Understanding “Democratization” in NLP Research

Anonymous ACL submission

Abstract

Despite increasing discussion of the “democratization” of natural language processing and machine learning research, the use of this term and its connections to democracy have not been thoroughly studied. Given the rich history of democracy, understanding what AI researchers mean by “democratization” is important for ensuring that we are accurately representing public participation in and control of the field. Thus, we conduct a large-scale, mixed-methods analysis of every use of democracy-related terms among all papers published in the ACL Anthology or at ICLR, ICML, or NeurIPS ($N = 507$ papers); we do this to uncover the themes, values, and concepts that researchers associate with democracy. In addition, we examine how deeply papers that mention democracy engage with the concept via their text and citations. Ultimately, we find that “democratization” mostly signals broadening access or use of technologies, especially without expertise. In contrast, researchers’ conceptualizations of democracy are diverse and grounded in theories of deliberation and debate. Moreover, we observe that papers that mention democracy often do not meaningfully treat democracy or draw on democratic theories from outside NLP. Based on our findings, we urge responsible use of the term “democratization” and greater engagement with theories of democracy towards enriching our discussions of AI access and governance.

1 Introduction

As the influence of language technologies grows around the world, including outside academia, it has become increasingly popular to discuss “democratization” in natural language processing (NLP) and machine learning (ML) research (Seger et al., 2023; Zaremba et al., 2023; Ganguli et al., 2023). Indeed, the number of papers mentioning democracy has seen a rapid increase as NLP technologies have become more powerful (see Figure 1). Responsible use of the term is critical for accurately representing progress in NLP and ML with respect to capturing democratic values and distributing power. However, the treatment of democracy in artificial intelligence (AI) literature, and in particular the term “democratization,” have not been carefully investigated thus far. Therefore, our paper asks the following questions: What does “democratization” in NLP actually mean and how is it connected to “democracy”? Moreover, when people use the word “democratization,” how do they operationalize it?

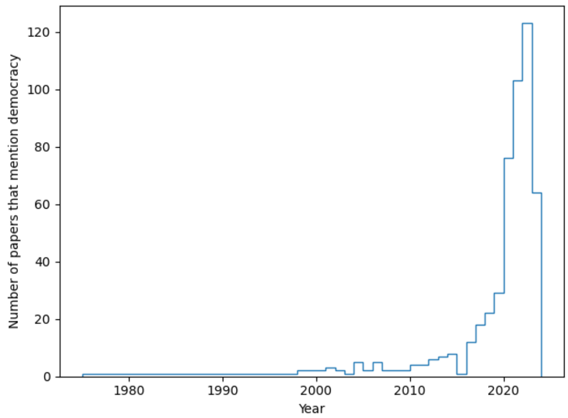

Figure 1: Number of papers mentioning democracy by year among all papers published in the ACL Anthology or in ICLR, ICML, or NeurIPS.

To answer these questions, we conduct a large-scale, mixed-methods analysis (§4) of every use of “democratization,” “democracy” and related words among all papers published in the ACL Anthology or at ICLR, ICML or NeurIPS (prior to November 24, 2023). Specifically, we uncover the themes, values and concepts that authors associate with these words. We find that the use of “democratization” mostly signals broadening access or use of something, especially without expertise, whereas literature discussing democracy in other contexts is grounded in theories of deliberation and debate.

Next, we examine the depth of engagement of papers that mention democracy[1] both in their own text and via their citations (§5). The vast majority of papers invoke democracy only once, outside the main paper sections, and engage minimally with extra-disciplinary literature.

Finally, we dig deeper into the differences between how "democratization" and democracy are discussed, finding that while authors generally associate "democratization" with various positive values related to access and reducing costs, they almost never explicitly operationalize "democratization."

Without clearly indicating our meanings, goals and plans for "democratization," and in particular the connections (or lack thereof) to democracy, we risk misrepresenting public control of the field. We thus urge more deliberate use of the word "democratization" and encourage NLP and ML researchers to improve their citational praxis and enrich their work by drawing on the over 3000 years of scholarship on democracy and democratization from the social sciences.

2 Background

Democratization has had a long history of study and consideration starting from 1100 BCE in ancient Phoenicia (Glassman, 2017). More recently, research has considered the links between technology and democracy (Mumford, 1964). In brief, this area of work has argued that technology can either afford agency, access, and distribute power, i.e., be democratic, or consolidate power within a small set of actors, i.e., be authoritarian. More recently, in relation to discriminatory ML, Kalluri (2020) argued that search for fair ML can serve as a distraction to considering how ML distributes power. Here, we consider select theories of democracy to serve as a basis for which we consider how NLP research has understood and operationalized democratization.

These democratic theories can enrich the democratization of NLP and ML by making democratic discussions representative and efficient, diversifying forums for democratic dialogues, and dismantling barriers to participate in democratic processes.

Deliberative democracies Deliberation and inclusion in the democratic process are often highlighted as goals for democratic societies and technologies. Indeed, as we find from our analysis (see §4), democratic deliberation often appears in our surveyed papers.

Deliberative democracy is a form of democracy that emphasizes a process where participants can debate a particular object (e.g., policy or technology, in the case of NLP) on its merits and collectively come to a decision about its implementation or integration (Goodin, 2000). Deliberative democracy thus provides an avenue for research to engage wider publics in conversation about research artifacts and their application, thereby obtaining more legitimacy of the outcome of the deliberation (Rosenberg, 2007). While some objects may be relevant to an entire population, other objects only require smaller groups. For example, policy on national healthcare or the use of NLP tools in judicial systems may pertain to entire national populations, compared to policy changes within a municipality. Thus, for a legitimate decision, competent and relevant publics must be considered, otherwise the outcome of the deliberate democratic process may be a rejection of the decisions (Parkinson, 2003).

Democratic spheres Considering the goals and mechanisms for technologies as well as arenas for successful democratic dialogues is essential towards achieving goals of democratization. While in some instances, a singular democratic arena, or sphere, may suffice, e.g., in a small-scale direct democratic process, larger and more complex structures such as societies require a greater number of democratic spheres through which different publics can engage (Fraser, 1990).

In her work "Rethinking the Public Sphere" (Fraser, 1990), Fraser discusses the idea of the public sphere as described by Habermas and Burger (1991). While Habermas and Burger argue for the existence of a single public sphere, Fraser argues that a functional democracy that seeks to be inclusive of its population must seek a *plurality* of public spheres. Drawing on Spivak (1988), Fraser posits that a single public sphere relegates many communities to the margins of the public sphere and gives weight to the loudest and majoritarian voices. In contrast, one can imagine a plurality of public spheres, which seek to represent smaller communities. Fraser argues that similar tendencies for the loudest voices to be heard also exist in such a plural-democracy, however, by virtue of multiple public spheres in which one can find representation and participate in, a plurality of public spheres

[1]By "mention democracy," we mean the usage of democracy-related terms, including "democratization."

minimizes the risk of marginalization and increases the space for otherwise excluded and marginalized communities to participate.

Radical egalitarian democracies Understanding and combating barriers to public participation in democratic processes, as well as uneven distributions of power, are paramount for operationalizing the democratization of NLP. Wright (2010) posits that in late-stage capitalist societies, people often feel limited efficacy in participating in democratic processes, and many decisions feel insufficiently democratic because they are dominated by elites and tainted by private property. Thus, he argues that a radical democracy must shield political processes by instituting strong mechanisms against translating private economic power into political power. Situating political justice in the NLP and ML landscapes, the development of language technologies, and indeed the operation of democratic processes for these technologies, are heavily controlled by the interests of private companies (Zaremba et al., 2023; Ganguli et al., 2023; Talat et al., 2022).

In addition, Wright (2010) argues that democratic egalitarianism requires that all humans must have equal access (not just equal opportunity) to participate in democratic processes, and in turn, these processes should institute programs that dismantle systems of oppression. To ensure equal access, it is necessary to identify where suffering and inequality exist, and diagnose its roots in mechanisms of oppression. Thus, the democratization of NLP must attend to and mitigate social conditions that prevent equal access.

3 Data

To investigate the use of terms related to "democracy" and "democratization" in NLP, we perform a large-scale mixed-methods analysis of *all* 507 papers (prior to November 24, 2023) that mention these terms in the ACL Anthology and three major ML conferences (ICML, ICLR, NeurIPS).

All excerpts First, we collect the metadata and text from open-access PDFs of all these papers using the Semantic Scholar API (Kinney et al., 2023). We then use the punkt NLTK sentence tokenizer (Bird and Loper, 2004) to decompose the full text of the paper (i.e., the title, abstract, and body) into sentences. We collect all the sentences that contain the substring "democra" (excluding "democrats") for a total of 3411 excerpts across 1537 papers.

Filtering for relevant excerpts In order to get at the specific excerpts that reveal authors' conceptualizations of "democratization" and "democracy," we exclude unrelated "democra" mentions, such as those that are part of named entities (e.g., "Center for Media and Democracy"), hypothetical examples (e.g., of textual entailment), modelling examples (e.g., word2vec clusters, LDA topics), or example data (e.g., a tweet for sentiment classification). We additionally exclude mentions that are primarily in a language besides English, and references. To do this filtering, we apply a two-stage procedure: automatic filtering and manual annotation for relevance.

In particular, we first leverage a curated list to filter out uses of "democra" words that are either part of named entities (e.g., "Syrian Democratic Forces," "Croatian Democratic Union," "ANR Democrat"), or terms that always appear in examples in papers (e.g., example tweets containing "#democracy"). Our full list of exclusion terms is shown in Appendix A, and the excluded excerpts from this stage of filtering were verified by one author.

Then, we manually annotate the remaining 2273 excerpts, focusing on finding instances where the authors deliberately use words containing "democra" substrings as part of their argument or evidence, including citations. If it is unclear whether the isolated excerpt is relevant or irrelevant, we look up the sentence in the original PDF and examine it in context to make a decision. Our two-stage filtering leaves us with 923 sentences from 507 different papers, which we subsequently analyze.

4 Conceptualizations of Democracy

In order to understand how democracy is conceptualized in NLP papers, we perform a large-scale, mixed-methods analysis of the 923 filtered excerpts to surface the overarching themes discussed in the literature, as well as the values and concepts that authors associate with democracy.

4.1 Methods

Two authors annotated the first 300 excerpts independently for themes, concepts and values, as explained in detail below. We then discussed our annotations and attempted to resolve inconsistencies in themes and normalize concept names, before annotating the remaining excerpts independently. For each paper, the themes, concepts and values from all of its excerpts were grouped together into single sets, i.e., a union operation was performed.

Themes We first inductively analyze the excerpts to identify salient, overarching themes that characterize how democracy is discussed in the papers (Saldana, 2021). Four major categories emerged after a first pass over all the excerpts:

- *necessary/beneficial*: things that are necessary for or beneficial to democracy (e.g., discourse, majority, voting)
- *danger*: dangers to democracy (e.g., misinformation)
- *democratization*: use of the words "democratize" or "democratization" (e.g., of ML)
- *math*: mathematical or ML ways to operationalize democracy (e.g., democratic networks, democratic matrices, mathematical models of democracy)

Two authors then systematically annotated every excerpt with an explicit and, if applicable, an implicit theme. An explicit theme was assigned to excerpts that explicitly state, e.g., that something is necessary for or a danger to democracy, something is being democratized, etc.; otherwise, it is classified as *other*. In contrast, the implicit theme requires annotators to make inferences about how authors think about democracy.

For example, the excerpt: "The most democratic option is to give each tagger one vote (Majority)," was assigned an explicit theme of *math* by both annotators, as it discusses a way to operationalize NLP taggers in a "democratic" way. Both annotators also inferred that the authors believe majority voting to be necessary for democracy, hence *necessary/beneficial* was assigned as an implicit theme.

Values and concepts In a final pass over the data, the authors also annotated each excerpt for values (e.g., "consensus," "equality") and more broadly, concepts (e.g., "misinformation," "elections") associated with democracy, with the goal of further exploring how authors conceptualize democracy; values are a subset of concepts.

4.2 Results

Of our four themes, *democratization* is by far the most frequent one with 213 papers, followed by 67 for *necessary/beneficial*, 59 for *danger*, and 35 for *math*. We find a total of 110 concepts and 77 values associated with democracy, with each paper containing on average 1.162 themes and 1.036 concepts. Annotation was highly consistent, with annotators only differing, on average, on: 0.0374 explicit themes, 0.0178 implicit themes, and 0.787 concepts, per paper. Given the minimal disagreement between annotators, we henceforth do not distinguish between explicit and implicit themes.

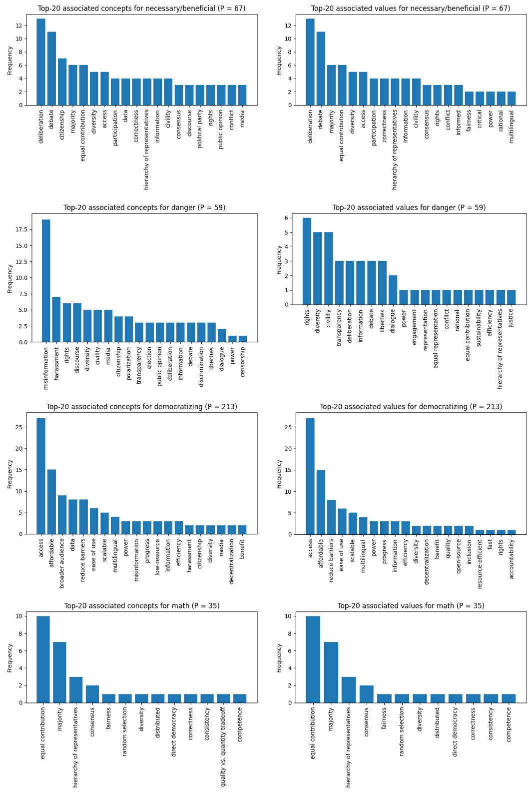

Figure 2: Frequency of concepts (left) and values (right) associated with democracy in papers, stratified by paper themes. For each theme, P refers to the number of papers annotated as having that type of theme.

Values associated with democracy in NLP Full lists of values and concepts associated with democracy are shown in Appendix B, and we focus here on the most frequent ones that we found during our qualitative analysis (see Figure 2). Notably, we found that some values contradict each other. For instance, treating "random selection" as democratic is incompatible with choosing by "consensus" which in turn is incompatible with "majority" decision-making. Yet researchers operationalize AI systems in all of these different ways and call them "democratic."

As Figure 2 shows, there are also big differences in the values and concepts associated with democracy when split by paper themes. “Democratization” is overwhelmingly associated with access, affordability and reducing barriers, while *math* papers concern themselves with values for decision-making (typically with multiple input features or models), e.g., equal contribution, majority, consensus, etc. Papers that discuss what democracy needs or is endangered by have more overlap in the values they associate with democracy, including deliberation, debate, and diversity. For the most part, the top values for our four major themes are also the top concepts, except for *danger* papers, which focus on threats to democracy, e.g., misinformation, harassment (Coeckelbergh, 2024).

In sum, these vastly different thematic clusters of how AI researchers tend to talk about democracy show that they associate it with different and sometimes even conflicting values. Next, we examine the depth of their engagement with ideas and prior literature about democracy to understand how these may inform the different conceptualizations of democracy observed in this section.

5 Engagement with Democratic Theories

One of our objectives is to quantify the extent to which papers that talk about “democracy” engage with it deeply, and reference theories of democracy and the literature outlined in Section 2. This section presents our mention and citation graph analysis to answer this question.

5.1 Methods

As a measure of the depth of engagement with democracy, we count how often democracy is mentioned per paper, as well as which sections of papers these mentions appear in. We extract section names with the Semantic Scholar API and apply basic cleaning to normalize them across papers (e.g., singularization such as “related works” → “related work,” merging similar sections like “conclusion” and “conclusion and future work”).

To analyze engagement with theories of democracy, we study the references they cite: the fields they belong to, the venues they were published in, the location and numbers of citations, and the *intent* of the citation, i.e., whether the citation is used to provide background, inform the methodology of the paper, or is related to the results. We obtain field, venue and intent metadata using the Semantic Scholar API, and we classify references as intra-disciplinary if they are from Computer Science, Mathematics, or Linguistics.

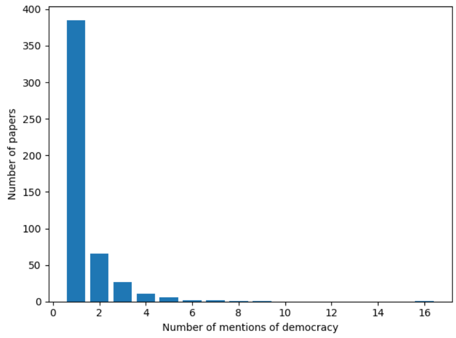

Figure 3: Frequency of numbers of mentions of democracy per paper.

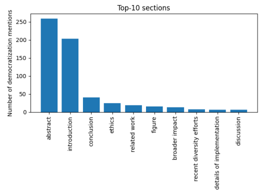

Figure 4: Frequency of paper sections in which mentions of democracy occur.

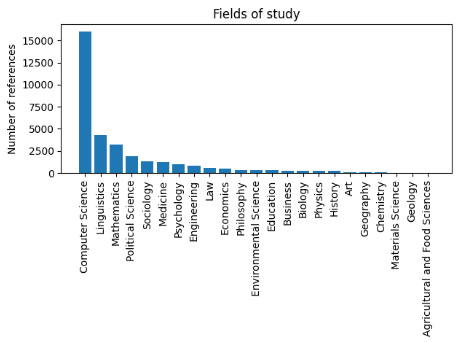

Figure 5: Frequency of fields of study of references cited by papers that mention democracy.

5.2 Results

Where and how often is democracy invoked in papers? Figure 3 shows that the vast majority of papers that do mention democracy only mention it once, suggesting superficial engagement with the concept. This is further substantiated by Figure 4, which reveals that most mentions (84.8%) occur in the abstract, introduction, and conclusion sections of papers.

What kind of papers are cited and why? Figure 5 shows that papers overwhelmingly cite work from Computer Science. The next biggest category is Linguistics, cited three times less often, followed by Mathematics, and finally Political Science. Sim-

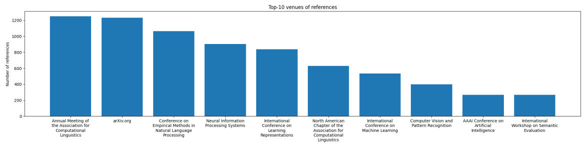

Figure 6: Frequency of venues of references cited by papers that mention democracy.

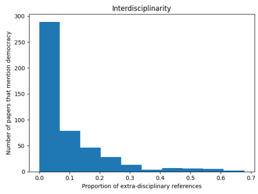

Figure 7: Frequency of proportions of extra-disciplinary references cited by papers that mention democracy.

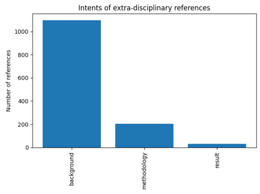

Figure 8: Frequency of intents of extra-disciplinary references cited by papers that mention democracy.

ilarly, when considering the venues of references, Figure 6 shows that the majority of references are from NLP and ML conferences, or arXiv. This suggests low levels of interdisciplinary engagement, which we find surprising for papers that invoke a term with such a rich academic history.

Indeed, as Figure 7 shows, the modal paper in our corpus cites a few or no extra-disciplinary references; 177 papers cite zero extra-disciplinary references, and 87 papers only cite one extra-disciplinary reference. After this, there is a long tail of papers that engage more extensively with literature outside NLP and ML.

Focusing on extra-disciplinary citations, we find, as expected, that most of them come from the social sciences, and in particular, political science. However, when examining citation intents in Figure 8, we find that most of these references are only for background. This means that even when papers related to democracy and democratization do engage more with extra-disciplinary scholarship, they tend not do so in their methods and results, which might indicate gaps in translating theories of democracy to our field.

6 "Democratization" in AI

Having observed that "democratization" papers comprise the largest proportion of our data and have noticeably distinct concepts and values, we focus on and further explore papers that explicitly mention "democratization." In addition to examining the differences between "democratization" papers and the other papers in our data, we ask: What is being democratized? How, and to what end?

6.1 Methods

One author annotated all excerpts with an explicit theme of "democratization" for:

- Causes (*how is something being democratized, or what is engendering its democratization?*);
- Targets (*what exactly is being democratized?*);
- Goals (*why, or to what ends, is something being democratized?*)

For example, take the following quote from an excerpt: "gaining more knowledge on AutoML and NAS could lead to improved democratisation of deep learning models to non-experts as they automate ML pipelines that previously could require immense human expertise." Here, the target of democratization is deep learning models (DL), the cause is knowledge (of AutoML and NAS), and the goal is use without expertise.

We additionally confirm the results of our excerpts-based analysis by sampling 30 papers to read fully. We use the HuggingFace `all-mpnet-base-v2` sentence transformer (Reimers and Gurevych, 2019; Wolf et al., 2020) to embed all excerpts related to democratization. Then, we apply spectral clustering to the embeddings (see Figure 11 in Appendix A) and we select 3 clusters using the spectral gap heuristic. We choose 5 papers from each of the cluster centers and boundaries, for a total of 30 papers.

6.2 Results

Figure 9 shows histograms of concepts and values associated with "democratization" compared to the associations with all other mentions of democracy. The top values and concepts for "democratization" papers are about increasing access and ease of use, and reducing costs and barriers. This is in stark contrast to non-*democratization* papers, which focus on values and concepts that are more recognizably related to both folk and theoretical notions

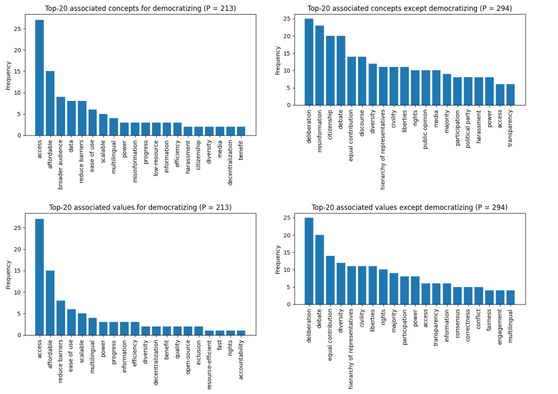

Figure 9: Frequency of concepts and values, split by *democratization* papers and all other papers. Associations with *democratization* (left) are different from associations with all other mentions of democracy (right). P refers to the number of papers with the given theme.

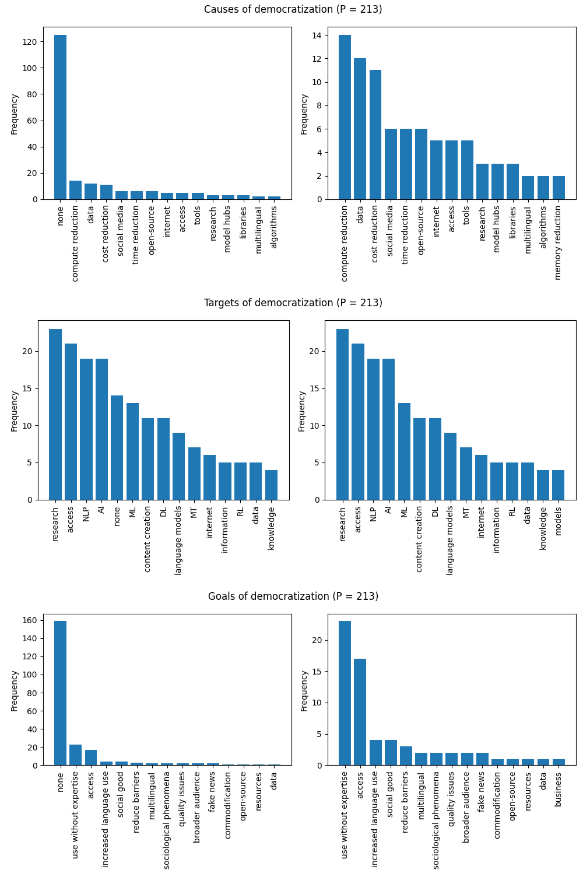

Figure 10: Frequency of causes, targets, and goals of democratization in papers. Figures on the right show frequencies with "none" removed from the x axis.

of democracy such as decision-making, deliberation, debate and diversity. Contestably, "democratization" papers share some values with radical egalitarian theories of democracy, but do not adequately distinguish between equal access and equal opportunity, or equal access to models vs. access to democratic processes. This mismatch in values and concepts shows that **NLP researchers conceive of *democratization* as something quite different from democracy.** The primary similarities appear to be that research generally view both "democratization" and "democracy" positively.

Having established that *democratization* in NLP is a distinct phenomenon more closely related to access and costs (computational, financial or otherwise), we now examine the causes, targets and goals of said *democratization* more granularly in Figure 10. 125 papers do not state the causes of democratization and 159 do not state the goals; sometimes, authors write about democratization as a separate, autonomous process that is not affected by the authors, or is minimally aided by their research contributions. Other authors write about how their research democratizes a technology without concretely expanding on how that occurs, e.g., in terms of digital infrastructure, governance structures, participatory structures, etc. When stated, popular causes for democratization are reductions in computation, time and cost; targets of democratization tend to be nebulous and big, e.g., AI, NLP, research and access; and the main goals of democratization are increasing access and use, particularly without expertise.

Fully reading the 30 sampled papers confirmed our analysis from the excerpts; none of the selected papers appear to lay out a plan for democratization, and indeed very few even comment on democratization outside of the excerpts. This strengthens the conclusions of our excerpts-based analysis.

7 Related work

Analyzing scholarly textual data Numerous prior works have extracted insights about how researchers conceptualize topics from intersectionality to power, from the text of their papers. For example, Blodgett et al. (2020) analyze 146 NLP papers to understand how their authors think about "bias." Birhane et al. (2022) annotate 100 ML papers to identify prominent values in the field. Ovalle et al. (2023) inductively and deductively surface patterns in how AI papers about "intersectional fairness" fail to engage with the critical framework of intersectionality. Wahle et al. (2023) analyze the diversity of citations in NLP with respect to their interdisciplinarity. In our work, we use similar methods to examine how NLP and ML researchers conceptualize "democracy" and "democratization," and their engagement with theories of democracy.

Conceptions of democracy in AI Seger et al. (2023) discuss the multiplicity of AI "democratization," positing that differing uses of the term causes people to not recognize the possibly shared "goals, methodologies, risks, and benefits" of their conceptions. They draw from news articles and talks to identify four conceptions of "democratization:" use, development, benefits, and governance. Based on a qualitative survey of 35 articles, Rubeis et al. (2022) study how "democratization" is used in relation to AI and its connection to democracy in the context of medicine and healthcare. They uncover diverse conceptions of democratization, from democratizing access to data to enabling people to govern AI. Unlike both of these papers, we conduct a large-scale, mixed-methods analysis of the text of NLP and ML papers.

Ahmed et al. (2020) identify criteria for "democratizing" the use of AI, e.g., affordability, accessibility, fairness, and Ahmed and Wahed (2020) empirically analyze the "democratization" of AI development, showing that a divide in compute access between tech companies and non-elite universities correlates with a divergence in AI research output. However, these works do not elucidate possible connections between "democratization" and "democracy." Nonetheless, their perspectives support our findings that researchers center model access (e.g., use, development) in their conceptualizations of "democratization."

Yet other works focus on AI governance and increasing public control of AI development and deployment. For example, Gilman (2023) posits that public participation is critical for democratizing AI, calling on institutions to include participation in all stages of AI development and budget for it. Siddarth (2023) describes a case study of "democratic" AI where a group of human representatives train a large model to align with a constitution based on their values. Collective Intelligence Project (2024) presents a roadmap to achieve "democratic" AI, including connecting open source and democracy communities and increasing the geographic diversity of public input processes. Mun et al. (2024) propose a "democratic" framework to gather AI uses, harms, and benefits from the public to guide the evaluation and regulation of AI.

8 Discussion and Conclusion

Our in-depth mixed-methods and citation graph analyses show that we have a long way to go when it comes to using "democracy" in our work as NLP and ML researchers. We find low levels of interdisciplinary engagement, infrequent operationalization of what "democratization" actually entails, and vastly different ways of viewing what "democracy" means. In Appendix B, we present additional results analyzing the authors, institutional affiliations, and funding bodies acknowledged in the papers in our data, as well as the sources of extra-disciplinary references. These additional analyses further characterize the politics of how NLP and ML reseachers treat "democratization" and "democracy."

Overall, our results show that when invoking democracy, NLP and ML researchers need to engage further with the centuries of rich literature from philosophy and the social sciences that discuss it. In addition, it is important for researchers who use the term "democratization" to describe precisely what they mean by it and their plan to operationalize it, especially detailing any connections, or lack thereof, to democracy. Without this, we risk misrepresenting public control of the field.

Indeed, some efforts by AI researchers, e.g., OpenAI's call for democratic inputs to AI (Zaremba et al., 2023) and Anthropic AI's Collective Intelligence Project (2024), seem to engage more deeply with definitions and implications of democracy for AI. However, on the whole, we must urgently "reflect on [our] engagement with other fields" (Wahle et al., 2023). In addition, instead of using democratization to mean increasing access, we echo Seger et al.'s (2023) call to simply use the word "access" rather than "normatively loaded language" like "democratization."

Limitations

In our analysis, we may have missed relevant NLP and ML literature that treats "democratization" or "democracy" through our focus on the ACL Anthology, ICLR, ICML and NeurIPS. In addition, our filtering of excerpts based on keywords like "democra" may have caused us to exclude important discussions of democracy-adjacent concepts that do not use the word. This may have been worsened by parsing errors stemming from our methods and the Semantic Scholar API. The Semantic Scholar API can also fail to correctly predict scholarly metadata, including fields of study and intent, which may affect our results. Furthermore, our discussion of theories of democracy (§2) is far from exhaustive, given the rich history of the subject.

Ethical Considerations

Our analysis complies with the terms of usage of Semantic Scholar. Our paper emphasizes careful consideration and usage of the term "democratization," especially given its relation to democracy, and urges drawing from extra-disciplinary literature on democratic theories. This is important for accurately representing the distribution of power, public control, and progress in NLP. In light of our findings, we stress that our analysis only captures a snapshot in time and that researchers' perspectives on "democratization" and "democracy" can evolve over time; moreover, the text of papers may not wholly reflect the perspectives of their authors, given the diversity of opinions among authors and reviewing incentives.

References

Nuri Mahmoud Ahmed and Muntasir Wahed. 2020. The de-democratization of ai: Deep learning and the compute divide in artificial intelligence research. *ArXiv*, abs/2010.15581.

Shakkeel Ahmed, Ravi Mula, and Soma S. Dhavala. 2020. A framework for democratizing ai. *ArXiv*, abs/2001.00818.

Steven Bird and Edward Loper. 2004. NLTK: The natural language toolkit. In *Proceedings of the ACL Interactive Poster and Demonstration Sessions*, pages 214–217, Barcelona, Spain. Association for Computational Linguistics.

Abeba Birhane, Pratyusha Kalluri, Dallas Card, William Agnew, Ravit Dotan, and Michelle Bao. 2022. The values encoded in machine learning research. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '22, page 173–184, New York, NY, USA. Association for Computing Machinery.

Su Lin Blodgett, Solon Barocas, Hal Daumé III, and Hanna Wallach. 2020. Language (technology) is power: A critical survey of "bias" in NLP. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 5454–5476, Online. Association for Computational Linguistics.

M. Coeckelbergh. 2024. *Why AI Undermines Democracy and What To Do About It*. Polity Press.

Collective Intelligence Project. 2024. A roadmap to democratic ai.

Nancy Fraser. 1990. Rethinking the public sphere: A contribution to the critique of actually existing democracy. *Social Text*, (25/26):56–80.

Deep Ganguli, Saffron Huang, Liane Lovitt, Divya Siddarth, Thomas Liao, Amanda Askell, Yuntao Bai, Saurav Kadavath, Jackson Kernion, Cam McKinnon, Karina Nguyen, and Esin Durmus. 2023. Collective constitutional ai: Aligning a language model with public input.

Michele Gilman. 2023. Democratizing ai: Principles for meaningful public participation.

Ronald M Glassman. 2017. *The origins of democracy in tribes, city-states and nation-states*, 1 edition. Springer International Publishing, Basel, Switzerland.

Robert E. Goodin. 2000. Democratic deliberation within. *Philosophy & Public Affairs*, 29(1):81–109.

J. Habermas and T. Burger. 1991. *The Structural Transformation of the Public Sphere: An Inquiry into a Category of Bourgeois Society*. Studies in Contemporary German Social Thought. MIT Press.

Matthew Honnibal and Ines Montani. 2017. spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing. To appear.

Pratyusha Kalluri. 2020. Don't ask if artificial intelligence is good or fair, ask how it shifts power. *Nature*, 583(7815):169–169.

Rodney Kinney, Chloe Anastasiades, Russell Authur, Iz Beltagy, Jonathan Bragg, Alexandra Buraczynski, Isabel Cachola, Stefan Candra, Yoganand Chandrasekhar, Arman Cohan, et al. 2023. The semantic scholar open data platform. *arXiv preprint arXiv:2301.10140*.

Lewis Mumford. 1964. Authoritarian and Democratic Technics. *Technology and Culture*, 5(1):1.

Jimin Mun, Liwei Jiang, Jenny Liang, Inyoung Cheong, Nicole DeCario, Yejin Choi, Tadayoshi Kohno, and Maarten Sap. 2024. Particip-ai: A democratic surveying framework for anticipating future ai use cases, harms and benefits. *arXiv preprint arXiv:2403.14791*.

Anaelia Ovalle, Arjun Subramonian, Vagrant Gautam, Gilbert Gee, and Kai-Wei Chang. 2023. Factoring the matrix of domination: A critical review and reimagination of intersectionality in ai fairness. In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*, AIES '23, page 496–511, New York, NY, USA. Association for Computing Machinery.

John Parkinson. 2003. Legitimacy Problems in Deliberative Democracy. *Political Studies*, 51(1):180–196.

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*. Association for Computational Linguistics.

Shawn W. Rosenberg. 2007. Rethinking Democratic Deliberation: The Limits and Potential of Citizen Participation. *Polity*, 39(3):335–360.

Giovanni Rubeis, Keerthi Dubbala, and Ingrid Metzler. 2022. “democratizing” artificial intelligence in medicine and healthcare: Mapping the uses of an elusive term. *Frontiers in Genetics*, 13.

Johnny Saldana. 2021. *The coding manual for qualitative researchers*, 4 edition. SAGE Publications, London, England.

Elizabeth Seger, Aviv Ovadya, Ben Garfinkel, Divya Siddarth, and Allan Dafoe. 2023. Democratising ai: Multiple meanings, goals, and methods. *ArXiv*, abs/2303.12642.

Divya Siddarth. 2023. How ai and democracy can fix each other.

Gayatri Chakravorty Spivak. 1988. *Can the Subaltern Speak?* Communications and culture. University of Illinois Press.

Zeerak Talat, Aurélie Névéol, Stella Biderman, Miruna Clinciu, Manan Dey, Shayne Longpre, Sasha Luccioni, Maraim Masoud, Margaret Mitchell, Dragomir Radev, Shanya Sharma, Arjun Subramonian, Jaesung Tae, Samson Tan, Deepak Tunuguntla, and Oskar Van Der Wal. 2022. You reap what you sow: On the challenges of bias evaluation under multilingual settings. In *Proceedings of BigScience Episode #5 – Workshop on Challenges & Perspectives in Creating Large Language Models*, pages 26–41, virtual+Dublin. Association for Computational Linguistics.

Jan Philip Wahle, Terry Ruas, Mohamed Abdalla, Bela Gipp, and Saif Mohammad. 2023. We are who we cite: Bridges of influence between natural language processing and other academic fields. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 12896–12913, Singapore. Association for Computational Linguistics.

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander M. Rush. 2020. Huggingface’s transformers: State-of-the-art natural language processing.

Erik Olin Wright. 2010. *Envisioning Real Utopias*. Verso Books, London, England.

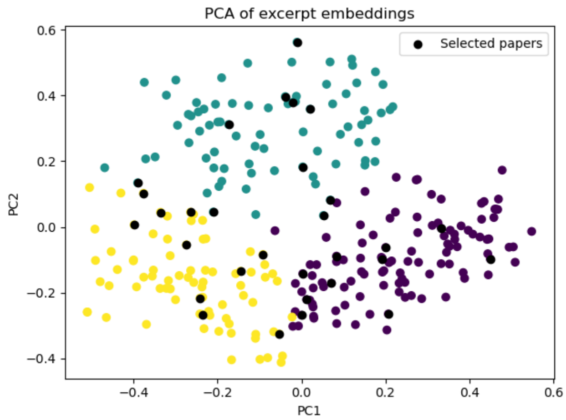

Figure 11: PCA and clustering of excerpt embeddings, along with selected papers.

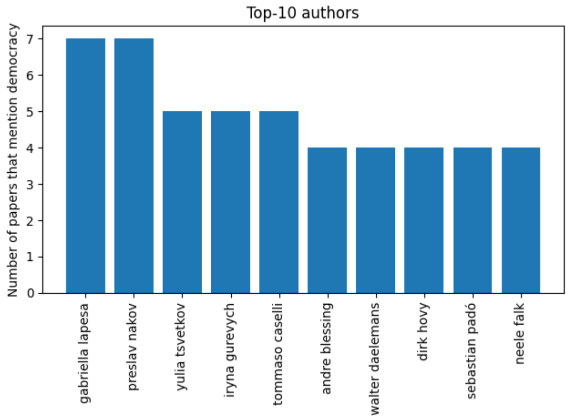

Figure 12: Frequency of authors of papers that mention democracy.

Wojciech Zaremba, Arka Dhar, Lama Ahmad, Tyna Eloundou, Shibani Santurkar, Sandhini Agarwal, and Jade Leung. 2023. Democratic inputs to ai.

A Methodological Details

Table 1 lists all false positive terms that we used in our first stage of manual filtering. Figure 11 shows the results of our PCA and clustering of embedded excerpts, with the darkest colour indicating the papers selected for reading and annotating fully.

B Additional Results

B.1 All concepts and values

Tables 2 and 3 shows all concepts and values we found during excerpt annotation.

B.2 Who is studying democracy?

We present additional results analyzing the authors, institutional affiliations, and funding bodies ac-

democrat	Liberal Democratic Party	Democrat system
Republican Democrat	Democratic Party	Description, Modélisation et Détection Automatique Des Chaînes de Référence
Democrat Republican	German Democratic Republic	DEMOCRAT
Republican and Democrat	Getman Democratic Republic	Democratic
Democrat and Republican	Democratic People's Republic of Korea	christian democratic parliamentary group
Republicans and Democrats	Christian Democratic Union	#democracy
Democrats and Republicans	Democratic Alliance	Democracy party
Republican or Democrat	United Democratic Front	Democrazia Cristiana / Christian Democracy
Democrat or Republican	Democratic Governors Association	#democratic_party
Republicans or Democrats	China Democracy Party	social-democratic political party
Democrats or Republicans	Christian Democrat	social-democratic leader
the Republican and the Democrat	Democratic primary	Center for Media and Democracy
the Democrat and the Republican	Democratic primaries	democratic president candidate
the Republicans and the Democrats	Somali Democratic Party	Stichting Democratie and Media (Democracy & Media Foundation)
the Democrats and the Republicans	New Democratic Party	Swedish social democratic politician
the Republican or the Democrat	Democratic Socialist Party	democratic congressman
the Democrat or the Republican	Liberal Democrat	social democratic movement
the Republicans or the Democrats	Democratic Left Alliance	Christian democratic
the Democrats or the Republicans	Alliance for Democracy in Mali	social democratic, centre-left political party
democratic and republican parties	Syrian Democratic Forces	Democratic Labour Party
Democratic Party of Japan	Democracy Now!	democratic republic of germany
Liberal Democratic Party of Japan	Movement for Democratic Change	Historical Press of the German Social Democracy Online
Social Democratic Party	Democracy Week	Forum voor Democratie, 'Forum for Democracy'
Democratic candidate	Democratic-controlled	centre-right party New Democracy
Democratic candidates	Croatian Democratic Union	Partito Democratico
Democratic republic of the Congo	Kurd Democratic Party	Social Democracy (S)
Democratic presidential candidate	New Democratic Union	Forum Migration and Democracy (MIDEM)
Democratic presidential candidates	ANR Democrat	
Democratic National Committee	Project ANR Democrat	

Table 1: False positives when matching "democra" in corpus.

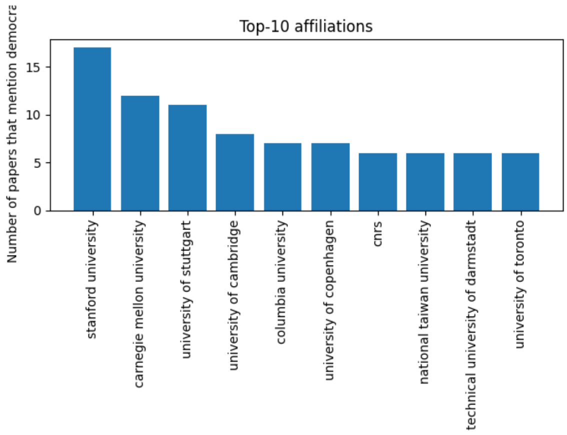

Figure 13: Frequency of affiliations of authors of papers that mention democracy.

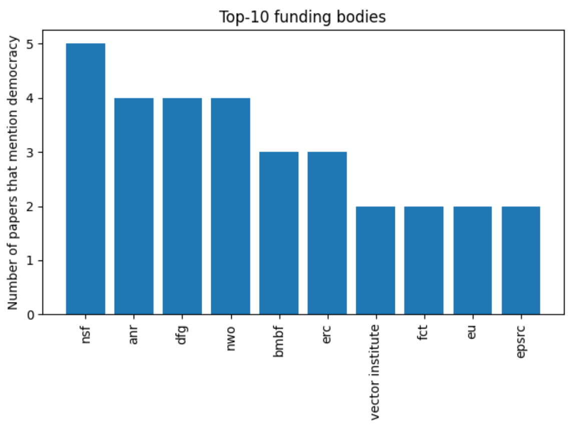

Figure 14: Frequency of funding bodies in acknowledgments of papers that mention democracy.

knowledged in the papers in our data, as well as the sources of extra-disciplinary references.

Methods We extract author names and author affiliations from the Semantic Scholar API. We apply basic cleaning to the affiliations, e.g., removing country and department names in order to normalize them. For each paper, each unique affiliation counts once to the overall frequencies, i.e., if multiple authors of a paper share the same affiliation, this affiliation counts once; if an author has multiple affiliations, each of these affiliations counts once. To extract funding bodies, we first locate paper sections using the Semantic Scholar API and then filter for sections with the substring "acknowledg," "funding," or "disclosure." Only 54 papers had such sections. We then use spaCy (Honnibal and Montani, 2017) to perform named-entity recognition on the texts and collect organizational entities. We exclude some false positives using dependency parsing and filtering out entities that are described as the "corresponding author" or "contact author," or are the subject of phrases like "is supported by." We then normalize the names of all organizational entities, e.g., by converting variants of governmental body names to their acronyms.

generalizability	protection	dialogue
literacy	debate	decentralization
public opinion	freedom	sustainability
fairness	moderation	emotion
WEIRD	replicability	justice
liberties	environment	voting
anti-power	integrity	citizenship
equal contribution	resource-efficient	low-resource
interaction	engagement	broader audience
hierarchy of representatives	multilingual	scalable
rights	news	efficiency
governance	transparency	caution
acceleration	disagreement	civility
reduce barriers	protest	anxiety
discrimination	progress	data
translation	quality	access
happiness	reasoning	power
constitution	harassment	accountability
questioning	majority	consistency
competence	value	social good
reflection	open-source	cohesion
equal representation	evolving	polarization
informed	argument	campaign
fast	available	cooperation
representation	trust	information
responsibility	random selection	inclusion
diversity	quality vs. quantity tradeoff	direct democracy
political party	election	bill writing
correctness	affordable	choice
conflict	ease of use	discourse
equality	distributed	media
education	misinformation	discussion
privacy	participation	propaganda
complexity	critical	benefit
proficiency	censorship	AI
rational	consensus	lack of prejudice
disinformation	deliberation	

Table 2: All associated concepts found when annotating excerpts.

Results Figures 12, 13 and 14 show the frequencies of authors, affiliations and funding bodies, respectively. Many NLP and ML research papers related to democracy and democratization appear to be from well-funded research institutions in countries in North America and Europe, and are often funded by the governments of nations in the Global North as well.

B.3 Where do extra-disciplinary references come from?

For a different view on our results on extra-disciplinary citations, we plot histograms of the most frequent venues and the most frequently cited references. Figure 15 confirms that the most common venues for extra-disciplinary references are political science and social science journals. Figure 16 shows the most frequently cited extra-disciplinary texts are cited for methods, e.g., content analysis, agreement computations, discourse network analysis, or related to fake news and polarization.

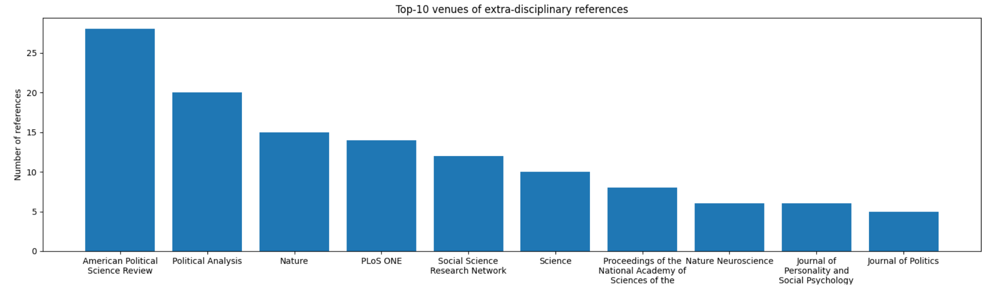

Figure 15: Frequency of venues of extra-disciplinary references cited by papers that mention democracy.

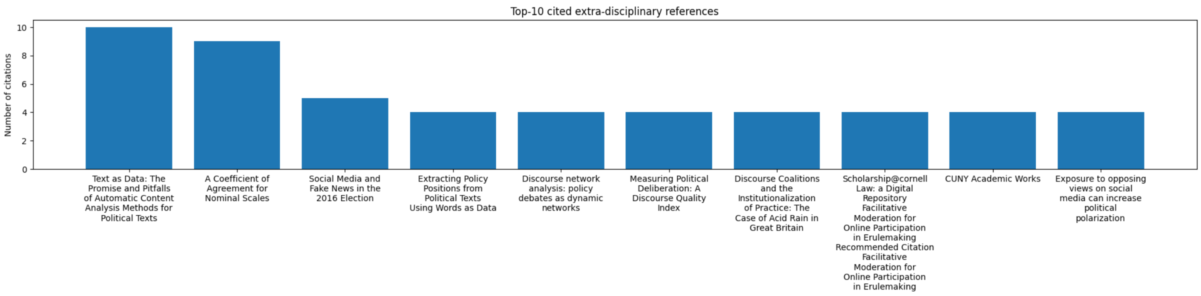

Figure 16: Frequency of extra-disciplinary references cited by papers that mention democracy.

sustainability	disagreement	moderation
fairness	caution	reduce barriers
argument	choice	justice
progress	optimality	direct democracy
trust	participation	rational
random selection	proficiency	resource-efficient
consensus	inclusion	diversity
available	critical	liberties
multilingual	engagement	cooperation
reasoning	interaction	efficiency
generalizability	benefit	open-source
integrity	accountability	reflection
literacy	transparency	access
social good	evolving	decentralization
civility	cohesion	informed
conflict	equal representation	equal contribution
majority	replicability	representation
correctness	equality	debate
privacy	power	distributed
quality	hierarchy of representatives	protection
deliberation	lack of prejudice	affordable
information	rights	discussion
ease of use	dialogue	happiness
responsibility	fast	anti-power
education	value	consistency
scalable	competence	

Table 3: All associated values found when annotating excerpts.