholders depending on their salience to successfully implement public projects through participatory processes [26, 27, 28].

Policymakers, however, face a significant challenge in analyzing the inputs received from stakeholders during large-scale deliberation processes. Participatory data is often unstructured, atomic, and divergent, making it heterogeneous in content and format [1, 2, 3, 29]. Policymakers struggle with identifying potential gaps in data, assessing its quality, and using it effectively to respond to citizens promptly [30, 3, 31, 22, 32]. As a result, policymakers often use various heuristics to filter information when analyzing complex civic data [4]. For instance, policymakers may determine that the public cannot contribute significantly to the process, or they may only select those aligned with their beliefs. This tendency can lead policymakers to dismiss participation outputs without providing reasonable explanations, often because they need help making sense of it [4, 33, 34].

2.2. NLP and Public Participation

NLP can complement policymakers' work and help them to analyze complex data generated from public participation processes (see [9] for a review). NLP techniques can help to summarize the contents of public contributions and create visualizations to share the results with the public [6, 10, 35]. There are limitations, however, to using NLP techniques, and different types of algorithms have varying levels of effectiveness [5, 10].

NLP techniques can be supervised, semi-supervised, or unsupervised. Supervised learning can better support policymakers in categorizing contributions [e.g., 36, 5, 10] but requires existing knowledge, dictionaries, and higher-quality data [5, 10]. Semi-supervised methods, such as active learning, show promising results [5]. However, their low applicability to heterogeneous and long texts and runtime dependency on the GPU can be prohibitive. Unsupervised learning algorithms [e.g., 37, 6, 7] have also been used to summarize public contributions, but most of them require an intensive manual analysis to define the number and meaning of topics that make sense to policymakers [5, 10].

Latent Dirichlet Allocation (LDA) has gained popularity in settings with limited resources, knowledge, or infrastructure because it can be cost-effective and achieve high-quality topics when rigorously trained, despite requiring iteration and manual inspection [7, 38]. This article assesses the effectiveness of LDA in identifying the main topics discussed in the public deliberative process leading to the drafting of Chile's AI policy. Our aim is not to develop new methods but to understand how policymakers could use LDA without intensive data preparation in a language different than English (Spanish) and with no significant adaptations to the algorithm.

3. Case: The Chilean AI Policy

In August 2019, Chile's President Piñera sanctioned the creation of an AI policy led by the Ministry of Science, Technology, Knowledge, and Innovation (MinCTCI). This Ministry assembled an advisory committee composed of experts from academia, industry, and civil society, which suggested organizing the policy discussion around three main themes: (1) enabling factors, (2) development and adoption, and (3) ethics, legal and regulatory issues. Each theme was further divided into sub-themes.

MinCTCI launched a participatory process that included self-organized and regional roundtables to foster deliberation between multiple stakeholders. Over 1,300 people from academia, industry, the public sector, and civil society participated in the self-organized roundtables, and over 400 people in the regional roundtables. The data collected through this process varied in quality, depth, and length. Those who organized their roundtables submitted their opinions via Google Forms with open text entry boxes to provide insights on the different policy themes. The self-organized roundtables had few requirements, resulting in heterogeneous participation data.

Participants collaborated on a board (Mural.co) in the regional roundtables to discuss the themes. Contributions varied in word count, quality, writing style, topics covered, and textuality (ranging from textual transcripts to participant summaries). No budget was allocated for the process, so free online tools were used to collect and analyze data. Two interns manually coded the data from the regional tables, which senior Ministry officials later reviewed to identify emerging topics and tensions.

This analysis of the self-organized and regional roundtables took around four months, and progress reports were regularly presented to the advisory committee. An interdisciplinary team of MinCTCI professionals with backgrounds in engineering, law, psychology, and social sciences conducted the data analysis. None of them were experts in computer science.

With inputs from this participatory process, a draft was released in December 2021 for two months of public consultation. Citizens and organizations could evaluate each sub-theme of the draft using a 5-point Likert scale. More than 200 institutions (e.g., NGOs, corporations) and 6,500 citizens evaluated the policy draft, which received more than 80% support. The advisory committee reviewed the final policy document before publication, which included an action plan. Finally, President Piñera presented Chile's National AI policy in October 2021.

4. Methods

We utilized NLP to analyze the citizen deliberation data from the self-organized and regional roundtables. In particular, we used LDA, which is an unsupervised, non-parametric, and generative method that analyzes words as generated in probabilistic samples, providing an estimated probability of a topic being contained in a document, and the probability of a word being representative of that topic [39, 7].

Two datasets contained the citizen participation data. The first dataset contained the responses of the 69 self-organized roundtables. The second dataset contained an equal number of responses for the regional roundtables. We used two spreadsheets to consolidate the information of the

roundtables, which were organized in columns identifying the anonymized ID of the participant, type of stakeholder (i.e., academia, industry, civil society), theme(s) of focus, sub-theme(s) of focus, and textual data. The spreadsheet containing the information on the self-organized roundtables had 597 rows; the regional roundtables had 2,092 rows.

For each dataset, we applied the following data processing procedure. First, we used the MALLET Topic Modelling package [40] and the NLTK library [41] in its Spanish version to remove stopwords. Second, we aggregated the contents of the citizen responses into sub-themes, as some of the responses contained less than ten words. As a result, the number of rows in the spreadsheet of the self-organized roundtable was reduced to 146, and the regional roundtables to 273. Third, we run LDA for each database several times, generating between 2 and 30 topics. For each run, we obtained the coherence score, which measures the degree of semantic similarity between high-scoring words in a topic [42]. Starting with the set of topics with the highest coherence score, we iteratively and manually analyzed the contents of the LDA-generated topics for meaningfulness. If the contents of the topics were too broad or overlapped, we discarded that set and moved to the next with the highest coherence score. For the self-organized roundtables, the number of topics that maximized coherence and meaningfulness was 10; for the regional roundtables, 27. We obtained a keyword list and a cluster of documents containing the stakeholders' opinions on each topic.

Two researchers inspected the keyword list and the three quotes that contributed the most to each topic to assign meaning to the topic. Each researcher independently coded the quotes using MAXQDA software and axial coding techniques to interpret the meaning of the keyword list. The researchers then compared these concepts and converged on a sentence that captured the underlying meaning of each topic. A third researcher with policymaking experience revised and adjusted the proposed sentences to fit AI policy language. Finally, one researcher manually coded the Chilean AI Policy objectives using the LDA topics as the Codebook and highlighted those objectives that did not fit any of them. A second researcher, with a policy background, revised the coding. The three authors discussed discrepancies found by the second researcher until they reached a consensus.

5. Results

We found 37 topics in the data. Self-organized roundtables, which had an unstructured methodology, resulted in 10 topics. Regional roundtables, which the MCTCI organized with a structured methodology, resulted in 37 topics. The topics are in the online appendix on GitHub: https://tinyurl.com/t3255xf7. Chile's AI policy draft contained 26 objectives in 3 themes and several sub-themes, which can be found translated in the online appendix on GitHub https://tinyurl.com/2xsknd8y. To assess the effectiveness of LDA in identifying relevant topics for policymakers, we compared topics to the contents of the AI policy draft. We analyze the match between topics, the policy for each roundtable type, and the whole dataset.

The LDA-generated topics could be matched to 20 of 26 objectives as shown in figure 1. We manually inspected the data to differentiate those objectives that were not discussed (not in the data) from those that were not found by LDA (they were in the data, but no topic was created from it). Three of the six objectives not covered by the LDA were never discussed (2.1.4, 2.1.6,
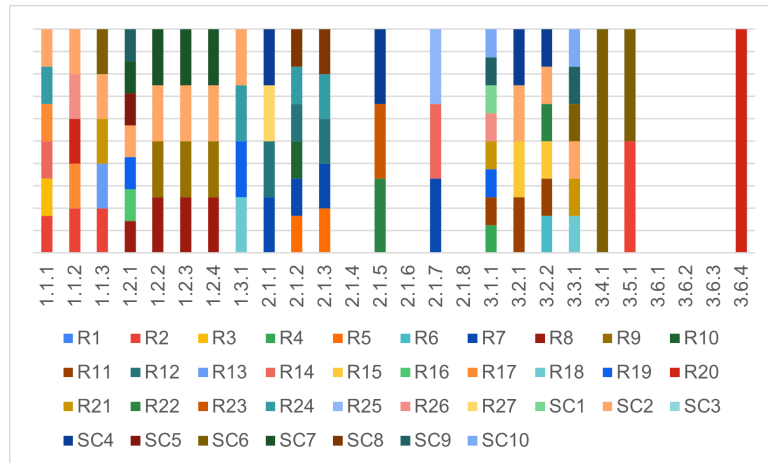
Figure 1: Objective Coverage by LDA

and 3.6.2), and three were not detected (2.1.8, 3.6.1, and 3.6.3). Only two topics detected by the LDA were not associated with any objective in the AI policy draft (SC3 and R1). We adjusted the statistics by removing the three topics not discussed (not in the data) and those not included (not incorporated by policymakers).

LDA was 87% effective in identifying the relevant topics in the participatory data, being capable of detecting 20 of the 23 policy objectives discussed in the regional and self-organized roundtables that were considered for the AI Policy draft. The average topic per objective is 3.826, suggesting that topics could be complementary, each contributing nuances to the objective.

The average number of objectives per topic was 2.588, and the median was 2. The latter made sense since many objectives were related, and they could be informed from both SC and R topics. The latter suggests that although the topics were not specific enough to be matched to only one objective, they could be specific to inform policymakers at an objective level. The only exception was SC2, which informed 11 objectives, suggesting it was too broad. Figure 2 shows the distribution of topics over the objectives.

LDA was robust to structured and unstructured participation formats. The self-organized roundtables were unstructured, and LDA topics informed 78% (18/23) of the objectives in the AI policy draft. The discussion in the regional roundtables was structured, and LDA informed 83% (19/23) of the objectives in the AI policy draft. Structured participation formats, however, yielded more topics as participants were required to provide their opinions on different matters, resulting in multiple sub-themes that were more specific than those of unstructured roundtables. This can be seen in the average topics per objective, which is higher for structured, suggesting more specific topics. The opposite happens with the ratio objective/topic, in which the average is higher, suggesting broader and more generic topics capable of covering multiple objectives. Table 1 summarizes the comparison between approaches.

LDA could reduce policymakers' time reviewing the comments to draft objectives and short papers. Instead of manually inspecting 419 comments, we only required the manual inspection of 26.5% of the documents (i.e., 111 documents, three documents per 37 topics) to assign meaning
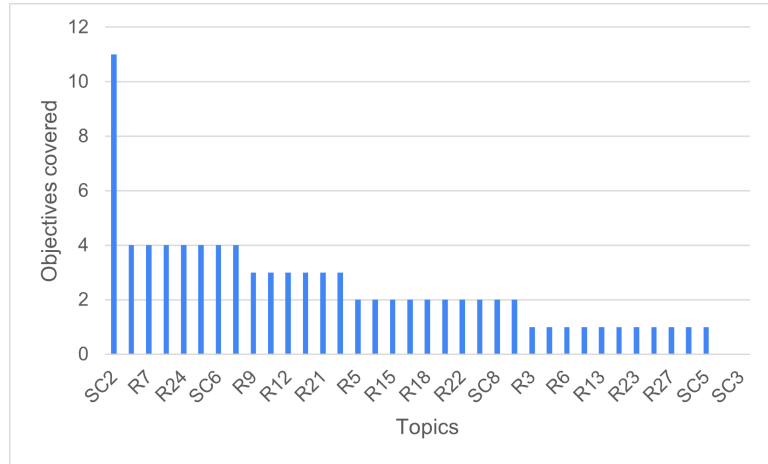
Figure 2: Topics over objectives distribution

Table 1
Comparison of structured and unstructured approach

Roundtable	Coverage	Topics/Objectives	Objectives/Topic
Unstructured (SC)	78%	1.391	3.556
Structured (R)	83%	2.435	2.154
Both	87%	3.826	2.514

to the LDA-generated keyword list, which informed most of the objectives in the AI policy draft.

6. Discussion

Using LDA in a low-resource setting can be an effective tool for policymakers to analyze participation data. Despite their current limitations [9], our findings suggest that LDA can effectively reduce the need to manually inspect deliberation transcripts to identify the most relevant policy topics. This finding is consistent with previous work [7, 38]. Furthermore, the algorithm was robust regarding structured and unstructured participation formats, which could enable policymakers to use it without significantly changing methods or requiring extra resources to prepare the data. However, LDA performed better on the structured roundtable, so our results suggest that participatory methodologies impact performance. Moreover, by combining both approaches, the best results were obtained, suggesting interesting venues for future research on designing participatory processes to improve automatic analysis performance.

Using LDA to complement expert analysis can reduce the cognitive load of systematizing and clustering the data, freeing up time to respond more timely to stakeholders, thereby increasing the legitimacy of the policymaking process. Without tools for systematizing information, policymakers rely on heuristics that can reflect personal biases, affecting the legitimacy of the process [4, 33]. Releasing time and cognitive capacities in policymakers can enable them to work on strategies to manage stakeholders depending on their saliency during the processes

[12, 26, 27]. Thus, enhanced capabilities in limited-resource settings enable policymakers to strategically manage multiple stakeholders to co-create value in technology policy's complex ecosystem [28].

Using LDA in multi-stakeholder deliberation is not free of limitations. Literature has raised concerns and challenges about using AI systems in the public sector [8]. Policymakers need to be aware of the limitations of the algorithms, and adequate guardrails need to be in place to reduce the risks of harming citizens. Using topic modeling is a helpful tool to complement policymakers' work but not a replacement for their political agency and accountability.

LDA does not replace policymakers' task of analyzing stakeholders' contributions. The algorithm exhibits a substantial level of generality, rendering it insufficient to delineate precise policy directives. While it provides a comprehensive overview of stakeholders' discourse, supporting their management, it fails to yield specific guidelines for the practical implementation, utilization, advancement, or integration of AI, which is consistent with previous research [9]. The use of other algorithmic approaches can partially remediate these issues [e.g., 5, 6, 37], but they have other limitations such as computational requirement, coding complexity, and data quality. Choosing the right tool will require policymakers to asses restrictions in time, technological infrastructure, knowledge, and data quality, among others. Still, our findings suggest that LDA can complement the policymaker revision tasks even with data from messy processes that are not prepared for automated analysis. Consequently, using LDA to enhance the revision of stakeholders' contributions enhances policymakers' capacity to make strategic and well-informed decisions for stakeholder management.

7. Conclusions and Future Work

Policymakers face several challenges in strategically managing multiple stakeholders when designing technology policy. One approach is opening deliberation processes for stakeholders to discuss visions and concerns. However, failing to respond promptly and adequately can erode legitimacy and trust. A particular challenge policymakers face is being able to analyze structured and unstructured data simultaneously. Our study reveals that topic modeling can effectively and efficiently support policymakers by systematizing and giving general insights about the issues stakeholders raise.

We build upon work on using NLP for public participation processes, showing that even simple applications can improve policymakers' work and that the gap between theoretical research and practical application is not that wide for complementary tasks. Providing policymakers with tools to strategically design stakeholders' journeys in limited resource contexts is relevant to improving policies and citizens' well-being. Our findings show that current LDA algorithms can help policymakers in systematization and more visual organization, which results in less time wasted analyzing massive data. LDA is not a replacement for human analysis, but it reduces cognitive overload, freeing the capacity to think strategically about stakeholders.

However, using NLP for public participation processes and stakeholder management still has many challenges. More research is needed to understand the dynamics within government organizations using AI to manage multi-stakeholder processes. As part of this ongoing research, we will study how policymakers evaluate topic modeling and its value to manage stakeholders.

To do so, we will conduct semi-structured interviews with the policymakers involved in the Chilean AI Policy process to distill their assessment of the potential value of using NLP in such contexts. Moreover, many questions remain: How do different internal stakeholders (e.g., politicians, policymakers, technical staff) interact with AI systems? What explanations are required for each group of stakeholders about the system? How are AI systems designed and acquired? How do external stakeholders design strategies to take advantage of algorithms? How can large language models with their own challenges and opportunities be used in participatory processes?

We want to highlight some limitations in our research. First, it is context-specific regarding a developing country (Chile), which might not be generalizable to other contexts. Second, we used LDA because it is simple to implement, and previous literature has shown that it yields good results in similar contexts; however, using other techniques or even combining more than one could deliver better stakeholder journeys in the public sector. Finally, to find more generalizable insights, longitudinal studies along many participation processes can help understand how these tools might be used in different contexts.

Acknowledgments

References

[1] T. Aitamurto, K. Chen, A. Cherif, J. S. Galli, L. Santana, Civic CrowdAnalytics: making sense of crowdsourced civic input with big data tools, in: Proceedings of the 20th International Academic Mindtrek Conference, AcademicMindtrek '16, Association for Computing Machinery, New York, NY, USA, 2016, pp. 86–94. doi:10.1145/2994310.2994366.

[2] S. Chun, S. Shulman, R. Sandoval Almazan, E. Hovy, Government 2.0: Making Connections Between Citizens, Data and Government, Information Polity 15 (2010) 1–9. doi:10.3233/IP-2010-0205.

[3] M. Janssen, Y. Charalabidis, A. Zuiderwijk, Benefits, Adoption Barriers and Myths of Open Data and Open Government, Information Systems Management 29 (2012) 258–268. doi:10.1080/10580530.2012.716740.

[4] K. Chen, T. Aitamurto, Barriers for Crowd's Impact in Crowdsourced Policymaking: Civic Data Overload and Filter Hierarchy, International Public Management Journal 22 (2019) 99–126. doi:10.1080/10967494.2018.1488780.

[5] J. Romberg, T. Escher, Automated Topic Categorisation of Citizens' Contributions: Reducing Manual Labelling Efforts Through Active Learning, in: M. Janssen, C. Csáki, I. Lindgren, E. Loukis, U. Melin, G. Viale Pereira, M. P. Rodríguez Bolívar, E. Tambouris (Eds.), Electronic Government, Lecture Notes in Computer Science, Springer International Publishing, Cham, 2022, pp. 369–385. doi:10.1007/978-3-031-15086-9_24.

[6] M. Arana-Catania, F.-A. V. Lier, R. Procter, N. Tkachenko, Y. He, A. Zubiaga, M. Liakata, Citizen Participation and Machine Learning for a Better Democracy, Digital Government: Research and Practice 2 (2021) 1–22. doi:10.1145/3452118.

[7] L. Hagen, Content analysis of e-petitions with topic modeling: How to train and evaluate LDA models?, Information Processing & Management 54 (2018) 1292–1307. doi:10.1016/j.ipm.2018.05.006.

[8] L. Tangi, C. van Noordt, A. P. Rodriguez Müller, The challenges of AI implementation in the public sector. An in-depth case studies analysis, in: Proceedings of the 24th Annual International Conference on Digital Government Research, DGO '23, Association for Computing Machinery, New York, NY, USA, 2023, pp. 414–422. doi:10.1145/3598469.3598516.

[9] J. Romberg, T. Escher, Making Sense of Citizens' Input through Artificial Intelligence: A Review of Methods for Computational Text Analysis to Support the Evaluation of Contributions in Public Participation, Digital Government: Research and Practice (2023). doi:10.1145/3603254, just Accepted.

[10] M.-H. Weng, S. Wu, M. Dyer, AI Augmented Approach to Identify Shared Ideas from Large Format Public Consultation, Sustainability 13 (2021) 9310. URL: https://www.mdpi.com/2071-1050/13/16/9310. doi:10.3390/su13169310, number: 16 Publisher: Multidisciplinary Digital Publishing Institute.

[11] L. S. Flak, J. Rose, Stakeholder Governance: Adapting Stakeholder Theory to E-Government, Communications of the Association for Information Systems 16 (2005). doi:10.17705/1CAIS.01631.

[12] J. Rose, L. S. Flak, �Sæbø, Stakeholder theory for the E-government context: Framing a value-oriented normative core, Government Information Quarterly 35 (2018) 362–374. doi:10.1016/j.giq.2018.06.005.

[13] M. Yildiz, E-government research: Reviewing the literature, limitations, and ways forward, Government Information Quarterly 24 (2007) 646–665. doi:10.1016/j.giq.2007.01.002.

[14] Z. Irani, T. Elliman, P. Jackson, Electronic transformation of government in the U.K.: a research agenda, European Journal of Information Systems 16 (2007) 327–335. doi:10.1057/palgrave.ejis.3000698, publisher: Taylor & Francis _eprint: https://-doi.org/10.1057/palgrave.ejis.3000698.

[15] J. Rowley, e-Government stakeholders—Who are they and what do they want?, International Journal of Information Management 31 (2011) 53–62. doi:10.1016/j.ijinfomgt.2010.05.005.

[16] J. Holgersson, F. Karlsson, Public e-service development: Understanding citizens' conditions for participation, Government Information Quarterly 31 (2014) 396–410. doi:10.1016/j.giq.2014.02.006.

[17] T. Aitamurto, J. Saldivar, Examining the Quality of Crowdsourced Deliberation: Respect, Reciprocity and Lack of Common-Good Orientation, in: Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems, CHI EA '17, Association for Computing Machinery, New York, NY, USA, 2017, pp. 2314–2321. doi:10.1145/3027063.3053248.

[18] V. A. Schmidt, Democracy and Legitimacy in the European Union Revisited: Input, Output and 'Throughput', Political Studies 61 (2013) 2–22. doi:10.1111/j.1467-9248.2012.

`00962.x`, publisher: SAGE Publications Ltd.

[19] K. Koc-Michalska, D. Lilleker, Digital Politics: Mobilization, Engagement, and Participation, Political Communication 34 (2017) 1–5. doi:`10.1080/10584609.2016.1243178`.

[20] J. Macnamara, Organizational listening: Addressing a major gap in public relations theory and practice, Journal of Public Relations Research 28 (2016) 146–169. doi:`10.1080/1062726X.2016.1228064`.

[21] A. Fung, Putting the Public Back into Governance: The Challenges of Citizen Participation and Its Future, Public Administration Review 75 (2015) 513–522. doi:`10.1111/puar.12361`.

[22] I. Mergel, R. K. Rethemeyer, K. Isett, Big Data in Public Affairs, Public Administration Review 76 (2016) 928–937. doi:`10.1111/puar.12625`.

[23] M. Sloane, E. Moss, O. Awomolo, L. Forlano, Participation is not a Design Fix for Machine Learning, 2020. doi:`10.48550/arXiv.2007.02423`, arXiv:2007.02423 [cs].

[24] F. Delgado, S. Yang, M. Madaio, Q. Yang, The Participatory Turn in AI Design: Theoretical Foundations and the Current State of Practice, in: Proceedings of the 3rd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization, EAAMO '23, Association for Computing Machinery, New York, NY, USA, 2023, pp. 1–23. doi:`10.1145/3617694.3623261`.

[25] J. Pierre, R. Crooks, M. Currie, B. Paris, I. Pasquetto, Getting Ourselves Together: Data-centered participatory design research & epistemic burden, in: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, CHI '21, Association for Computing Machinery, New York, NY, USA, 2021, pp. 1–11. doi:`10.1145/3411764.3445103`.

[26] �Sæbø, L. S. Flak, M. K. Sein, Understanding the dynamics in e-Participation initiatives: Looking through the genre and stakeholder lenses, Government Information Quarterly 28 (2011) 416–425. doi:`10.1016/j.giq.2010.10.005`.

[27] M. R. Johannessen, �Sæbø, L. S. Flak, Social media as public sphere: a stakeholder perspective, Transforming Government: People, Process and Policy 10 (2016) 212–238. doi:`10.1108/TG-01-2015-0003`, publisher: Emerald Group Publishing Limited.

[28] O. R. Ashaye, Z. Irani, The role of stakeholders in the effective use of e-government resources in public services, International Journal of Information Management 49 (2019) 253–270. doi:`10.1016/j.ijinfomgt.2019.05.016`.

[29] M. A. Pirog, Data Will Drive Innovation in Public Policy and Management Research in the Next Decade, Journal of Policy Analysis and Management 33 (2014) 537–543. URL: https://www.jstor.org/stable/24033344, publisher: Wiley.

[30] J. Font, S. P. d. Amo, G. Smith, Tracing the Impact of Proposals from Participatory Processes: Methodological Challenges and Substantive Lessons, Journal of Deliberative Democracy 12 (2016). doi:`10.16997/jdd.243`, number: 1 Publisher: University of Westminster Press.

[31] A. Macintosh, Characterizing e-participation in policy-making, in: 37th Annual Hawaii International Conference on System Sciences, 2004. Proceedings of the, 2004, pp. 10 pp.–. doi:`10.1109/HICSS.2004.1265300`.

[32] K. Yang, K. Callahan, Assessing Citizen Involvement Efforts by Local Governments, Public Performance & Management Review 29 (2005) 191–216. doi:`10.1080/15309576.2005.11051865`.

[33] O. Perez, Complexity, Information Overload, and Online Deliberation Online Consultation and Democratic Communication, I/S: A Journal of Law and Policy for the Information

Society 5 (2008) 43–86. URL: https://heinonline.org/HOL/P?h=hein.journals/isjlpsoc5&i=53.

[34] P. G. Roetzel, Information overload in the information age: a review of the literature from business administration, business psychology, and related disciplines with a bibliometric approach and framework development, Business Research 12 (2019) 479–522. doi:10.1007/s40685-018-0069-z.

[35] I. Yovanovic, I. Goñi, C. Miranda, Remote Usability Assessment of Topic Visualization Interfaces with Public Participation Data: A Case Study, JeDEM - eJournal of eDemocracy and Open Government 13 (2021) 101–126. doi:10.29379/jedem.v13i1.640, number: 1.

[36] T. Aitamurto, H. Landemore, Crowdsourced Deliberation: The Case of the Law on Off-Road Traffic in Finland, Policy & Internet 8 (2016) 174–196. doi:10.1002/poi3.115.

[37] A. Simonofski, J. Fink, C. Burnay, Supporting policy-making with social media and e-participation platforms data: A policy analytics framework, Government Information Quarterly 38 (2021) 101590. doi:10.1016/j.giq.2021.101590.

[38] L. Hagen, �Uzuner, C. Kotfila, T. M. Harrison, D. Lamanna, Understanding Citizens' Direct Policy Suggestions to the Federal Government: A Natural Language Processing and Topic Modeling Approach, in: 2015 48th Hawaii International Conference on System Sciences, 2015, pp. 2134–2143. doi:10.1109/HICSS.2015.257, iSSN: 1530-1605.

[39] D. M. Blei, A. Y. Ng, M. I. Jordan, Latent Dirichlet Allocation, Journal of Machine Learning Research 3 (2003) 993–1022.

[40] A. McCallum, Mallet. A machine learning for language toolkit. (2002). URL: http://mallet.cs.umass.edu.

[41] S. Bird, E. Klein, E. Loper, Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit, "O'Reilly Media, Inc.", 2009. Google-Books-ID: KGIbfiiP1i4C.

[42] K. Stevens, P. Kegelmeyer, D. Andrzejewski, D. Buttler, Exploring Topic Coherence over Many Models and Many Topics, in: Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, Association for Computational Linguistics, Jeju Island, Korea, 2012, pp. 952–961. URL: https://aclanthology.org/D12-1087.