

WASHINGTON UNIVERSITY IN ST. LOUIS

School of Arts & Sciences
Department of Political Science

Dissertation Examination Committee:

Betsy Sinclair, Chair
Dino Christenson
Andrew Reeves
Christopher Lucas
Janet Box-Steffensmeier

A Big Voice in a Small Democracy: Three Papers on American Local Politics
by
Lucas Boschelli

A dissertation presented to
Washington University in St. Louis
in partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

May 2024
St. Louis, Missouri

© 2024, Lucas Boschelli

Table of Contents

List of Figures	v
List of Tables.....	vi
Acknowledgements.....	viii
Abstract	xi
Chapter 1: Is a More Accessible Local Government a More Representative One? The Effects of Online Public Meetings.....	1
1.1 Accessibility and Political Participation	2
1.2 Meeting Participation Across Local Contexts:	4
1.3 Public Meeting Data:	6
1.3.1 Coding Meeting Accessibility:.....	9
1.4 Do Online Meetings Encourage Greater Participation?	12
1.4.1 Measuring Effects on Meeting Participation:	12
1.4.2 Did Accessibility Have Differential Effects?.....	16
1.5 Do Online Meetings Lead to More Diverse Participants?	20
1.5.1 Obtaining Commenter Characteristics:	20
1.5.2 Characterizing Commenters Across Local Contexts	21
1.5.3 Measuring the Difference Between Online and Offline Commenters	23
1.6 Discussion:.....	25
Chapter 2: Is it All Good in the Neighborhood? How Partisanship May Shape Evaluations of Municipal Services	27
2.1 The Process of Polarization	29
2.1.1 How Partisanship Shapes Evaluations	30
2.1.2 Why Schools and Police.....	31
2.2 Data and Methods	32
2.2.1 Local Opinions and Individual Characteristics:	33
2.2.2 Measuring School and Police Performance.....	34
2.3 Full Sample Results.....	36

2.3.1	Polarized Evaluations	37
2.4	A Closer Analysis of Large Cities	39
2.4.1	Reexamining Schools and Police	41
2.4.2	An Alternative to Schools	42
2.4.3	Party Position Taking or Elite Valence?	45
2.5	Conclusion	47
Chapter 3: Utilizing Large Language Models To Transform Messy Text-Based Sources In Useful Data.....		49
3.1	Why is Transforming Text-Based Sources Difficult?	51
3.2	What are LLMs and How Can They Help?	55
3.3	The Proposed Pipeline.....	57
3.3.1	Preparing Text-Based Sources	59
3.3.2	Selecting a Prompt and Context	59
3.3.3	Processing, Iterating, and Validating	61
3.4	Applications and Validation.....	62
3.4.1	Extracting Commenters From Local Meeting Minutes	63
3.4.2	Identifying Interest Groups in Amicus Curiae Briefs	66
3.5	Discussion.....	69
References		71
Appendix A: Appendix For Is a More Accessible Local Government a More Representative One? The Effects of Online Public Meetings		86
A.1	Additional Sample Details:.....	86
A.1.1	Included Local Governments:.....	86
A.1.2	Demographic Characteristics:.....	87
A.1.3	Access By Level of Government.....	88
A.2	Full Tables For Figures:.....	89
A.2.1	Figure 2 Regression Tables:	89
A.2.2	Figure 3 Regression Tables	91
A.2.3	Figure 4 Regression Tables	91
A.3	Supplementary Analyses:.....	93
A.3.1	Simultaneous Comparisons of Commenters to Voters	93
A.4	Text-Analysis of Comments	94

A.4.1	Examining the Content of Public Meeting Comments	94
A.4.2	Do Online Meetings Elicit More Negative Comments?	97
A.5	Survey Details:	98
A.5.1	Question Wording:	98
A.5.2	Survey Responses:	99
Appendix B:	Appendix For Is it All Good in the Neighborhood? How Partisanship May Shape Evaluations of Municipal Services	101
B.1	Survey Details	101
B.1.1	Variable Distributions	101
B.2	Supplemental Analyses	103
B.2.1	Full Regression Results For Table 1	103
B.2.2	Alternative Modeling Specifications	105
B.2.3	Alternative Model Specifications	106
Appendix C:	Appendix For Utilizing Large Language Models To Transform Messy Text- Based Sources In Useful Data	108
C.1	Example Applications	108
C.1.1	Local Meeting Minutes:	108
C.1.2	Amicus Curiae Briefs:	112
C.2	Alternative BERT-based Approach	114

List of Figures

Figure 1.1: Public Meeting Formats Overtime	11
Figure 1.2: Effect of Meeting Format on Public Meeting Participation.....	14
Figure 1.3: Effect of Switching Away From Accessible Meetings By School Board Groups...	18
Figure 1.4: Effect of Switching Away From Accessible Meetings By City Groups.....	19
Figure 1.5: Difference Between Voters and Commenters	22
Figure 1.6: Difference Between Voters and Commenters By Meeting Format.....	24
Figure 2.1: Included Cities Map	40
Figure 3.1: Precinct Level Election Results Example	54
Figure 3.2: Framework for using Large Lanaguage Models to transform text-based sources. .	58
Figure 3.3: Prompt to Extract Commenters from Public Meeting Text Using GPT-3	65
Figure 3.4: Prompt to Extract Signers from Amicus Curiae Briefs Using GPT-3	68
Figure A.1: Access to City Meetings From 2018-2021	88
Figure A.2: Access to School Board Meetings From 2018-2022.....	88
Figure A.3: Access to County Council Meetings From 2018-2022.....	89
Figure A.4: Average Comment Sentiment By Meeting Format.....	98
Figure A.5: Response to Non-Meeting Attendees Question	99
Figure A.6: Response to Meeting Attendees Question.....	100
Figure C.1: Meeting Text Example #1	109
Figure C.2: Meeting Text Example Output #1	110
Figure C.3: Meeting Text Example #2	110
Figure C.4: Meeting Text Example Output #2	111
Figure C.5: Amicus Curiae Brief Example Output #1	112
Figure C.6: Example Model Output #1	113

List of Tables

Table 1.1: Summary of Comments Database	8
Table 2.1: OLS Regression of Local School and Police Evaluations on Partisan Identity.....	38
Table 2.2: Large City Analysis Using CBPS Weighting of Partisanship on Local Evaluations	42
Table 2.3: Hierarchical Linear Regression of Local Roads Evaluation on Partisan Identity: Full Results	44
Table 2.4: Hierarchical Linear Regression of State and Local Co-partisanship on Local Service Evaluations	46
Table A.1: Average Demographic Characteristics of Sample Governments	87
Table A.2: Full Regression Results for Figure 2: Virtual.....	89
Table A.3: Full Regression Results for Figure 2: Livestreamed.....	90
Table A.4: Full Regression Results for Figure 2: Online.....	90
Table A.5: Full Regression Results of Switching Away from Virtual Meetings	91
Table A.6: Full Regression Results of Switching Away from Online Meetings	91
Table A.7: Full Regression Results of Switching Away from Virtual Meetings	91
Table A.8: Full Regression Results of Switching Away from Online Meetings	92
Table A.9: Pooled Commenters Compared to Voters.....	93
Table A.10: Online and Offline Commenters Compared to All Voters	94
Table A.11: STM: Comment Topics By Level of Government.....	95
Table A.12: County Topics of Online Versus Offline Comments	96
Table A.13: City Topics of Online Versus Offline Comments	96
Table A.14: School Board Topics of Online Versus Offline Comments	97
Table B.1: Descriptive Statistics of Independent and Dependent Variables: Full CCES	101
Table B.2: Difference in Covariate Means Between CCES and Large Cities Subset.....	102
Table B.3: OLS Regression of Local School and Police Evaluations on Partisan Identity: Full Results	103
Table B.4: Hierarchical Linear Regression of Local Roads Evaluation on Partisan Identity: Full Results	104
Table B.5: OLS Regression of Local School and Police Evaluations on Partisan Identity: Full Results and Independents.....	105

Table B.6: Hierarchical Linear Regression of 7-Point Party ID on Local School and Police Evaluations	106
Table B.7: Hierarchical Regression Results With Alternative Measures of Police Performance	107
Table C.1: LLMs Compared to Alternative Name Entity Recognition Approach	114

Acknowledgments

Not to start too cheesy, but I would like to thank each member of my committee. Each of them has helped me at a different point along my journey, and without them, I would not have come close to finishing this work. Janet Box-Steffensmeier and Christopher Lucas, whom I admit I should have been in way better contact with, were incredibly supportive of my work and an absolute delight to work with. From struggling to teach undergraduates statistics or battling reviewers over the importance of mergers and acquisitions, working with these two had an immeasurable positive impact on me. I truly appreciate Andrew Reeves for providing positive feedback and critique since day one and for the fantastic opportunities to explore a subject I love. Without Andrew, I would have never discovered my passion for local politics. None of this dissertation would be possible without the encouragement and guidance from Betsy Sinclair, my chair. Betsy was and still is incredibly supportive of me and my interests, and I cannot thank her enough.

Finally, I thank Dino Christenson for helping me become a true *Dream-Liver*. Whether it was surviving the jungle of intro to American politics or foolishly jumping headfirst into well-established fields of research, every moment was some of the most fun I had during my entire six-year long journey. I am eternally grateful for these experiences and, most importantly, his friendship.

I would also like to thank my family and friends for their love and support. I could not do this without my parents' and brothers' help. Thank you so much for challenging me to go further. I would especially like to thank Ben, Ipek, Tony, and Jeremy for letting me waste hours of their productive time with my ramblings. I owe so much to my best friend Lily and partner Dani for everything they have done to support me. Thank you.

Last but certainly not least, I must thank my dog, Bucky. I adopted him at the start of my graduate school journey, and he has been awkwardly staring at me every step of the way. Without

that chunky dog, I would not have survived the journey.

Lucas Boschelli

Washington University in St. Louis

May 2024

Dedicated to everyone who helped me thrive.

ABSTRACT OF THE DISSERTATION

A Big Voice in a Small Democracy: Three Papers on American Local Politics

by

Lucas Boschelli

Doctor of Philosophy in Political Science

Department of Political Science

Washington University in St. Louis, 2024

Professor Betsy Sinclair, Chair

An individual's greatest political voice is often within America's smallest democracies. Through three papers, this dissertation explores three key dimensions of political participation within America's local governments: representation, evaluation, and coverage. The first paper explores the question of whether a more accessible local government is a more representative one. By leveraging the switch to online meetings induced by the COVID-19 pandemic, I analyze whether the shift to more accessible meetings promotes healthy democratic participation or amplifies pre-existing inequities in local political participation. To accomplish this, I first construct a novel dataset of public meeting minutes from 2018 until 2022 for county, municipal, and school board governments within the St. Louis region. I then combine public participation data from these minutes with existing political and demographic datasets. Through a time-series cross-sectional analysis, I find that virtual meetings significantly increased participation in some local contexts but did not affect or decreased participation in others. Importantly, I find evidence that while virtual meetings experience similar inequalities in who participates as in offline meetings, they garner greater participation from women in county and school board governments.

Before individuals decide whether to voice their concerns to a local government, they must, in some way, evaluate its performance. In the second paper, I examine how individuals form these evaluations and whether and how an individual's partisan identity may shape their evaluations. I

argue that for some services, for which parties hold ideologically distanced and distinctly homogeneous opinions, evaluations of local services will be biased toward their party's average position regardless of a service's actual performance. To test this theory, I examine individual's evaluations of their local schools and police. Through a cross-sectional analysis of national survey respondents matched to measures of their local conditions, I find that for polarized services such as policing, individuals have a systematic bias in favor of their party's position regardless of the service's objective performance. Additionally, this bias exists irrespective of the partisan control of state and local governments.

Finally, in the third paper, I introduce a promising framework to address a problem often hindering our coverage of local political data: the transformation of text-based sources. Political and legal texts include many meaningful data we need as researchers to quantify; however, these texts are often too complex to transform automatically and require slow and costly hand coding to do so correctly. I propose a novel approach: a human-in-the-loop pipeline that leverages Large Language Models (LLMs) to crowdsource and transform text-based sources into quantifiable datasets. Through case studies in local and judicial politics, I demonstrate that LLMs are a cost-effective alternative, producing results comparable to professional coders while significantly reducing transformation time. By easing the costs to process text-based sources, this framework promises to improve the accessibility of research and the coverage of important political outcomes.

Overall, this dissertation makes significant contributions to understanding political participation in America's local governments by examining representation, evaluation, and coverage. It not only sheds light on the complexities and inequities of local democratic participation within an increasingly nationalized context but also introduces innovative methods to expand our understanding of local political dynamics.

Chapter 1:

Is a More Accessible Local Government a More Representative One? The Effects of Online Public Meetings

In response to the growing COVID-19 pandemic, the Denver City Council began hosting their general public meetings online, allowing citizens to attend and participate in discussions virtually. This switch was largely positive for the council, seemingly increasing attendance and active engagement with its meetings. In light of these benefits, and at the request of the public, the council decided to maintain online access to their meetings even after switching back to in-person meetings. Denver was not alone in this continuation of pandemic-era online meetings. Numerous other local governments across the country, citing improved access, still allow citizens to participate or at least watch public meetings virtually.¹ Despite this evidence of success, there has been relatively little work examining the empirical effects these online meetings have on political participation within local governments.

Public meetings, in particular, are a critical component of American local democracy. They allow citizens to participate directly in the policy-making process (Adams 2004) and act as a conduit through which local officials can learn constituent preferences (Einstein and Glick 2018; Cho et al. 2021; Fung 2006). However, participation in these meetings is rare, and those who attend tend to be socially or economically advantaged compared to their surrounding communities (Einstein, Palmer, and Glick 2019; Yoder 2020). Scholars and policy advocates have theorized that by reducing participatory costs, online meetings could address these inequities and increase the number and diversity of meeting participants (ICMA 2020). Yet, in one of the few studies of online meetings Einstein et al. (2022) finds that, in the case of planning and zoning meetings, online access

¹ Some notable examples include municipal and county governments in Chicago, IL, St. Louis, MO, and Austin, TX

fails to engage under-represented groups, and participants are similarly misrepresentative of their surrounding communities. In light of this work, however, we still know relatively little regarding the long-term impacts online meetings had on American local democracy, especially outside the context of the pandemic. Moreover, it remains to be seen how such structural changes to local meetings may have impacted other forms of local governments, such as county councils or school boards.

Through this article, I examine online meetings' long-term effect on political participation within local public meetings. By leveraging the staggered switch toward and away from online meetings in the years after the pandemic, I estimate the impact online meetings had on the amount and diversity of meeting participants within county, city, and school board governments. Overall, I find that online meetings had mixed effects on local democracy. They increased the average number of public commenters in the aggregate but did not do so universally across types of local government. Some local governments experienced drastic increases in participation following their switch to online meetings, while others experienced no change. While commenters vary significantly between forms of local governments, online commenters were equally misrepresentative of their surrounding community as their offline counterparts. In the rare instance when online meetings alleviated disparities in who participates, other dimensions of inequities, such as race, were exacerbated. Overall, my results suggest that online meetings are potentially an effective tool to expand access to local governments, but in isolation, they are insufficient to address pre-existing inequities in who participates within local democracy.

1.1 Accessibility and Political Participation

Attending a local political meeting is a potentially burdensome form of political participation for many individuals. Meetings are often held late into the evening on weekdays in schools or government centers. Individuals likely have to arrange for things such as transportation or childcare to attend. By allowing individuals to participate virtually, online meetings alleviate many of these

associated costs, as individuals can watch and participate from anywhere with internet access. Improving political participation by minimizing its social, monetary, or time-related costs is far from a new concept in the study of political science. A broad literature exists documenting the benefits of improved accessibility on political behavior and representation. In the context of elections, vote-by-mail and same-day registration, initiatives can increase turnout, especially among low-propensity voters (Barber and Holbein 2020; Bonica et al. 2021). Online access, in particular, can substantially lower the costs of participation as individuals do not have to be physically present to participate. In some instances, digital participation can increase the turnout of traditionally underrepresented groups and attract groups previously disengaged in politics (Krueger 2002; Dalton 2017). Thus, it is reasonable to expect online meetings to have similar long-term effects as they minimize many of the costs associated with commenting at local public meetings. However, evidence exists that increased accessibility does not necessarily translate into positive political outcomes.

While improvements in political accessibility may make participating easier, they may not foster diverse participation. Without efforts to mobilize under-represented groups in particular, disparities in participation may be reinforced or amplified (Verba, Schlozman, and Brady 1995). Notable examples of this occurring are early voting reforms (Burden et al. 2014), online political activism (Chan 2021), and 311 services (Wichowsky, Shah, and Heideman 2022). In each instance, the reforms designed to make participation more accessible largely reinforced pre-existing socioeconomic divides. Notably, Einstein et al. (2022) conclude in their analysis of the early implementation of online meetings that this lack of engagement was one of the critical reasons online meetings failed to diversify public meetings. Thus, online meetings—even in the long term—may be inadequate to improve the diversity of commenters within local meetings unless accompanied by sufficient mobilization and outreach efforts.

1.2 Meeting Participation Across Local Contexts:

As required under most state laws, local governments hold meetings open to the public, and most offer an opportunity for citizens to participate directly in the policy-making process. Despite this, previous studies of participation in public meetings have primarily focused on the context of city governments (e.g., Einstein, Palmer, and Glick 2019). And they have done this for a good reason. Cities play a prominent role in individuals' lives and are the local government they are most likely to interact with directly. They also tend to offer a plethora of opportunities for input in local policy-making processes (Nabatchi and Amsler 2014; Pin 2022; Cho et al. 2021; Wichowsky, Shah, and Heideman 2022) and often rely on these public meetings to gauge public opinion and learn of issues for their constituents (Einstein and Glick 2018). In terms of records, cities are also likely to keep high-quality meeting minutes transcripts, which are vital to constructing public meeting participation datasets. However, cities are just some of the local governments to hold public meetings to learn their citizens' preferences and concerns. Both school boards and county governments offer similar opportunities.

School boards often use public meetings as a vehicle to guide school closures and curriculum changes. Much akin to municipal governments, school boards host various public meetings and forums to inform constituents of upcoming changes or gather community feedback. These meetings and the community connection they provide can be vital for the success of school boards (Morel 2021; Stone 2001; Tracy and Durfy 2007), and public policy scholars have paid increased attention to the impact these meetings have both on policy outcomes (Morel 2018; Hartney and Finger 2022) and the individuals who participate within them (Nuamah 2021; Collins 2021; Nuamah and Ogorzalek 2021). While invaluable to our understanding of direct democracy and public education, most studies only examine individual case studies commonly surrounding school closures and rarely examine how the structure/accessibility of the meetings may influence participation (although see Collins 2021). Understanding whether school boards exhibit the same levels of inequality well-documented at the municipal level is vital, especially as critical issues surrounding race, gender,

and public health are increasingly decided at the school board level.²

Despite counties' vital role within the American government, they have primarily remained the "forgotten government" of local politics. We know little about how public participation and opinion shape county-level politics (Benton 2005). Previous work surrounding counties has primarily examined how the composition and characteristics of public officials have shaped county outcomes (e.g., Mohr et al. 2019). Crucially despite the fact that counties are often constrained by both state and municipal governments (Peterson 1981), they are responsive to constituent preferences and susceptible to traditional political influences (de Benedictis-Kessner and Warshaw 2020b). Thus public meeting participation may, as it does at other levels of local government (Einstein, Palmer, and Glick 2019), play an influential role in county politics. Counties, much like other local governments, hold regular meetings open to the public. Many of these meetings allow the public to comment directly on agenda item issues or other concerns. County governments often use these meetings to gauge the public's support for new budgets, large infrastructure projects, or electoral changes. During the COVID-19 pandemic, these meetings became a vital source of information and expression for citizens surrounded by ever-changing health guidelines and restrictions (Lynch, Bernhard, and O'Neill 2022). We know very little regarding who attends these meetings, the issues they bring up, or the impact of their participation. Do counties experience similar levels of participation found in other local governments, or are they also forgotten in the eyes of their constituents?

Whether participation in public meetings—or the impacts of improved accessibility—should differ across these local government levels is unclear. In their book examining elections and democracy within local politics, Oliver, Ha, and Callen (2012) suggest that political participation fundamentally changes as local governments increase in size, scope, and bias.³ As local governments grow, individuals tend to engage with politics less as their connection to political issues and political efficacy decrease (Oliver 2001; Schlozman, Verba, and Brady 2012). Thus, in the context of public

²As an example, see Hartney and Finger (2022)

³As defined by the authors, bias refers to a government's ability to distribute resources among its constituents.

meetings, we should expect differences in participation between school boards, cities, and counties simply due to their differences across these dimensions. In theory, individuals should participate the most in school board meetings, as their ability to enact direct policy change is the greatest and their ties to school board issues the most tangible. Likewise, participation within county governments should be the least due to the wider variety of issues and constituent preferences they must address.

However, these early conclusions contrast the historically low participation levels across all levels of local government (Hopkins 2018). Local political issues are often parochial and temporal (Oliver, Ha, and Callen 2012). The habitual nature and value of repeat participation in national-level politics may not exist in local politics (Aldrich, Montgomery, and Wood 2011). While events such as school closings, zoning decisions, or election administration are significant political events for these governments, they likely do not affect individuals on a consistent enough basis to warrant intense engagement (see Nuamah and Ogorzalek 2021). Moreover, work such as Hopkins (2018) or Hersh (2020) suggest that individuals care more about national issues and are often disengaged from local political happenings. Therefore, it may not matter how these local governments differ because the American public is equally unengaged across all levels, and participation is dominated by a subset of advantaged individuals (e.g., Einstein, Palmer, and Glick 2019; Yoder 2020). Understanding how participation differs between local governments is a critical but unaddressed question in political science. A question this paper hopes to address.

1.3 Public Meeting Data:

In order to examine the effects of online meetings and expand coverage beyond previous local boundaries, I construct a novel dataset of public meeting participation from the school board, city, and county governments within St. Louis County, Missouri. I chose St. Louis because of the quality of public records, partly due to Missouri's detailed sunshine law, and the county's many overlapping local governments. By focusing on close or overlapping local governments, I can examine the differential effects online meetings may have while controlling for similar regional,

state, or federal factors. All of these are especially important given the surrounding context of the COVID-19 pandemic, which elicited a cooperative response from almost all levels of the American government. In total, I collect records of public participation from 22 school boards, 15 municipal governments, and the county itself. These local governments cover a broad spectrum of sizes, with the largest government being St. Louis County (1,001,982) and the smallest being Brentwood City (8,198). Notably, my sample's city and school board governments are relatively diverse and vary in terms of racial makeup, median income, and home ownership. I provide a more detailed breakdown of my sample's demographics and the selection method used to construct the sample within the Appendix. While far from exhaustive, the novel dataset covers 22% of the county's municipal populations, 52% of its school board populations, and the entirety of the county.

For each of these governments, I download and digitize their general meeting minutes from 2018 until late 2022. These are the routine public meetings held by each government. By law, these meetings are open to the public, provide opportunities for public comments, follow a set schedule, and maintain detailed minutes of all topics discussed. Given their routine nature and general purpose, these meetings provide a roughly equivalent comparison point between the three levels of local government within my sample.⁴ While each government's meeting minutes vary in terms of how detailed they are, most record the names, addresses, and comments of all individuals who participate in the meeting.⁵ In a few rare instances, local governments only record the total number of commenters present. For these cases, I treat each commenter as a unique individual. I present the detailed breakdown of my final dataset within Table 1.1. Overall, my dataset includes over 10,000 unique comments submitted by over 8,000 individuals across approximately 2,213 local political meetings.

⁴Local governments hold other meetings addressing specific policy issues, such as zoning or industrial development. These meetings are also often open to the public and have been the focus of previous studies of public meeting participation (see Einstein, Palmer, and Glick 2019).

⁵This includes individuals who submit their comments digitally.

Table 1.1: Summary of Comments Database

	Level of Government		
	County	City	School Board
Years:	2018-2022	2018-2021	2018-2022
Number of Governments:	1	15	22
Number of Meetings:	218	1029	966
Population Coverage:	100%	22%	52%
First Name:	✓	✓	✓
Last Name:	✓	✓	✓
Address:	✓	✓	✓
Number of Comments:	7069	1206	1892
Number of Individuals:	5621	926	1555
Commenting Rate:	0.56%	0.20%	0.19%

Note: The table above provides the summary of the universe of comments within the data set. Population coverage refers to the estimated proportion of the population of St. Louis County covered by the sample.

From this initial dataset, we can glean some insight into public participation across these levels of government. Participation within these meetings is rare across all three forms of local government. Only 0.56% of residents comment at county level meetings over my sample's four years of coverage, with municipal (0.20%) and school board (0.19%) governments experiencing slightly lower rates of participation. Individuals also appear to only comment once within meetings and then disappear from the dataset. In fact across all three levels of local government only 15% of individuals ever submit more than one comment.⁶ On average individuals appear to comment regarding a specific issue and then stop actively participating within local meetings. This pattern falls in line with previous conceptions of local democracy and suggests that participation within this realm is deliberate and motivated (Oliver, Ha, and Callen 2012; Nuamah and Ogorzalek 2021). Individuals reach out to their local governments when faced with a pertinent issue and then resume

⁶These low levels of repeat participation are not unique to my dataset. Other datasets of meeting participation such as that from Yoder (2020) have similarly low levels and cover a far more extensive 10 year timeframe.

their lives once it has been addressed. In the broader context of this paper, meeting accessibility may be more about encouraging new groups to participate rather than fostering repeat participation. Given the unique overlapping nature of my dataset, I can also examine participation across levels of government.⁷ Are individuals active at one level of local government also active in others? The short answer appears to be no. Cross-level commenting appears exceptionally rare with a little over 1% of commenters ever commenting at more than one level of local government. While this finding may be a product of the relative short time frame covered by the data, it may provide further evidence of deliberate and conscious participation within local democracy. Individuals appear to only engage the local government pertaining to their specific issue.⁸

1.3.1 Coding Meeting Accessibility:

Access to local public meetings took many different forms and styles following the pandemic. Each allowed members of the public to watch and participate to varying degrees in these meetings. I classify each meeting in my sample into three broad categories based on the relative level of accessibility they provide to the general public. First, *Offline* meetings are those with traditional access. Individuals must physically attend these meetings to watch and participate. In most cases, local governments will make audio or video recordings available a few days after the meeting, but individuals must be present to participate. Second, *Livestreamed* meetings are those in which local governments stream the meeting to some internet source, and individuals can watch meetings in real-time. For some governments, individuals can submit comments before the meeting and then watch them be addressed live. Finally, *Virtual* meetings are those in which individuals can attend and participate entirely virtually. These were commonly conducted over videoconferencing software, but some local governments allowed the public to call into the meeting and comment. These categories are agnostic to how local governments conduct their meetings (i.e., whether they

⁷To match individuals across levels of local government I fuzzy string match on commenters first and last name. My estimates of cross-level commenting are likely larger than the true rates for my sample.

⁸With in the Appendix I provide further evidence of this through text-analysis. Individuals appear to approach local governments with specific and often jurisdictionally relevant issues.

use videoconferencing software, email, phone calls, etc.) and focus instead on what they allow citizens to do without being physically present.

Notably, there are apparent similarities between livestreamed and virtual format meetings. Both improve access to public meetings and allow individuals to attend meetings without being physically present. Individuals do not have to travel to government centers and can theoretically attend from anywhere. Given this similarity, I drop this distinction in some future analyses and codify these meetings as simply *Online*. One critical difference between the two, however, is that virtual meetings offer the chance for deliberation between commenters and local public officials. Active deliberation within public meetings can deepen democratic ties (Bryan 2010; Gastil, Deess, and Weiser 2002; Fishkin 2009), produce more responsive policies (Roberts 1997; Fung 2006), and increase trust in local institutions (Carpini, Cook, and Jacobs 2004). In fact, experimental evidence finds that simply watching public meetings containing deliberation between officials and the public can increase both individuals' trust in public officials and willingness to attend future meetings (Collins 2021). An effect that may only increase for individuals who actually participate (Gastil, Deess, and Weiser 2002; Minozzi et al. 2023). Thus if accessibility matters for public participation in local public meetings, we should see the largest changes in virtual meetings, given their additional capacity for deliberation.

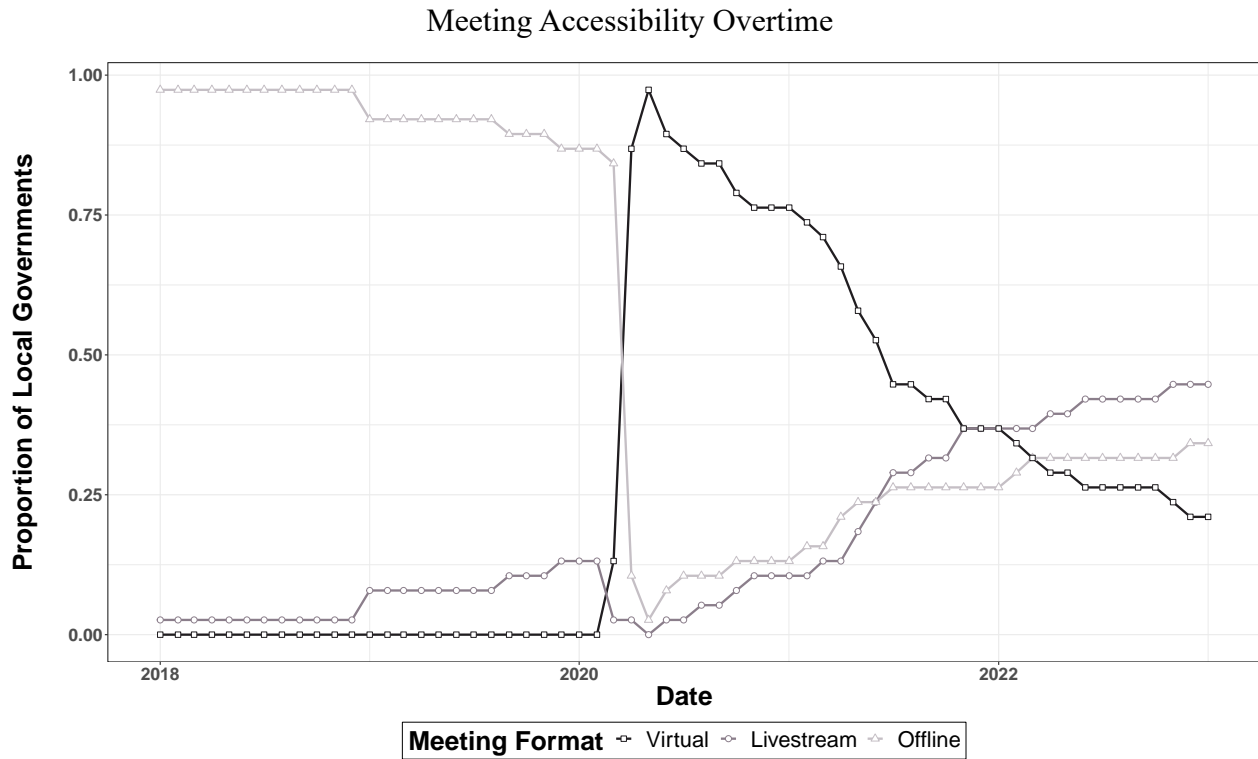


Figure 1.1: The figure presents the format of local political meetings by month from January 2018 until May 2022.

For each meeting within my dataset, I classify them into one of the three categories outlined above. As local governments conduct meetings on different schedules and cadences, I aggregate participation and access at the monthly level. I code accessibility based on whether most meetings during a given month belong to one of the three categories. Figure 1.1 displays access to the local government's meetings from 2018 until late 2022. I provide similar figures for each level of government within the Appendix. While most governments within my sample held offline meetings before the pandemic, a few cities (18%) and school boards (6%) experimented with holding livestreamed meetings beforehand. No governments in my sample held virtual meetings before the pandemic. Following the pandemic, however, every local government appeared to switch to some form of online meeting format. Governments then varied in terms of whether and when they switched away from online meetings, with many choosing to continue hosting online meetings in the years following the pandemic. Interestingly, livestreamed meetings became the modal meeting

format for most governments, with offline meetings being the second most common. These trends, however, varied by level of government, with most cities eventually switching back to offline style meetings and both school board and county governments tending to livestreamed meetings. Virtual meetings, while the least common across all three levels of government, were still implemented by a few school board and municipal governments by the end of my sample coverage. In the following section of this paper, I leverage this staggered adoption to (and away from) online meetings to analyze whether the additional accessibility they provide translates to increased and more diverse participation.

1.4 Do Online Meetings Encourage Greater Participation?

1.4.1 Measuring Effects on Meeting Participation:

In response to the developing COVID-19 pandemic, all governments within the state of Missouri were required to hold some form of online meeting from April 2020 until May 2020.⁹ Following the month-long period, local governments could decide whether to switch back to offline meetings or continue hosting their meeting online. This initial shock for local governments provides near perfect opportunity to measure the impact of online meetings on public participation as I can: (1) estimate *within* government effects by comparing governments before and after their switch to online and (2) estimate *between* government effects by comparing governments as they adopt differing meeting formats following the pandemic. The critical question is whether the increased accessibility afforded by online meetings increased rates of public commenting. These meetings should lower the individual costs of participation and hopefully engage additional citizens. To test this initial analysis, I employ the following two-way fixed effects (TWFE) design:

$$Y_{it} = \beta_0 + \beta_1 D_{it} + \gamma_i + \alpha_t + \epsilon_{it} \quad (1.1)$$

⁹The exact order was referred to as the “Stay Home Missouri” Order.

where Y_{it} is the total number of participants and D_{it} is an indicator of whether government i held primarily livestreamed, virtual, or online meetings during month t . The two-way government (γ_i) and month (α_t) fixed effects to control for the time-invariant characteristics within local governments as well as monthly shocks, which may influence the number of participants at local meetings. Finally, given that I focus on the number of participants within local meetings, I model the above equation using a negative binomial regression.

To examine whether the effects of online meetings are consistent across levels of local government, I estimate both a pooled and government-level model for each local government in my sample. I utilize the same model definition for both school board and city governments but subset the data to each respective type of government. For inference at the county level, however, my sample only includes a single county government. This fact makes it difficult to employ any form of difference-in-difference design. In order to maintain comparability and interpretability between models, I opt to employ the following simplified model:

$$Y_t = \beta_0 + \beta_1 D_t + \delta_t + \epsilon_t \quad (1.2)$$

where I forgo the government and month-fixed effects and substitute δ_t to capture year-based fixed effects. While this model loses the casual leverage provided by the TWFE modeling approach, it should still provide observational evidence as to whether accessibility matters for county meetings.

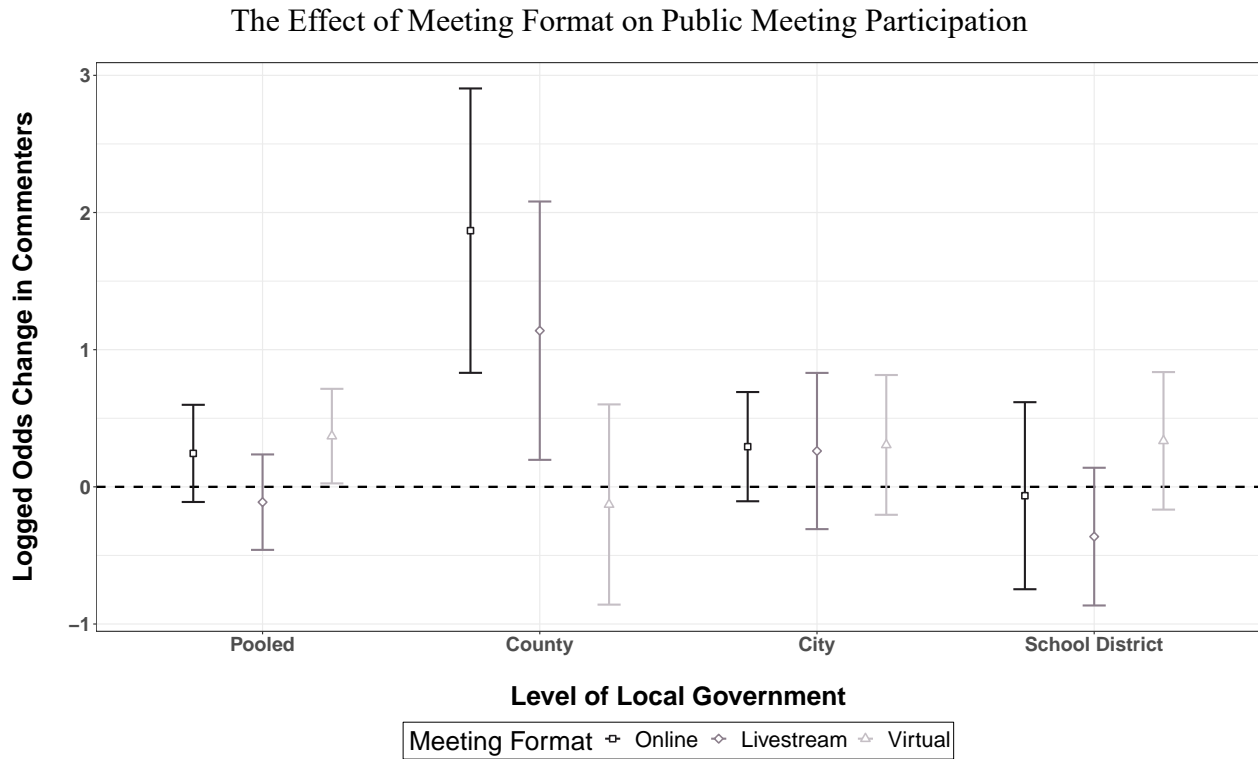


Figure 1.2: Full regression results are available within the Appendix

Figure 1.2 presents the estimated coefficients of virtual, livestreamed, and online meeting formats from my pooled and government-level models. For the pooled model, only the virtual meeting coefficient appears positive and statistically significant, with a value of 0.368. This estimate translates to a sizeable 45% increase in the average monthly commenters or roughly 2-3 additional commenters per month.¹⁰ While seemingly minor, these additional commenters could have a significant impact given the smaller scale and the increased individual political efficacy of local democracy (Oliver, Ha, and Callen 2012). Virtual meetings appear unique in their effect on meeting participation. Both online and livestreamed meetings appear not to significantly affect meeting participation. Intuitively this makes sense as virtual meetings allow individuals to participate without having to submit their comments ahead of time or physically attend. This added flexibility likely overcomes the traditional costs of meeting participation. Additionally, virtual meetings'

¹⁰These effects appear robust to alternative modeling strategies such as lagged dependent variable regression and the inclusion of government-level fixed effects. However, the effects lose significance but maintain direction when using more robust estimation strategies such as Imai and Kim (2021)'s weighted linear fixed effects model.

capacity for deliberation may make participation more appealing and foster increased commenting (Collins 2021). Interestingly, the coefficient surrounding livestreamed meetings is negative, albeit not significant. This potentially negative result could be due to livestreamed meetings' potential as a substitution rather than a promotion of meeting participation. Individuals can theoretically watch local discussions without having to participate themselves and voice their concerns only if they are not appropriately addressed within the meeting. Pooled together I find evidence that virtual meetings, in particular, fostered additional participation within local meetings, but the effect of these meetings may change between levels of local government as they vary in size and scope.

For the county model, livestreamed and online meetings appear to significantly increase meeting participation with coefficients of 1.139 and 1.868, respectively. Virtual meetings alone did not significantly affect meeting participation for the county. These abnormally large and significant coefficients translate to increases in meeting participants of 212% for livestreamed meetings and 547% for online meetings in general. Counties are geographically larger and serve more expansive constituencies; thus, they may benefit most from the added accessibility of online meetings. Whether meeting accessibility is the sole cause of these large effects, however, is unlikely. Given the county model's reduced form, within year time-varying confounders—namely the spread of the COVID-19 virus and discussions of racial inequities in policing—likely inflate the estimates. Examining the contents of the meeting minutes confirms this suspicion. Both meetings in which the county experienced the greatest number of participants involved discussions on reopening schools during the pandemic and the continuation of the county's mask mandate. These meetings alone drew over 2,500 comments from citizens, far above the average 49 the county tended to receive before then. The simple fact that so many citizens *could* participate may be evidence enough that accessibility matters for participation in county meetings, even if their exact causal impact is beyond the design of my simplified model.

Turning to the city and school board models, the effects of online meetings become muddled. Online meetings of either kind do not significantly influence participation. The effect of virtual

meetings is in the same direction as both the pooled and county models, but the effects of livestreamed and overall online meetings point in opposite directions for each level of government. These null effects may suggest that for these smaller governments, accessibility may matter far less than other political factors. While online meetings may remove the burden of participating in meetings, individuals may not have a reason to attend these meetings or are not engaged in local events (Hopkins 2018). Given that I find significant effects at the aggregate level, data restrictions may also be to blame as my dataset does not include enough local governments to detect the potentially nuanced effects surrounding meeting accessibility. Another possibility for the difference in effects between local government levels is that online meetings impacted some local governments and not others.

1.4.2 Did Accessibility Have Differential Effects?

Could online meetings affect local governments differently depending on their individual characteristics? Government and community norms may shape engagement and moderate the effects of online meetings (e.g., Cho et al. 2021; Nuamah and Ogorzalek 2021). Traditional TWFE models assume homogenous treatment effects and do not allow for group-level estimates of the average treatment effect (ATE). Instead, I turn to an alternative multi-period difference-in-difference model proposed by Callaway and Sant’Anna (2021). This alternative model provides two key benefits beyond the traditional TWFE design. First, it allows for heterogenous treatment effects between groups and can estimate a separate ATE for each government in my sample.¹¹ Thus, I can estimate the effects of online meetings based on when individual (or groups of) local government(s) switch away from online meetings. Second, the model allows for the staggered adoption of treatment and therefore avoids the potential inferential concerns which can occur when applying traditional TWFE models to multi-period difference-in-difference designs (Imai and Kim 2021; Sun and Abraham 2021).

¹¹Small groups can cause estimation problems for this model, so I follow the authors’ advice and average the ATE for each group across all time periods in which they are treated.

Notably, the model does require that once a unit is treated, it remains treated for the remainder of the data. Thus, I alter my previous design and leverage that all local governments had to hold some form of online meeting following the Stay Home Missouri order. In doing so, I effectively flip my treatment indicator and examine the effect of switching *away* from online meetings.¹² If online meetings *increase* participation, then I should find *negative* treatment effects as local governments effectively lose participation by switching back to offline meetings. To leverage all possible changes in meeting accessibility, I estimate the effect of entirely switching away from online meetings and the effects of switching away from only virtual meetings. Given my focus on heterogeneous effects within local government, I estimate separate models for city and school board governments. One last assumption of the model is that units do not anticipate treatment or at least have limited anticipation of it. Local governments could theoretically anticipate switching away from online meetings and notify their constituencies, thus influencing the total number of commenters. This anticipation would bias my results and cause the estimator to become inconsistent. However, I find no evidence of anticipation when examining various lags and leads of treatment or the inclusion of various anticipation terms.

¹²No government in my sample switched back to online meetings after switching away from them.

Effect of Switching Away From Accessible Meetings By School Board Groups

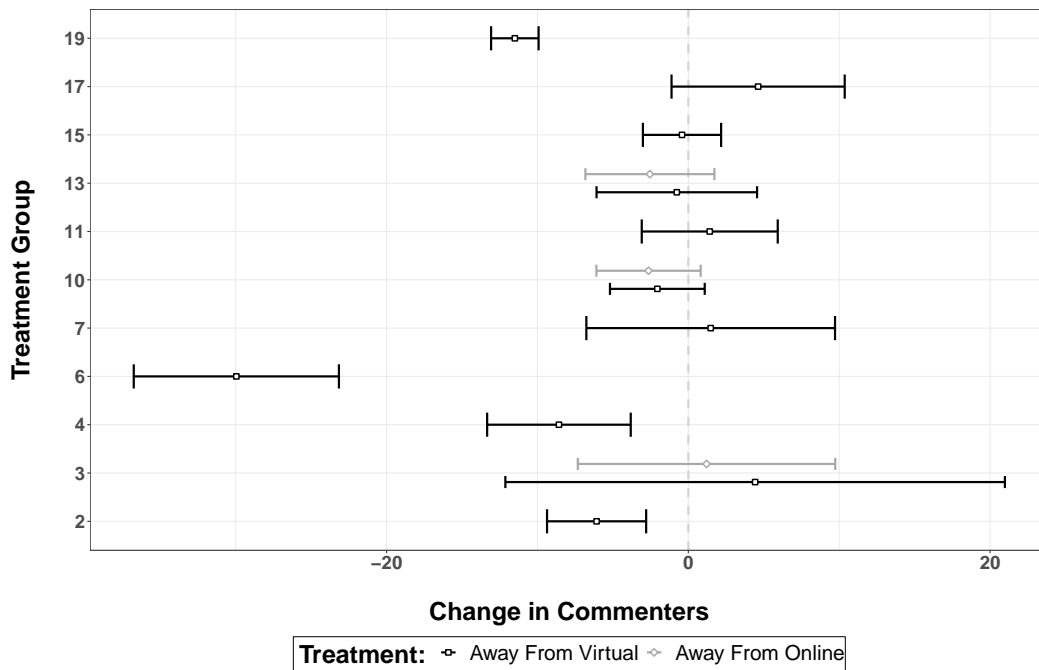


Figure 1.3: Effects are calculated using multi-period difference-in-difference design. X-axis indicates the ATE for a group averaged across all treated periods. Y-axis indicates the treatment group and the number of months passed before treatment. Standard errors are doubly robust and clustered at the group level.

Figure 1.3 and Figure 1.4 present the differential effects of switching away from online meetings for cities and school boards within my sample. The y-axis indicates the treatment group and the number of months before it was treated. For example, group two in Figure 1.3 switched away from virtual meetings two months after the Stay Home Missouri order started. Across both models, all but two groups consisted of a single government. The x-axis captures the group specific ATE in the change of monthly commenters. Immediately apparent, virtual meetings appeared to have differential effects within and between school boards and city governments. Most school boards saw no change in their number of monthly commenters. However, three experienced significant losses in meeting participation after switching away from virtual meetings, with one school board experiencing an estimated decline of 29 monthly commenters. The opposite appears true for cities. Most cities similarly experienced no change in their average monthly commenters; however, in the case of two cities, switching away from online meetings had a positive and significant effect

on meeting participation. For these cities, switching away from online meetings increased their monthly commenters by roughly a single additional commenter.

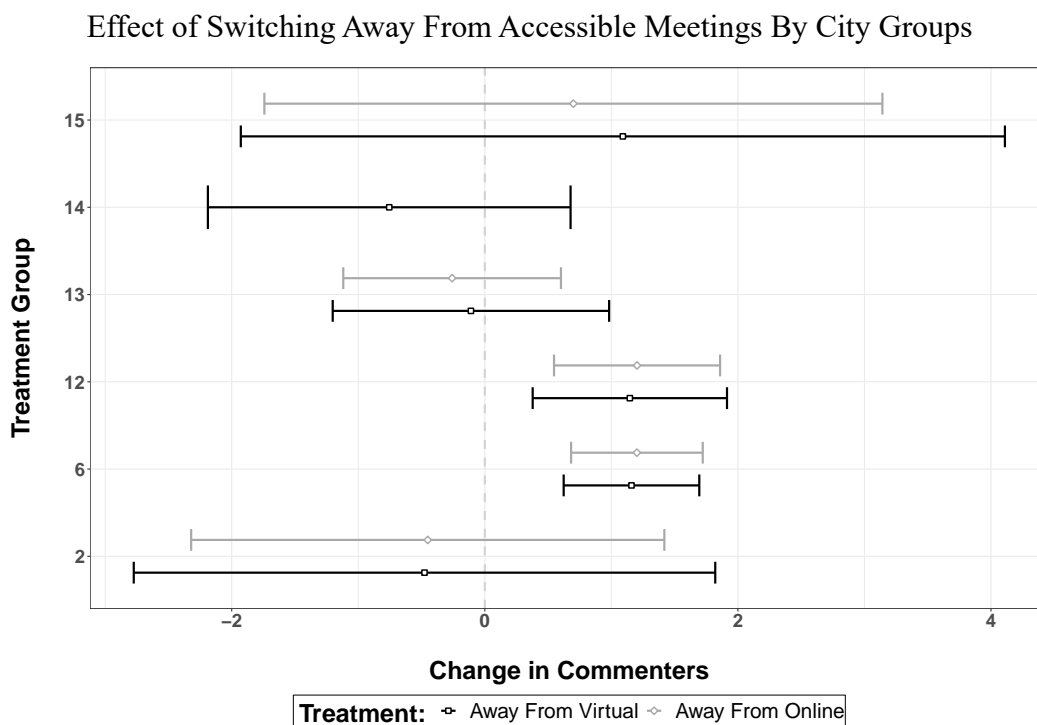


Figure 1.4: Effects are calculated using multi-period difference-in-difference design. X-axis indicates the ATE for a group averaged across all treated periods. Y-axis indicates the treatment group and the number of months passed before treatment. Standard errors are doubly robust and clustered at the group level.

Beyond group-level estimates of the effect of online meetings, these models can also provide additional insight regarding the effect on cities and school boards as a whole. Following Callaway and Sant’Anna (2021), I average the group-level effects to obtain overall estimates of the ATE. Doing so provides a significant effect of -4.30 for school boards switching away from virtual meetings and a non-significant effect of -1.3 for switching away from online meetings altogether. For cities, both the effect of switching away from virtual meetings (0.324) and online meetings (0.32) appear positive but not significant. While the effect of virtual meetings contrasts those found within the school specific TWFE model, it is similar in size to the estimated effects found within the pooled model.

Since I group governments by treatment timing, I also can examine how said timing may

influence the effects of online meetings. Theoretically, how quickly a government switches away from online meetings may moderate their experience as citizens and public officials have varying amounts of time to become familiar with the new meeting styles. In the case of both school boards and cities, I find some evidence that those governments who switch sooner experience more significant changes, but the effect appears to be inconsistent. Not accounting for those with null effects, school boards who switched away from virtual meetings 2-4 months after the Stay Home Missouri order experienced roughly the same participation loss as those who switched a year later. No apparent pattern exists for cities changing away from virtual or all online meetings. These findings suggest that the effects of online meetings are likely more centered on individual government characteristics rather than the timing of their choice to switch.

Overall, this evidence, combined with that of my primary models, suggests that meeting accessibility positively affected participation within local democracy. However, this boost was inconsistent across and within levels of local government. Some governments experienced a drastic increase in participants, while others experienced no change or a slight decline. Given this potential for increased participation, it is essential to understand whether accessibility brings diverse voices or further instills pre-existing inequities in local participation.

1.5 Do Online Meetings Lead to More Diverse Participants?

1.5.1 Obtaining Commenter Characteristics:

In the previous section, I found that more accessible meetings can foster increased participation, albeit the effect varies across local contexts. Are these meetings bringing new voices to local meetings or instilling previous inequities in participation? To answer this question, I match each commenter in my dataset to a commercial voter file provided by L2. This file contains both previous records of political activity and demographic estimates for each individual. I utilize the `fastLink` package in R to probabilistically match commenters on first name, last name, and address

(Enamorado, Fifield, and Imai 2019). While probabilistically matching commenters can lead to higher rates of false positive matchings in which I incorrectly match commenters to voters, it decreases rates of false negatives and increases my overall match rate. The process also allows me to include measures of matching uncertainty within my later analyses. Given the voter file holds more than 650,000 records for the St. Louis region alone, I block voters first on government boundaries and then on zip code.¹³ This process aims to narrow down the universe of possible matches and improve matching rates. I follow previous work matching commenters to administrative records and match commenters with a match threshold of 0.85 (Yoder 2020). In other words, two records are considered a match if the posterior probability of said match is greater than 85%. Before examining whether online meetings encouraged diverse meeting participation, my sample provides a novel opportunity to characterize commenters across local contexts.

1.5.2 Characterizing Commenters Across Local Contexts

Figure 1.5 displays the difference in means between commenters and all voters living within the region. For this initial analysis, I pool commenters across all meeting formats. Regarding school boards and cities, I subset to only voters residing within the jurisdiction of my sample. To account for the uncertainty in my matching procedure, I down-weight each estimate by my match threshold (0.85) as prescribed by Enamorado, Fifield, and Imai (2019). Previous studies of commenting within municipal politics find that individuals who comment are whiter, wealthier, older, and more likely to be homeowners than their immediate communities (Einstein, Palmer, and Glick 2019; Yoder 2020). I find broadly consistent findings when examining cities in my sample. City commenters stand out as predominantly whiter and older, with a higher likelihood of being homeowners and actively participating in national and local elections. Interestingly, I do not find a significant overrepresentation of men as consistently documented in previous studies (e.g, Einstein et al. 2022), albeit the coefficient is in the same direction.

¹³When matching county commenters, I only block voters based on zip code

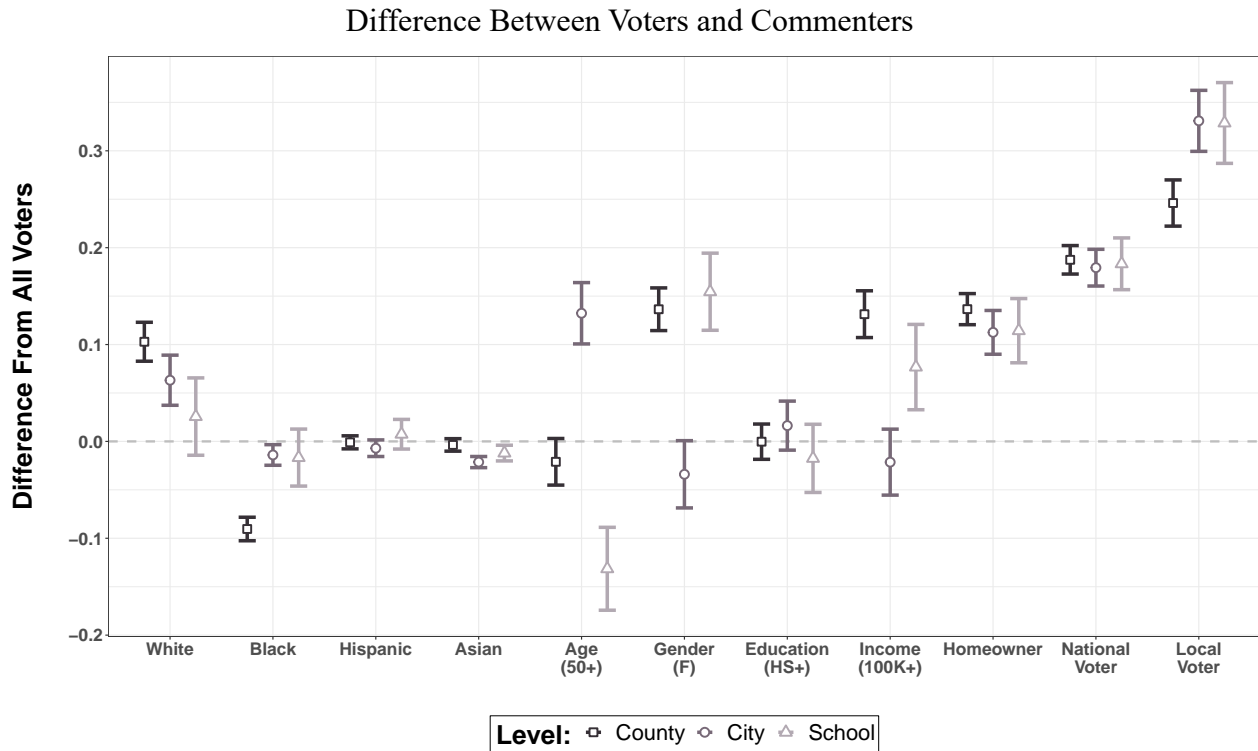


Figure 1.5: Commenters are pooled across all meeting formats. School and city comparisons are subset to their respective voters. Simultaneous comparisons are available in the Appendix.

County governments deal with inherently different issues and represent larger constituencies; however, they are still comparable to local governments in terms of services and powers. When examining the differences between county commenters and their surrounding communities, I find that they also suffer from similar participation inequities as in municipal politics, albeit with several key differences. Commenters are similarly whiter and wealthier, with a higher likelihood of being homeowners and politically active but are not so different from their communities regarding education or age. One of the most surprising differences is that significantly more women comment at the county level. These effects may be due to counties handling issues related to health and welfare, two policy areas women lawmakers are more likely to prioritize (e.g., Huddy and Terkildsen 1993; Volden, Wiseman, and Wittmer 2018; Payson, Fourniaies, and Hall 2023). Additionally, the pandemic's gender-specific impacts may have motivated women to participate at greater rates (Carreras, Vera, and Visconti 2023; Johnson and Williams 2020). I find similarly high rates of

women participating in school board meetings.

Commenters at school board meetings are younger and more likely to be women than their surrounding communities. Intuitively this makes sense, given the traditional and gendered role women tend to hold within households, and younger individuals are more likely to have school-age children. Thus, these two groups are more likely to have a vested interest in school board outcomes. These gendered effects, in particular, provide a new perspective on the difference in gendered political participation within local politics (Coffé and Bolzendahl 2010). Importantly they highlight yet another dimension along which the pandemic potentially impacted genders differently. Along other demographics, school commenters appear to vary from their communities in unique ways. While commenters are more likely to be wealthier, own homes, and vote in recent elections, they do not differ significantly in terms of race. Surprisingly, school board commenters largely resemble the racial makeup of their surrounding communities. This lack of difference may be partly due to the well-documented mobilization of minority communities to protect school resources (Morel 2018; Nuamah and Ogorzalek 2021; Kitchens 2021). However, this seemingly equal participation in public meetings may come at an increased cost of participation paid on behalf of Black individuals especially (see Nuamah 2021). Overall, school board, county, and city commenters appear significantly different from their surrounding communities but do so uniquely depending on the local context and policy priority of the local government.

1.5.3 Measuring the Difference Between Online and Offline Commenters

Online meetings should, in theory, lessen the burdens of political participation faced by different groups and potentially increase the diversity of commenters. However, previous work examining online meetings in the context of planning and zoning meetings has found them to largely attract the same demographics and fail to address inequities in meeting participation (Einstein et al. 2022). To examine the effects of online meetings across my sample, I calculate the difference in means between online and offline commenters relative to all voters living within my sample's geographic

coverage. Doing so allows me to compare commenters to those from alternative meeting formats as well as their communities. Figure 1.6 displays the difference means between each group, again adjusted by my match uncertainty.

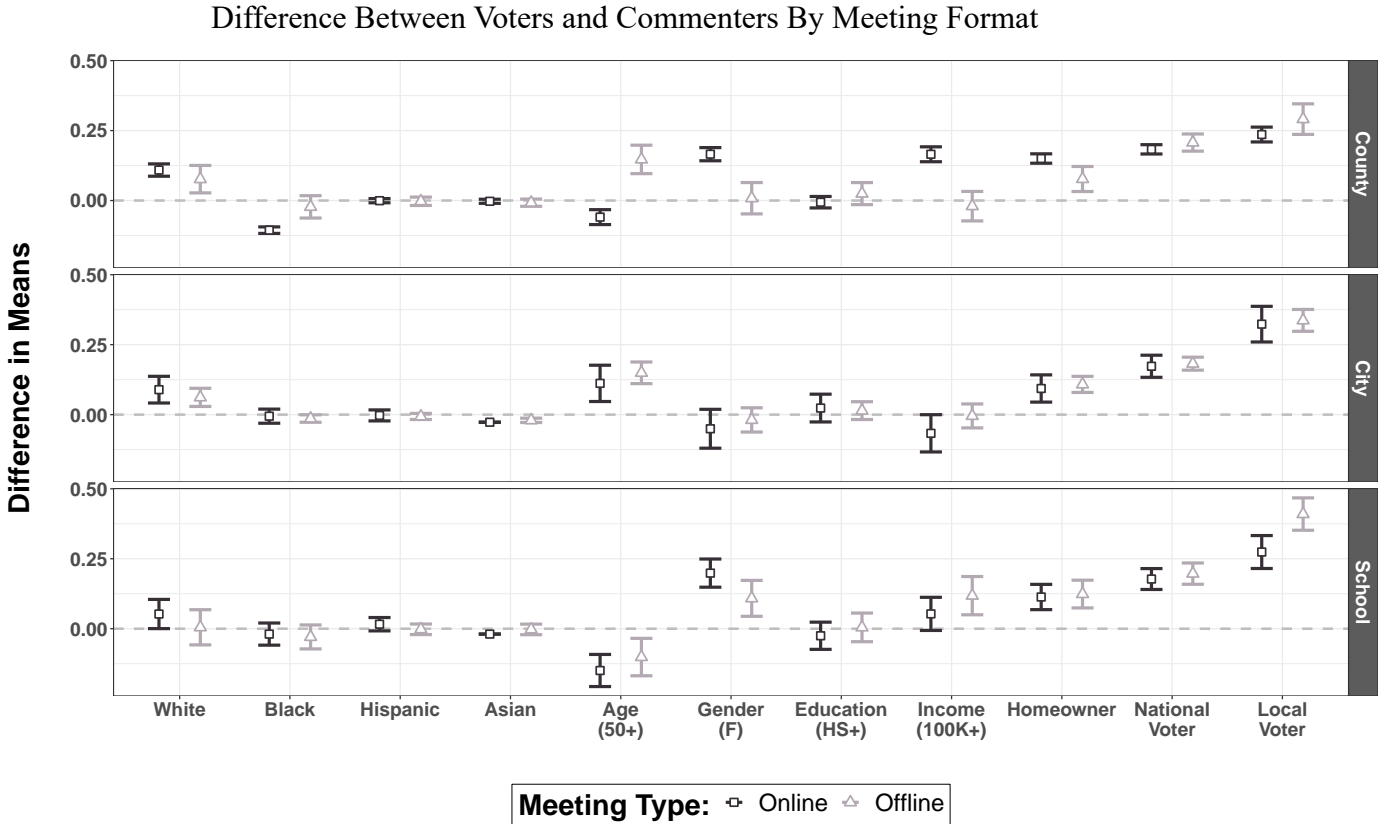


Figure 1.6: School and city comparisons are subset to their respective voters. Estimates are adjusted to account for uncertainty in the matching process. Simultaneous comparisons are available in the Appendix.

Overall, online commenters across all three local contexts are similar to their offline counterparts. They tend to be misrepresentative of their communities along dimensions of race, income, and political engagement. For the county, online meetings did attract significantly more women and younger commenters. Online meetings, in particular, appear responsible for the over-representation of women identified in my pooled analysis above. However, these benefits came at the cost of widened gaps in racial representation, income, and homeownership. In the case of cities, I document no noticeable change between online and offline commenters, with results similar to Einstein et al. (2022) examination of online meetings. Finally, for school boards, online meetings attracted

more individuals previously not engaged in local elections and potentially lessened the divide in income, albeit the estimate is not statistically distinguishable from that of offline commenters. Beyond these characteristics, I find no substantial differences between online and offline school board commenters.

Despite the theoretical benefits of online meetings, I find they had a mixed effect on the diversity of meeting participants. For the county and school boards in my sample, online meetings lessened some divides and incentivized an over-representation of traditionally underrepresented groups; however, many previous inequities in participation persisted within these meetings. Online meetings appeared even to exacerbate some disparities in the county's case. Notably, my approach above does not account for time or geographic varying confounders, which may alter specific demographics' propensity to comment. To partially address this concern, I utilize a logistic regression to examine differences between online and offline commenters simultaneously with added geographic controls. I present the results within the Appendix and find substantively similar results.

1.6 Discussion:

Online meetings were adopted out of necessity in response to the pandemic but offered a potential solution to historically low engagement in local political meetings. These meetings act as a crux for local democracy and allow individuals a direct connection to local public officials and policy-making processes. Given historic inequities in who attends and the resulting policy outcomes, increasing the quantity and diversity of meeting participants is vital to the healthy functioning of local democracy. Utilizing a novel dataset of meeting commenters, I find that online meetings are an effective tool to expand access to local governments, but in isolation, they are insufficient to address pre-existing inequities in who participates within local democracy. What's more, I find that inequities in participation previously found within municipal politics are present within both county and school board contexts. However, I find some silver lining for local democracy as some groups—namely women—appear to be overrepresented within some levels of local politics. This

finding suggests that some groups may effectively make up for underrepresentation in local politics by being overly active in other potentially more relevant local contexts. Future work should explore whether these potential increases in participation manifest into meaningful policy change for these often-underrepresented groups.

For policymakers, these results bring mixed news. On one hand, accessibility matters for local government and can boost the number of voices heard within local meetings. What's more these voices at least appear to be sincere. Section A.4 of Appendix I shows evidence that online participants share jurisdictionally relevant comments and, contrary to fears of trolling or increased incivility, do so with a similarly positive sentiment as their offline counterparts. However, improving access is no panacea for the deep-rooted inequities present within local politics. In fact, improving accessibility, when implemented alone, may do more harm than good by exacerbating existing inequities in political participation. To truly address these problems, policymakers likely need to find a way to empower individuals and their communities and improve their sense of political efficacy. If individuals do not believe their voice can impact local policy outcomes, allowing them to attend virtually will likely not persuade them to participate.

Chapter 2:

Is it All Good in the Neighborhood? How Partisanship May Shape Evaluations of Municipal Services

While extensive work has been done examining how voters evaluate the performance of federal and state governments, comparably less attention has been paid to evaluations in the context of local governments. In what work has been done, researchers often focused on small geographical areas and how information availability shaped evaluations (e.g., Berry and Howell 2007; Payson 2017). We still know relatively little about how voters' individual-level characteristics shape how they evaluate their local governments and their respective services. Nor do we know whether these evaluations correlate to the actual performance of local governments. Traditional conceptions of local politics might suggest that these evaluations are primarily a result of voters' day-to-day interactions with these services (e.g., Kaufmann 2004; Yoder 2020); however, this framework fails to take into account the growing role partisan identities play in voter's evaluations of government services at state and federal levels (Healy and Malhotra 2013). Furthermore, the increasing nationalization of politics means local political issues are no longer as insulated from traditional partisan discourse as previously conceived (Hopkins 2018).

In this article, I attempt to further our understanding of how voters evaluate their local government's performance and the role partisan identities play in constructing these evaluations. I argue that local services can become polarized when local issues become nationalized and distinct partisan cleavages form in the national discourse. For these polarized services, partisanship becomes a powerful identity through which individuals construct their local evaluations. This partisan lens then biases assessments of these services toward their party's average position regarding the service.

To test this argument, I use a nationally representative and geographically diverse survey that

asks individuals to grade two prominent local services, schools, and police. While both are primarily under the purview of local governments and receive significant political attention, they differ regarding their partisan division. Using administrative records from the Census Bureau and National Weather Service, I then match all respondents to their local governments. This matching allows me to, in addition to gathering respondent-level characteristics, (1) account for locality-based effects that may influence a respondent's evaluation and (2) assign each respondent an objective measure of both services to isolate any form of partisan bias.

Overall, I find a significant partisan bias in voters' evaluations of polarized services even after accounting for objective performance and prominent socio-economic features. Despite living in areas with similarly performing police agencies, Republicans systematically rate local police higher than Democrats and Independents. Notably, this bias is equivalent to a substantial change in objective outcome measures for police and the effects of income and race. While I find differences in how partisans evaluate local schools, the results are comparably small and outweighed by school performance. To test the robustness of my results, I subset my data set to only large cities to incorporate additional city-specific measures, alternative local services, and city and state levels of partisan control. I find support for my theory, and interestingly, my results suggest that the partisan bias is based on national partisan preferences rather than on party control of the government.

This research contributes to the growing literature on the role of partisanship within local politics and provides further evidence of the nationalization of American politics. Notably, my research is one of relatively few to map a geographically diverse collection of specifically local evaluations to their respective municipalities. My results highlight how local services, like national policy issues, can become polarized and susceptible to partisan biases. These partisan effects are especially problematic as they can influence an individual's evaluations more than a local service's objective performance. Additionally, these results have potential implications for the already low levels of accountability found with local politics, as voters may rely more on national opinion rather than local conditions when constructing their evaluations.

2.1 The Process of Polarization

I argue that some local services have become polarized due to the nationalization of American politics (Hopkins 2018) and the adoption of services as party-owned issues. Borrowing from both the polarization (e.g., Levendusky 2010; Sinclair 2006) and issue ownership literatures (e.g., Petrocik 1996; Egan 2013; Green, Palmquist, and Schickler 2004), I define a local service as polarized when parties hold ideologically distanced and distinctly homogeneous opinions regarding a service's intended purpose, appropriate support, and overall performance. Services can become polarized as the ideological distances of elite and media rhetoric surrounding a service increase (Prior 2013; Fiorina and Abrams 2008; Wilson, Parker, and Feinberg 2020). As more apparent partisan stances towards a service emerge, individuals' opinions of a service slowly converge to that of their party (Walgrave, Lefevere, and Nuytemans 2009; Walgrave, Lefevere, and Tresch 2012; Craig and Cossette 2020). A local service can become effectively polarized when this sorting becomes severe enough (i.e., there is little to no overlap in party positions). After polarization, individuals have apparent partisan affiliations and party positions to lean on when constructing their evaluations.

However, we know from past and recent research that not all aspects of local politics hold national significance or are inherently political (Anzia 2021). While some services and issues may intersect with national policy matters, such as education, law enforcement, and criminal justice, others remain distinctly local, like road management and zoning (Thompson 2020; Jensen et al. 2021; Sances 2021). As a result, we should not anticipate uniform polarization in local services. Instead, the process should occur on a service-by-service basis.

Furthermore, we understand that the extent to which any aspect of politics aligns with an individual's political identity depends on the salience of its respective issues areas (Bélanger and Meguid 2008). Local political outcomes, in particular, often require a catalyst for individuals to establish political connections and pay attention (Moore 2022; Nuamah and Ogorzalek 2021). Previous research on local accountability suggests that news coverage or a political spotlight on

a particular service is necessary for individuals to make political associations with that service (Payson 2017; Berry and Howell 2007; de Benedictis-Kessner 2022). Therefore, services with broad relevance to many individuals and receiving sufficient media and political attention should be the most likely to become polarized.

2.1.1 How Partisanship Shapes Evaluations

A fundamental assumption of my argument is that polarization and party identity shape individuals' evaluations of government performance. We know individuals craft their political opinions through their party identification (Campbell 1980; Zaller, John 1992) and that given the increasing alignment of ideological and policy preferences, an individual's overall evaluation tends to correspond to either that of their party's elites or their party's average position (Lenz 2009; Sinclair 2006). Past research in retrospective voting provides greater insight into why this is and its mechanisms. First, when individuals interpret policy outcomes, they must decide how to allocate blame or credit. Voters tend to attribute positive outcomes to their own party's performance and blame adverse outcomes on the other party (Healy and Malhotra 2013). In the context of a local service, such as police, which may not belong to a specific party, Republican individuals may associate them more with their party. Thus, any negative performance on their behalf may be attributed to other political actors.

Second, partisanship can be used as a powerful heuristic when it is unclear how to interpret policy outcomes or establish lines of responsibility. In a study examining retrospective voting in the wake of Hurricane Katrina, Malhotra and Kuo (2008) find that individuals were more likely to blame Republican officials for the deaths and damages resulting from the storm. However, these effects disappeared when respondents received the Republican official's title. There is also evidence that when information environments become polarized, individuals are more likely to adopt partisan cues even in the presence of contradictory objective information (Druckman, Peterson, and Slothuus 2013). Thus, when tasked with evaluating police performance, individuals may rely more heavily on their partisan identity than objective performance.

2.1.2 Why Schools and Police

Ideally, to analyze how local services become polarized, I could use longitudinal data on individual evaluations of various local services. Then, as services become polarized in the national sphere, I could measure the divergence between partisans. Unfortunately, no such data is sufficiently detailed to ask about local services or cover an extended time frame. An alternative way to measure polarization's effects on a service is to examine evaluations between an arguably polarized and non-polarized service and compare their differences. I argue that local schools and police are good candidates for this comparison.

First, schools and police are predominantly administered and funded by local and county-level governments. Most public schools are governed primarily through school districts, which derive over half of their total funding through local property taxes (Urban Institute 2020). School districts do have to adhere to state and federal guidelines to secure additional financing, but hiring practices, setting the teaching curriculum, modifying school boundaries, and adjusting property taxes are primarily at their discretion.¹ While police agencies exist at almost all governmental levels, over 80% of agencies are supervised directly by municipal governments. These local police agencies also employ roughly 67% of all full-time sworn officers (Bureau of Justice Statistics 2019). Much as is the case for school districts, police agencies must adhere to state and federal guidelines to secure outside grants. Still, most of their administrative, procedural, and hiring practices are left to their governing municipalities' discretion.

Second, local schools and police agencies are both politically salient and widely relevant services. Schools and police receive relatively consistent coverage from local and national news outlets and are often addressed by political elites. Beyond coverage, both services also play a visible socio-economic role within their communities. Each service plays a crucial role in their community (i.e., through direct educational services or public safety) and has downstream effects such as impacts on

¹I should note that while districts are allowed to develop their curriculums, federal programs such as Common Core, require districts meet specific educational standards. These requirements often mean many districts teach towards these standards to maintain or increase funding levels.

property values or income inequality. Thus, these services are likely already politically salient, and individuals likely have already formed some political opinions regarding them.

Third, schools and police differ regarding the degree of polarization surrounding them. In the wake of highly publicized police violence against racial minorities, opinions toward police have become increasingly partisan and differentiated. Republican party figures have come out primarily in support of police agencies, whereas Democrats have taken a far more critical viewpoint. Recent surveys have highlighted how partisans fundamentally have different views over police agencies' roles within their communities. Democrats are more likely to view police officers as enforcers, whereas Republican individuals tend to view police more as protectors (Pew Research Center 2017). While schools and education are not free from partisan division, parties are not as clearly divided. Republicans and Democrats differ along dimensions of social values and federal oversight but are similar along other dimensions, such as access to education and accountability within schooling (Jensen et al. 2021). Unlike the case for police agencies, there does not appear to be a clear partisan divide over all dimensions of schooling. Even amongst the most divisive aspects of schooling (e.g., sex education and prayer within schools), partisan differences are relatively small compared to other issues, such as abortion or gun rights (Shapiro et al. 2021). This difference in partisan polarization allows for a practical test of my proposed theory. If polarization does shape local evaluations, I should find that Republicans systematically evaluate police higher than Democrats. Given the lack of clear partisan sorting surrounding schools, I should see limited differences between partisans.

2.2 Data and Methods

Overall, I argue that local services have become polarized due to the nationalization of local issues. My goal is to highlight this phenomenon through a cross-sectional comparison of local school and police evaluations. If my theory is supported, I should find that voters are biased towards their party's position in their assessments of police, but not schools. To empirically test this difference, I rely on nationally representative public opinion data and school and police performance measures. I

also use voter's geographical information to account for municipal-level influences in their service evaluations.

2.2.1 Local Opinions and Individual Characteristics:

One of the primary challenges associated with measuring opinions at a local level is a lack of geographically diverse survey(s) surrounding predominantly local issues. However, the 2018 Cooperative Congressional Election Study (CCES) fielded several questions asking over 48,245 individuals to evaluate the local services within their surrounding communities.² Specifically, the survey asked respondents to grade their local communities regarding schools, policing, roads, and zoning. For this study, I focus on how respondents rated their local school districts and police agencies. Respondents assigned grades on a five-point scale ranging from -2, for poor, to 2, for excellent, with 0 representing an average evaluation.³ To account for each respondent's partisan identity, I collapsed a 7-point party ID into a 3-point party ID and recorded it with two binary indicators representing Republican and Independent. I use this approach rather than the traditional 7-point scale to account for the possibility that partisanship has a non-linear effect. I separate independents as a growing body of work has suggested that independents as an identity are differentiated from traditional partisans. Thus, their identity may uniquely shape local evaluations. In the Appendix, I replicate the primary results using a conventional 7-point party ID and find no substantive differences.

In addition to these evaluations, I also take several steps to account for other socio-economic characteristics that may influence a respondent's interaction with local goods and their partisan identity. Previous works, such as Trounstein (2010), found that local services and infrastructure are developed strategically along race and income.⁴ Many of these identities are also highly

²This sample size constitutes the total number of respondents who answered the question(s) of interest and the required demographic information below.

³In the original survey, respondents gave an alphabetic rating ranging from A to F. I standardized this rating to a -2 to 2 scale to ease interpretability and modeling

⁴See Einstein, Palmer, and Glick (2019), or Anzia and Moe (2014) for additional examples of this historical trend within local politics.

correlated with party ID. Therefore, to address concerns of potential confounding, I first include binary indicators for respondents' gender, race, and income. Second, I add indicators for both parenthood and homeownership, as these variables may moderate an individual's interactions with schools and police.

Finally, I attempt to account for geographic differences between partisans. Recent work has shown that conservative individuals live farther away from city centers and in more rural areas than their liberal counterparts (Gimpel et al. 2020). These decisions drastically shape each individual's interactions with both schools and police and the forms of the services themselves (Walsh 2012). To account for this potential problem, I include indicators for self-reported urban/rural residency, which I report as Rural. While this indicator may not capture the nuance in the strategic settlement, it attempts to address the geographical and partisan confounding. Descriptive statistics for these controls are available in the Appendix.

2.2.2 Measuring School and Police Performance

One standard measure of a school district's performance is its standardized test scores. Since 2010, school districts in all fifty states must report their Common Core test performance to state-level education departments. These scores are often widely accessible online and frequently picked up by media outlets. Previous work has found that voters use these scores to reward or punish incumbents in school board elections (Berry and Howell 2007; Payson 2017). Using the popular service, SchoolDigger, I aggregate the 2018 performance of all school districts covered by my CCES sample. I standardized all test scores to an increasing scale from 0 to 5 relative to their percentile performance within their respective state.⁵

Given these scores, however, I still need to correctly assign respondents to their respective school districts. Cities vary immensely regarding the number and layout of their school districts. Some smaller towns share a school district with other municipalities in their county, while others

⁵For more information regarding these scores, I direct the reader to SchoolDigger.com's methodology section

contain several interlocked and competing districts. These district lines often don't follow municipal or other geographical boundaries. To address this issue, I take the average test performance of all school districts within an individual's zip code. This process yields a single score, *School Rating*, which reflects the general performance of schools within each respondent's immediate area. The average school rating for each respondent was 2.651, with a standard deviation of 0.86. The next challenge I faced was finding a measure for local police agencies.

There is a significant debate in sociology and economics over how best to quantify police performance. One prominent way the public may evaluate their local police agencies is along the lines of crime prevention. While fine-grain crime data is available for a subset of larger cities, respondents within my sample reside in towns with populations ranging from under 1,000 to over 2 million. Thus, to include objective measures for most of my sample, I use the 2018 county-level violent and property crime rates provided by the Federal Bureau of Investigation's Uniform Crime Reporting Program (UCR). Specifically, I aggregate the rates of murder, nonnegligent manslaughter, rape, robbery, and aggravated assault and record them as one variable *Violent Crime Rate*. Similarly, I aggregate rates of burglary, larceny, and motor vehicle theft into one measure of *Property Crime Rate*. Both measures are standardized to incidents per 100,000 individuals to scale relative crime rates between areas of varying populations. I chose to use the current year's crime rate as past work in retrospective voting found that individuals tend to be myopic and overweight current conditions over long-term change (e.g., Achen and Bartels 2016).⁶ By using county-level measurements, I notably risk losing within and between city-level variation in crime rates. Additionally, county-level rates may misconstrue crime rates for unincorporated cities compared to their surrounding counties. Despite these potential pitfalls, I still chose to use these measures due to their coverage of my sample.

Beyond geographic availability, these scores could be more problematic for other reasons. First,

⁶I should note that almost all of this previous work examined retrospective voting primarily in an economic sense. There is a chance that crime rates differ from other economic conditions, and voters interpret them differently. To investigate this, I test several alternative specifications of crime rates, such as using a lagged measure of violent and property crimes and the annual difference in county crime rates. The results are available in the Appendix.

traditional crime rates are highly sensitive to reporting practices. These measures can depend just as much on the number of police officers in an agency or an area's developed trust in its police force as the amount of crime (Fielding and Innes 2006). Second, these measures only capture one dimension of police work: crime prevention. A significant portion of what police agencies do is to address non-criminal social problems surrounding mental health, poverty, and addiction; however, this work is almost entirely uncaptured by crime-based metrics (Hodgkinson, Caputo, and McIntyre 2019). Despite these problems, public officials and media outlets often pick up and publicize crime rates. Thus, individuals are likely to use these scores to inform their evaluations of their local police. While I acknowledge these scores are flawed and capture a minimal scope of police performance, they are available for the entirety of my sample and are often publicly salient.

2.3 Full Sample Results

Federal, state, and local policy decisions all influence school districts and police agencies. To capture this process and test my theory, I employ a series of methodological tests examining which factors influence an individual's evaluations and the mechanism through which any partisan bias may emerge. The section proceeds as follows. First, using my entire dataset, I test the core of my theory and examine whether a partisan bias exists in individuals' evaluations of local police compared to schools. Then, I subset my data to only large cities to replicate my results with more city-specific measures and alternative modeling strategies. Next, I examine whether my choice of comparison is robust by seeing if the pattern holds for another prominent local service, roads. Finally, by including mayoral and gubernatorial partisanship measures, I extend my work to see whether the partisan bias is one of elite valence or party positioning. Overall, I find general support for my theory of local polarization.

2.3.1 Polarized Evaluations

To model partisanship's effect on local perceptions of schools and police, I use a simple linear model with state-fixed effects and clustered robust standard errors. I include all controls mentioned in the data section within the modeling process. For the analysis below, I limit the scope of my analysis to only Democrats and Republicans for ease of interpretability.⁷ Table 2.1 presents the results of my linear model. Results are a subset of relevant covariates, with the detailed regression results available in the Appendix. From the first column of Table 1, we can see that the coefficient surrounding the Republican party ID is both positive and significant at the $p < 0.05$ level. This effect appears robust to the inclusion of both objective measures of police performance and other salient identities, such as race. Identifying as Republican translates to a one-fourth of a standard deviation increase in an individual's evaluation of local police. To contextualize this effect, it is the equivalent of moving from a county with approximately 90 violent crimes per 100,000 individuals to one with a rate of over 2735. This finding suggests that, for highly polarized local services, partisan biases can vastly outweigh differences in objective performance. Notably, the magnitude of this effect is also similar to that of race, an identity traditionally thought to dominate local politics and evaluations of police. Thus, in the context of a polarized local services, party ID appears to become a salient and influential identity.

⁷Within the Appendix, I replicate the results, including independents, and find no substantive differences in the results.

Table 2.1: OLS Regression of Local School and Police Evaluations on Partisan Identity

	<i>Service Evaluation:</i>	
	Police	School
Republican	0.264* (0.018)	-0.035 (0.028)
Property Crime Rate	0.00000 (0.00003)	
Violent Crime Rate	-0.0004* (0.0001)	
School Rating		0.340* (0.014)
Age	0.007* (0.0005)	0.0002 (0.001)
Gender (M)	0.004 (0.010)	0.034* (0.011)
Black	-0.282* (0.033)	-0.016 (0.035)
Income	0.025* (0.002)	0.016* (0.002)
State Fixed Effects	✓	✓
Additional Controls	✓	✓
Observations	29,020	34,812
R ²	0.078	0.103
Adjusted R ²	0.076	0.101

Note: The above table presents the results of a OLS regression with state fixed effects and clustered standard errors. Crime rates are standardized to incident per 100,000 citizens. Full regression results can be found within the Appendix. Standard errors are presented in parentheses. * $p < 0.05$

For my theory of polarized services to be fully supported, we should see that party ID has a smaller effect in the context of school evaluations. Examining the second column, we can see that my theory is supported. The coefficient surrounding the Republican party ID appears negative but is not significant at the $p < 0.05$ level. Unlike in the case of police above, how well an individual's local schools perform on standardized tests appears to be the most influential driver of their evaluation.

Thus, individuals rely more on the service's objective performance and relevant identities for non-polarized local goods rather than their partisan identity when forming their evaluations. Overall, these results appear to support my theory of local polarization; however, the relatively small effects of violent crime rates and lack of effect associated with property crime rates may lead some to worry about my use of county measures or the mechanisms behind the partisan bias. I aim to address some of these concerns in the following sections.

2.4 A Closer Analysis of Large Cities

While the complete CCES data set provides a geographically diverse sample to test my hypothesis, my measurements for local services must be inherently broad and limited. It is possible that my choices in service measures do not accurately reflect individuals' experiences or that they rely on alternative metrics to base their evaluations. However, finding detailed estimates for municipal services is difficult, and these measures often do not exist for small cities. To address this concern, I zoom in on a subset of large cities from my original dataset to gather additional measures of these local services, explore alternative political mechanisms, and compare against salient local services.

This subset consists of all cities for which I have at least 65 respondents and effectively limits my examination to medium and large cities, with the smallest city of the sample, Pensacola, FL, having a population of over 52,000.⁸ In total, my new dataset leaves me with 10,613 survey respondents from 72 unique cities.⁹ Figure 2.1 displays the cities covered with my new large cities subsample compared to my original sample. Importantly, I acknowledge that what I gain in terms of data granularity, I may lose in the generalizability of results.

⁸In the Appendix, I compare all covariate measures between the original CCES data set and my large cities sample.

⁹In total this represents roughly 10% of all cities with a population over 50,000

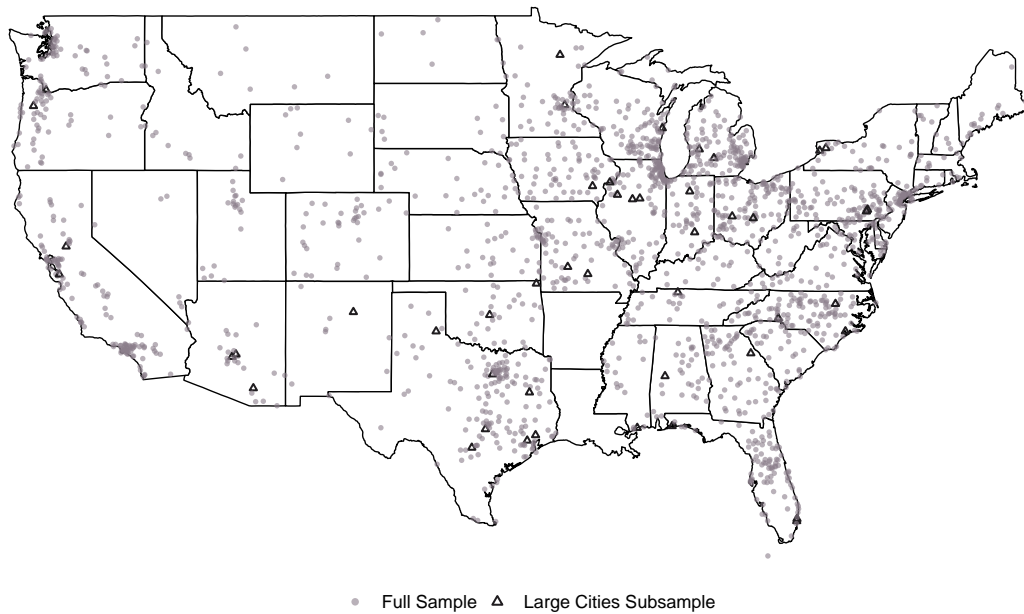


Figure 2.1: Cities included within my full sample (2,216) and large cities subsample (72).

As cities grow, the size and scope of both their internal governments and local services fundamentally change (Oliver 2001; Oliver, Ha, and Callen 2012). Individuals within sparse rural areas such as Carlinville, IL (4,000) may interact inherently differently with local government and their services than individuals from a large urban center such as Chicago, IL (2.75 million) (Moore 2022; de Benedictis-Kessner 2022; Lassen and Serritzlew 2011). Moreover, from the large body of work examining the urban/rural divide within the U.S., an individual’s partisan identity may also vary across these local contexts (Gimpel and Reeves 2023; Nemerever 2021; Scala and Johnson 2017). Thus, my results in the following sections may not map perfectly to these smaller American democracies. However, given that I identify clear trends within my larger sample and nearly 40% of Americans still reside within similarly large cities, my proceeding analyses should provide valuable insight into mechanisms of local polarization and the relationship between local

services and individuals' partisan identities.

2.4.1 Reexamining Schools and Police

For each of the 72 cities in my subset, I collected the reported 2018 violent and property crime rates reported by the city to the UCR program. While these measures help to alleviate concerns associated with county-level estimates, they still suffer from problems associated with using crime rates to gauge police performance. To supplement these rates, I also include the number of employees each city employs within its police department. I standardize these measures to the number of employees per 100,000 citizens to compare across municipalities of varying sizes. While the number of employees is not directly correlated to a city's ability to combat crime, it serves as a proxy for city-level spending on policing. For school performance, I use the identical standardized test scores used in the primary analysis.

One of the critical problems present in my primary analysis above is that an individual's partisanship is not assigned randomly. One's partisan identity is closely correlated with various other identities influencing access to local services and selective residency. In an effort to combat concerns of endogenous confounding, I employ a procedure of covariate balanced propensity score weighting (CBPSW) as presented in Imai and Ratkovic (2014). To improve matching, I match respondents on each of the previous controls included in the initial analysis in addition to city-specific objective performance measures. Using these weights, I then regress the effect of partisanship on evaluations of both local schools and police using both random state and municipality effects and present the results in Table 2.2. Due to the variability associated with choosing weights, I opt to bootstrap the standard errors for my estimates. I recalculate the matches and fit the hierarchical model for each bootstrapped sample.

Table 2.2: Large City Analysis Using CBPS Weighting of Partisanship on Local Evaluations

	<i>Service Evaluation:</i>	
	School	Police
Republican	−0.043 (0.027)	0.262* (0.029)
Constant	0.240* (0.063)	0.345* (0.042)
Observations	7,591	4,956
AIC	24,304.160	14,917.710

Note: The above table presents the results of a CBPS Weighting with bootstrapped standard errors for my large cities sample. Bootstrapped standard errors are presented in parentheses. * $p < 0.05$

From the table above, the coefficients surrounding *Republican* are again statistically significant at a $p < 0.05$ level for police, while a similar pattern is not seen for schools. Additionally, the effect of partisanship in the case of police appears to be similar to my initial analysis despite the inclusion of additional city-specific measures. I should note, however, that CBPSW does not completely solve all concerns of endogeneity as the process is still sensitive to large misspecifications of the modeling process. Accounting for all possible confounders associated with one's partisan identity and interaction with local services is nearly impossible for this form of observational study; however, this procedure provides more probable estimates than those generated by traditional forms of regression alone.

2.4.2 An Alternative to Schools

There may be some concern that schools do not provide an adequate contrast to test my theory of polarization. Given trends following nationwide shutdowns in response to the COVID-19 pandemic, some may argue that schools are far more polarized than I posit earlier in this paper. Luckily, given my focus on larger cities, I can examine another prominent local service: roads. While roads may receive less media or political attention, they are a salient service most individuals interact with to

some degree on a daily basis. One objective measure of roads is traffic congestion. Using historic traffic data provided by TomTom, a popular transportation and navigation service, I assigned each city an index measure for their average level of traffic for the year 2018. *Traffic Index* is measured on a 0 to 100 scale and calculated— through GPS data—by measuring the average additional travel time traffic congestion adds to a standard 30-minute trip within a given city. For ease of calculation and interpretability, I employ a hierarchical linear regression that includes state- and city-level random effects and display my results in the table below.

Table 2.3: Hierarchical Linear Regression of Local Roads Evaluation on Partisan Identity: Full Results

	<i>Service Evaluation:</i>
	Roads
Republican	0.030 (0.026)
Traffic Index	−0.003 (0.003)
Age	−0.005* (0.001)
Gender (M)	0.020 (0.022)
Black	−0.175* (0.033)
Income	0.016* (0.004)
Constant	−0.003 (0.077)
Observations	6,928
State Random Effects	✓
Municipality Random Effects	✓
Additional Controls	✓
AIC	18,191.980

Note: The above table presents the results of a hierarchical linear regression with random effects. All respondents are nested within municipalities within states. Full regression results can be found within the appendix. Standard errors are presented in parentheses. * $p < 0.05$

The results from Table 3 above suggest that one's partisan identity does not appear to play a significant role in their evaluations of local roads. However, the average traffic conditions do not significantly affect an individual's evaluations either. This finding could be because individuals use alternative cues, such as pavement conditions, to measure road quality. Alternatively, individuals may not have a conception of the annual performance of roads and instead base their evaluations on road conditions immediately surrounding when the survey was fielded. If this were the case,

annual measures of road conditions may be poor objective performance measures for local road quality. Despite this finding, the results surrounding partisanship support my theory of polarized local services. Additionally, these results align with previous work examining road conditions, such as Arceneaux (2005) and de Benedictis-Kessner (2018), which find an inconsistent correlation between road and traffic conditions and perceived mayoral performance.

2.4.3 Party Position Taking or Elite Valence?

While the results thus far have supported my claim that for polarized services, individuals' partisan identities bias their evaluations, it remains to be seen as the exact mechanism behind this bias. As mentioned previously in this paper, this bias could be individuals adopting their national party's stance towards a service (i.e., favorability in the case of policing); however, it could also be caused by some form of partisan penalty/reward depending on which party holds political control. To examine this possibility, I borrow from de Benedictis-Kessner and Warshaw (2020a) analysis of the effect of mayoral partisanship on municipal spending to record the partisanship of each city's mayor and governor as of 2018 for each city within my large city subsample. I opt to control for both state and local partisan control due to the potentially blurred lines of responsibility for some local goods. While work such as Arceneaux (2006) suggests that individuals can make meaningful distinctions between levels of government, they also highlight that this ability depends on an individual's perceived responsibilities of local, state, and federal governments (Arceneaux 2005). The growing trend of nationalization and politicization may lead individuals to view state and federal governments as increasingly responsible for some local policy areas. For the analysis, I employ a similar hierarchical regression as before and present the results in Table 2.4.

Table 2.4: Hierarchical Linear Regression of State and Local Co-partisanship on Local Service Evaluations

	<i>Service Evaluation:</i>		
	Police	Roads	Schools
Co-partisan Mayor	−0.038 (0.034)	0.013 (0.032)	0.042 (0.030)
Co-partisan Governor	0.015 (0.031)	0.045 (0.025)	0.052 (0.028)
Republican	0.246* (0.038)	0.023 (0.033)	−0.065 (0.033)
Constant	−0.057 (0.122)	−0.021 (0.081)	−0.630* (0.085)
Observations	4,750	6,846	6,908
AIC	12,635.960	18,002.590	18,953.690
State Random Effects	✓	✓	✓
Municipality Random Effects	✓	✓	✓
Objective Measures	✓	✓	✓
Additional Controls	✓	✓	✓

Note: The above table presents the results of a hierarchical regression for my large cities sample. Respondents are nested within cities and then states. Co-partisanship was determined by whether the respondent shared the party of the acting Governor or Mayor in 2018. Standard errors are presented in the parentheses. * $p < 0.05$

In the case of all three local services, co-partisanship with either the city’s mayor or governor does not significantly affect one’s evaluations of local services. Even in the case of police, an individual’s evaluations appear to be based far more on their partisan identity rather than their shared partisanship with local officials. While these null results don’t necessarily mean shared partisanship does not affect local evaluations, for polarized services, the party’s average position matters far more. These null results may be partially explained by the lack of party competition often found within municipal governments (Bucchianeri 2020).

2.5 Conclusion

At the beginning of this article, I argued that individuals' evaluations of municipal services are not necessarily based on the quality of the service or their access to it. Instead, some local services have become polarized, leading to partisan distortions of evaluations in the direction of party preferences. Through an analysis of local police and schools, I provided strong correlational evidence that the polarized discourse surrounding police has led Republicans to evaluate the service more positively than other outpartisans. Additionally, I showed how a similar process is absent in non-polarized cases of local schools and local roads.

My work adds to our understanding of retrospective evaluations within local politics in several ways. First, it highlights how increased nationalization and partisan sorting may inhibit retrospective voting within a local context. Second, the data used in my analysis departs from the traditionally geographically narrow approaches used to examine local politics and expands my analysis to a national sample. Lastly, it adds to our broader understanding of how polarization shapes electoral accountability. As individuals become increasingly sorted along party lines, my work suggests that retrospective evaluations may become less and less likely at all levels of government.

These findings suggest that partisanship may play a far more prominent role in the realm of local politics than previously conceived. For policymakers, this conclusion is potentially problematic as individuals may begin holding local services accountable for national opinions rather than their objective performance. By doing so, individuals may incentivize policymakers to prioritize polarizing aspects of local services at the potential detriment of the service as a whole. This analysis of local evaluations raises several vital questions surrounding local accountability and responsiveness. Notably, for two of the three examined services, objective measures of the said service failed to impact an individual's evaluation significantly. Future work should examine how individuals construct these perceptions of local services and which measures impact which group's evaluations. Additionally, while I've shown that partisanship may shape local assessments, it remains unclear whether this translates into any policy or electoral change. Examining how local

politics' nationalization and potential polarization affect municipal policy-making and spending should be an essential focus of future work.

Chapter 3:

Utilizing Large Language Models To Transform Messy Text-Based Sources In Useful Data

Often, political scientists are tasked with transforming text-based primary sources into quantitative datasets. Sources such as newspapers, legal records, and transcripts contain a myriad of valuable information ranging from simple names and counts to specific legislation and institutional features. These data form the backbone of many political datasets, but extracting said data has historically been laborious and time-consuming. Texts are often recorded in formats with human readers in mind, making automated approaches prohibitively challenging to apply to the wide range of styles in which text-based sources may appear. Given this challenge, most researchers extract their desired information by hand or use crowdsourcing to expedite the process (Sumner, Farris, and Holman 2020; Bohannon 2011). However, these solutions come at the cost of time for hand coding and money for crowdsourcing, which is especially problematic if researchers wish to rely on accurate information and require multiple crowdsources to compare results (Benoit et al. 2016; Carlson and Montgomery 2017). These costs force many researchers to limit their focus to a subset of features and texts that either share commonalities in formats or are simply manageable to transform. By doing so, researchers sacrifice potential substantive coverage and cannot update datasets to reflect changing trends.¹ Moreover, after the information is extracted from these sources, researchers are locked in and can only gather new data (even if it's within the same text-based sources) if they repeat the expensive process.

To alleviate this problem throughout political science, I propose a human-in-the-loop pipeline that leverages recent advances in Large Language Models (LLMs) to crowdsource and transform

¹For example, de Benedictis-Kessner and Warshaw (2016) highlight this exact problem for the field of local politics.

text-based sources into quantifiable datasets artificially. At its core, the process involves the iterative construction of instructions and context for a model and then the model's implementation of them to transform text into usable data efficiently. Researchers hand code a selection of their texts and then use it to measure LLM performance and adjust the contextual settings provided to the model. The pipeline relies on the foundational nature of LLMs (Bommasani et al. 2022), which allows them to complete various tasks outside (Brown et al. 2020; Radford et al. 2019) and within political science (Argyle et al. 2023; Häffner et al. 2023; Laurer et al. 2023; Wang 2023; Ornstein, Blasingame, and Truscott, n.d.). Importantly, LLMs are far faster than human coders and can cost an order of magnitude less. However, LLMs are imperfect and can suffer inaccuracy or hallucinations wherein the model creates data that is not present within the text (Bommasani et al. 2022; Bender et al. 2021; Karpinska and Iyyer 2023). These complications can make it difficult for researchers to gauge whether these models produce accurate data or whether false or unreliable results contaminate the final dataset. The iterative nature of the pipeline works to address this concern by providing researchers with a validation of the model's performance and allows them to find and understand potential points of friction (Grimmer and Stewart 2013). By leveraging both the researcher's judgment in the data quality process and the speed and adaptability of LLMs, researchers can more efficiently transform a greater variety of text-based sources into valuable datasets.

I demonstrate the pipeline with two examples from local and judicial politics. Both examples involve extracting meaningful information from text-based government and legal documents, a task common throughout numerous fields of political science. For each case, I hand-coded a small sample of documents and used them to train a model to transform the corpus into a quantitative dataset. I then compare the model's performance with expert coders assigned the same task. Overall, I find that LLMs produce results comparable to professional coders and, in a process, far more technically accessible than other machine learning solutions. Crucially, after a model was iteratively tuned to the task, LLMs could accomplish the coding tasks in mere hours rather than the days or weeks required to transform the data by hand.

In summary, this pipeline has the potential to significantly enhance the process of transforming and coding text-based sources. While it may not entirely eliminate the need for manual coding, it can considerably reduce the manual labor required to produce high-quality data. Additionally, it provides researchers with an adaptable framework for text transformation. By incorporating additional manually coded data, LLMs can quickly capture data that researchers may have missed or overlooked during the initial coding process. Beyond transforming datasets, this process contributes to the broader push for transparency in political science research by allowing researchers to clarify their data processing methods, facilitating replication and expansion of their work (Gaikwad and Herrera 2023).

The paper proceeds as follows. First, I expand on the difficulties of transforming text into data and why automating it has historically been challenging. Second, I provide additional information regarding LLMs, including what they are, how they are trained, and why they are perfect for this specific application. Third, I outline my pipeline, including best practices and example implementations. Fourth, I demonstrate the model with two unique case studies and conclude with potential applications and pitfalls of using LLMs as artificial crowd coders.

3.1 Why is Transforming Text-Based Sources Difficult?

Transforming text-based sources into quantifiable datasets is a deceptively simple problem. Our desired data is often contained directly within texts, and we only need to locate it, extract it, and format it into our final dataset. However, this process is often more complex than we might initially think. For example, if we want to extract all interest groups who sign an amicus curiae brief to the Supreme Court, much like Box-Steffensmeier and Christenson (2012). We need to locate the section of the brief that lists the interest groups that signed the brief, which may vary on a brief-by-brief basis. Next, we need only to extract the interest who signed the brief, not their legal counsel or other mentioned groups. Finally, we must store those interest group names in our desired data format. For researchers who either hand-code the data or utilize crowd-sourcing, this process is manageable as

they can easily see the organizations in the appropriate section of the brief and extract their names. Ideally, each brief would then be coded multiple times to ensure coders accurately extracted all interest groups. However, this process can be tedious, time-consuming, and potentially expensive, especially as the number of briefs grows.

In most cases, researchers have limited resources and may only code each document once, often by a singular coder, thus forgoing many benefits of crowd-sourcing data (Benoit et al. 2016). Hand coding and crowd-sourcing text-based sources have been the backbone of political science research. Still, their associated costs mean researchers are often limited in the amount and variety of data they reasonably process. Alternatively, researchers could turn to computer-assisted methods to transform text-based sources, but these methods bring additional challenges.

Transforming text-based sources into quantifiable data through computer-assisted means is far from a novel problem. Within political science alone, approaches such as regular expressions, named entity recognition (NER) (Kerkvliet, Kamps, and Marx 2020; Jalal et al. 2020), and optical character recognition (OCR) (Permaloff and Grafton 1992) have long been utilized to identify and extract meaningful information from text, including names, dates, locations, and amounts from text. In theory, each of these methods can efficiently transform many text-based sources present throughout political science but often require extensive fine-tuning and adjustments to transform data correctly. While researchers could lessen these burdens by utilizing out-of-the-box models or fine-tuning pre-trained models to extract their desired information or fine-tuned pre-trained models (Laurer et al. 2023), producing an automated process that works for a large variety of texts can be exceedingly challenging.

The core problem with transforming text-based sources is that the information we desire may take a myriad of forms, making it difficult to process automatically. While many individual text-based sources may follow consistent formatting, when researchers look across units (i.e., across different sources of texts), formats can vary wildly. This heterogeneity in how information is stored is not limited to between-unit examinations but is also present within units. Recording practices

may change over the years as maintainers of these records adopt new services or standards.² This problem is especially present when researchers examine longitudinal sets of texts where reporting practices may change over time, even if the core information remains present throughout. Figure 3.1 displays an illustrative example of formatting discrepancies within the precinct-level municipal election returns from the city of St. Louis and the county of St. Louis. Both returns record the same electoral information regarding the contest and vote breakdown but in starkly different formats.³ Each format could require a novel automated method or special modification to transform accurately. Even if automated methods successfully transform and extract a researcher's desired data, they may fail to consider the context of the information.

²A common example of this is with municipal records, which can vary depending on who is the municipal clerk.

³Precinct level electoral returns, in particular, have proven exceptionally difficult to gather and quantify. Even recent groundbreaking efforts such as de Benedictis-Kessner et al. (2023)'s local election database or previous advances such as Marschall, Shah, and Ruhil (2011)'s LEAP only cover a sample of all electoral races and results in the United States.

WD28 ALDERMAN DEM					
	Reg. Voters	Times Counted	Total Votes	LYDA KREWSON	WILLIAM HAAS
PROVISIONAL	590	0	0	0	0
HOUSE VOTES	590	0	0	0	0
Total	590	112	112	100 89.29%	12 10.71%
W 28 P 02					
Polling	490	69	69	67 97.10%	2 2.90%
ABSENTEE	490	12	11	9 81.82%	2 18.18%
PROVISIONAL	490	0	0	0	0
HOUSE VOTES	490	0	0	0	0
Total	490	81	80	76 95.00%	4 5.00%
W 28 P 04					
Polling	541	62	62	58 93.55%	4 6.45%
ABSENTEE	541	9	9	9 100.00%	0 0.00%
PROVISIONAL	541	0	0	0	0
HOUSE VOTES	541	0	0	0	0

(A) St. Louis City

			VOTES		PERCENT		WITH 6 OF 6		REPORTING	
ALDERPERSON BALLWIN WARD 1										
(Vote for) 1										
01 = JIMMY TERBROCK			408		98.31					
02 = INVALID WRITE-IN			7		1.69					

			01		02					

1209 LAF9			154		1					
1213 LAF13			17		0					
1218 LAF18			98		3					
1219 LAF19			58		2					
2815 WH15			37		1					
2832 WH32,38,46			44		0					
=====										

(B) St. Louis County

Figure 3.1: Snapshots of official precinct level results from the 2015 General Municipal elections. Both examples cover separate aldermanic races occurring that year.

Often, texts include a plethora of information that may not be relevant to our research question. We may only be interested in information when it pertains to specific concepts or is related to particular entities. For example, we may be interested in the kinds of citizens who participate in public town hall meetings and the types of comments they make. A possible text-based source to examine this question would be transcripts of the meetings from the Local View dataset (Barari and Simko 2023). We would want to extract each commenter's name and comment for each meeting. However, citizens are not the only individuals to participate in these meetings, and we would want to filter out public officials who may be leading the discussion. Citizens may also talk in groups or multiple times throughout a meeting, and we would like to associate each citizen with all the comments they make. Human coders could, albeit tediously, accomplish this task as they could

understand both the context as to whether an individual was a citizen and whether each transcribed portion of the meeting belonged to a specific individual. In the case of automated approaches, this task can quickly grow in complexity, as even if automated approaches can identify comments and commenters, researchers would be forced to devise a schema to filter out unwanted participates and to associate comments with individual commenters.⁴ The need for context around information need not even be so complex. It can be as simple as only wanting specific dates or monetary values from texts, both of which might require some level of contextual awareness to identify correctly.

Transforming text-based sources is not an impossible task, but hand coding can be costly in terms of time and resources, and automated approaches often require significant labor to get up and running and are inflexible to new cases. There is a need for an approach that combines the adaptability of hand coding with efficient automated approaches. I propose that we can employ LLM to meet this need and improve the number and variety of text-based sources we transform into usable datasets.

3.2 What are LLMs and How Can They Help?

The term LLM refers to a family of deep learning models built upon a *transformer* framework (Vaswani et al. 2017). Hallmarks of these models include immense training sets comprised of a variety of text sources ranging from web content, code, or even classical literature and billions (or, at the time of writing, trillions) of individual parameters, which allow these models to understand the human language on a previously unprecedented level (Adiwardana et al. 2020; Brown et al. 2020; Radford et al. 2019). While there currently exists a plethora of LLMs, each with its unique size, scope, and intended purpose, in the interest of this paper I will focus primarily on the application of large generative language models such as OpenAI’s GPT-3 (Brown et al. 2020), Google’s Lambda (Thoppilan et al. 2022) or the open-sourced LLama (Touvron et al. 2023) and BLOOM (Scao

⁴I should note that recent advances NER and OCR based approaches can successfully associate entities within texts, but they would likely still require some amount of training and fine-tuning to function successfully (Kim et al. 2022)

et al. 2023) models. At the most basic level, these larger models are trained to predict words and phrases in response to some initial prompt or context. For example, when presented with the question “What color is the sky?” the model may respond with “Blue” as this is the most probable response to the initial question.⁵

Despite the seeming simplicity of this design, these models have shown a remarkable ability to perform emergent capabilities ranging from replicating common natural language processing tasks (e.g., sentiment analysis, classification, and topic modeling) (Brown et al. 2020; Ornstein, Blasingame, and Truscott, n.d.), to performing information retrieval tasks (Zhou, Ethayarajh, and Jurafsky 2021), or even generating synthetic human responses to survey questions (Argyle et al. 2023). LLMs perform these emergent tasks by changing the core tasks into a series of next-word prediction tasks. Models are presented with some form of prompt or initial context and then tasked with generating a response in accordance with some set of rules. For example, a researcher interested in obtaining the sentiment of tweets might ask the model whether a tweet is positive, negative, or neutral in tone and then provide the tweet’s text. The model would then give the most probable response from that selection of options. To further improve a model’s performance, researchers can provide examples of desired responses, commonly referred to as one or few-shot learning examples (Lazaridou et al. 2022).

I propose that we can turn the transformation of text-based sources into a next-word prediction task and then have these large-scale models complete them as synthetic coders. By providing these models with our text-based source and a prompt detailing how we want it processed, we can leverage the advanced capabilities of these models to transform texts in an adaptable and efficient fashion. Specifically, I suggest we provide these models with instructions—much akin to instructions for human coders—regarding the type of information we want from texts and then have the model identify, extract, and format said information into quantifiable datasets. These models’ ability to understand the contextual features of text allows them to overcome many of the hurdles associated

⁵This is a general overview of how these models function. Models include a variety of other features, such as context windows or temperature. Models can also be trained with additional Reinforcement Learning with Human Feedback (RLHF), which influence a model’s ability to respond to specific contexts (Ouyang et al. 2022).

with prior automated approaches and produce near-human-quality coding responses. However, these models should not synthetically code all text-based sources without validation.

While impressive, LLMs have several limitations that may negatively affect their ability to code text-based sources. First, these models are trained on a large variety of human-generated data and, as such, can express many of the same biases and stereotypes present within these data (Garg et al. 2018; Caliskan, Bryson, and Narayanan 2017; Bender et al. 2021; Zhou, Ethayarajh, and Jurafsky 2021). These reflected biases may be especially concerning for political science research as models may systematically misidentify information pertaining to marginalized groups. For example, a model tasked with identifying government officials with a piece of text may inappropriately identify (or misidentify) individuals along the lines of gender or race. Second, these models can produce noisy results or hallucinate data not initially present within the original text (Lin, Hilton, and Evans 2022). In either case, the incorrect data may be difficult to spot and contaminate the final dataset in unseen ways. Given these potential problems with LLMs, researchers must have some way to validate the output of these models (Grimmer and Stewart 2013; Knox, Lucas, and Cho 2022). Therefore, in the following section, I propose a human-in-the-loop process in which researchers iteratively design the context for these models and then validate their performance on a subset of their data to identify—and ideally minimize—instances of error and bias.

3.3 The Proposed Pipeline

Following many of the principles from recent advances in adaptive machine learning (Enamorado, Fifield, and Imai 2019; Wiedemann 2019; Miller, Linder, and Mebane 2020), the framework relies on leveraging LLMs efficiency to process some text-based sources and then utilizing humans to correct and guide these models to the correct answer (Mozer et al. 2019). While most applications of adaptive machine learning rely on the models to iteratively update based on human input, this process follows an alternative approach. Much like a researcher would update instructions for human coders, the researcher updates parameters and inputs in response to the model’s performance.

The process consists of hand coding a subset of data and then using that data to improve the prompt and context sent to the LLM iteratively. Once the LLM achieves a researcher’s desired level of accuracy, it is used to transform the remaining data. Figure 3.2 provides an overview of each of the steps within the workflow. Notably, the process overall is agnostic to the exact LLM utilized, but as detailed below, the model’s characteristics may substantially affect a researcher’s choices throughout the process.

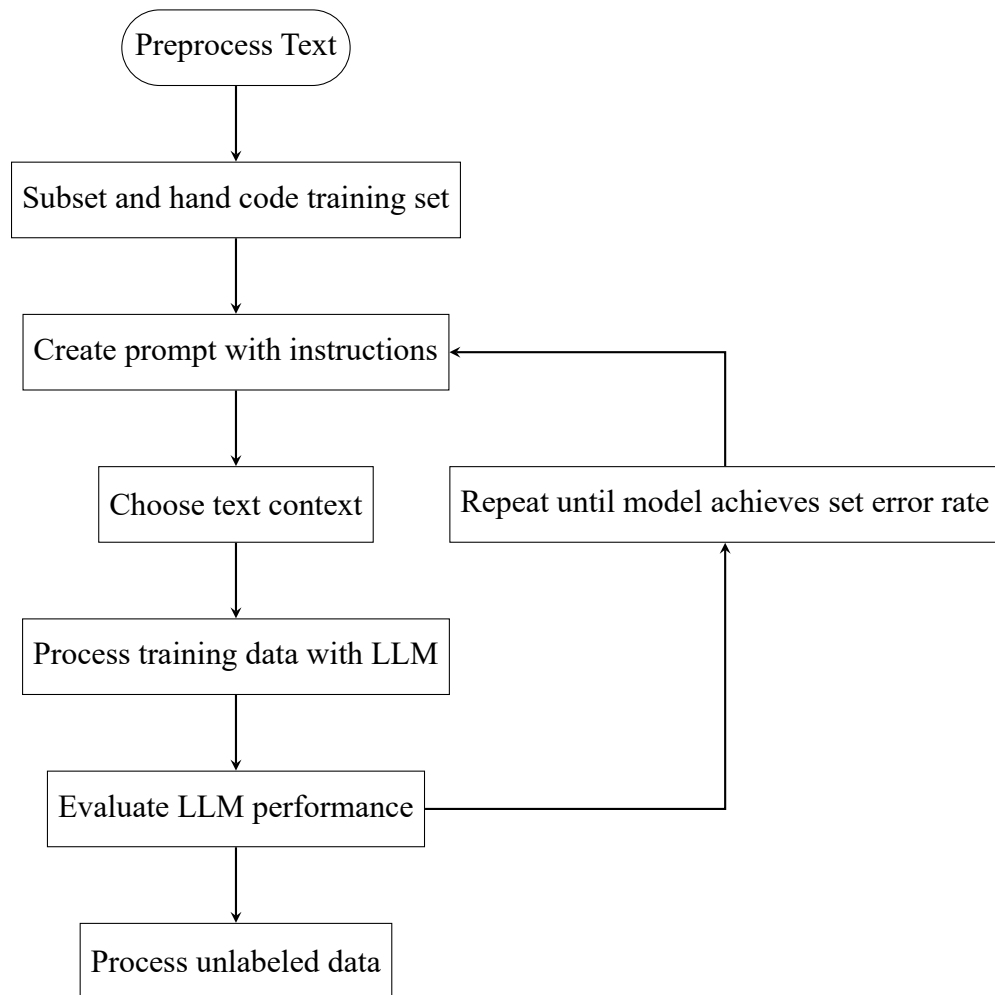


Figure 3.2: Framework for using Large Lanaguage Models to transform text-based sources.

3.3.1 Preparing Text-Based Sources

The first step of the workflow is to preprocess the text into some machine-readable format. Like most text-as-data approaches, how we preprocess the text can significantly affect our final results (Grimmer, Roberts, and Stewart 2022; Wilkerson and Casas 2017). However, unlike most preprocessing for NLP, we want to be careful not to remove too much from the texts. We want our text to maintain its rough format and maintain any important contextual features surrounding our desired data. For texts that require some form of OCR, as is the case with many legal and government documents, we need to be aware that the quality of the transformed text may affect their data of interest. Poorly converted data can inhibit even LLM’s performance by potentially obfuscating our desired data within the text (Van Strien et al. 2020).

Next, we must divide some of our texts into a training set. This training set aims to validate the LLM’s performance, so it should represent a wide variety of texts within our corpus. If we draw on texts from multiple sources, the training set should include examples from each source. Similarly, if we rely on text spanning numerous years, the training set should consist of some texts from each period present within the corpus. For each text within this set, we must manually code the text into our desired data format. This set will be used as the golden standard to compare the performance of the LLM. There is no hard-set rule regarding the size of a training set, although most conventional splits for machine learning rely on splitting at least 20 percent of the sample into a training set. Notably, larger training set sizes will allow researchers to better check the LLM’s performance at the cost of additional hand coding.⁶ After successfully dividing our corpus, we can begin constructing the prompt for our LLM.

3.3.2 Selecting a Prompt and Context

Much like human coders, models require some form of text-based instructions to generate responses. This prompt details the instructions for the model and will be our primary point of iteration and

⁶Several notable advances have been made in optimizing one’s choice in training sets (Kaufman 2024). Although it is not clear whether they could be applied in this context.

revision throughout the workflow. While the exact wording of our prompt will vary depending on our text and desired data, the prompt should include three key components. First, we should specify the information we wish to extract from the text. We should be as specific as possible in this stage, as more detailed descriptions tend to lead to higher-quality outputs from the model. Second, is the format we want the model to return our data in.⁷ Beyond making it easier to process the data output from the model, having a consistent output can make it easier for us to detect errors. In addition to these formatting instructions, we should provide instructions on what the model should do when the desired data is absent from a text. By doing so, we can reduce unexpected responses from the model and provide a clear null case during the validation step. Finally, a one or few-shot learning example of the desired output data. For example, again, in the case of obtaining sentiment from a tweet, we would want to include examples of tweets and their respective coding. These training examples can significantly improve an LLM's performance and help the model produce consistent results (Brown et al. 2020). If including a full training example is not feasible due to the type of desired data or length of the text, we can include an example output to obtain a similar effect on model performance and consistency.⁸

Our last choice in the workflow is the context of the text to share with the model. Specifically, how much of the text-based source should be included following the prompt? All LLMs have a maximum sequence window. These windows are primarily a factor of the model's number of parameters and effectively dictate how much text the model can intake and output at a time. Most models have context windows between 2048 and 8000 tokens.⁹ Given these limits, we must be careful regarding how we present our text to the model. For smaller texts, we can freely include the entirety of the text following the prompt. However, for larger texts, we may need to split the text into sections and have the model process each section. If we break our texts, we should consider how the splits may affect our desired information. We may inadvertently split a section of desired

⁷Given most LLMs are trained on a large body of code, we may opt to have the model return the results in a data format from our preferred coding language (e.g., a data frame within R, or a dictionary within Python).

⁸I show an example of this practice within both case studies later in this article.

⁹Tokens are sequences of characters representing meaningful units of text. Usually, tokens consist of one or more words, but different models utilize different tokenizers to preprocess text.

information and effectively cut it out of the text we provide the model. Beyond context windows, we may want to give the model specific sections of text to limit the chances of false positives or cut down on the total text sent to the model. For example, if we only care about information present at the beginning of a document, it would be faster and potentially safer only to provide said section to model rather than the entirety of the text. However, if our desired information could occur throughout a text, we would need to give the entirety of the text to the model.

Before we process our training data, we may need to decide on a few additional model parameters. These parameters vary between models but often include a maximum response length, a starting seed, or model temperature. Of these parameters, a model's temperature is the most likely to affect the model's capacity to process text-based sources. Conceptually, a model's temperature dictates the diversity in its responses. Often, individuals will equate this parameter to a model's practical creativity. In the context of next word prediction, a model may not choose the most probable word in every case and instead choose an alternative, less likely word or phrase. Given this workflow aims to produce consistent results from text-based sources, we will want to use lower model temperatures.

3.3.3 Processing, Iterating, and Validating

With our prompt, context, and ancillary parameters decided, we can finally ask the LLM to transform our training data. Most LLMs can be sent data directly through an API.¹⁰ For each text within our training set, we will send the model our specific prompt and context and then store its response. Since it can be difficult to anticipate whether our early prompts will be effective, we may want to start with a small portion of our training set and slowly incorporate more of our data as we finetune the model.

After the model processes our training set, we will evaluate its performance using standard accuracy and calibration metrics . While we will need the model's overall performance to report

¹⁰Popular models often also include an official or popular open-sourced package that facilitates interacting with their APIs in the most popular coding languages (see the `“text2data”` package from Ornstein, Blasingame, and Truscott (n.d.) as an example.)

our uncertainty later, we want to focus on instances in which the model failed to transform text or mislabeled data for the initial iterative process. Were there certain types of text the model failed to process? Does the model include irrelevant data in its response? Did the model show any obvious signs of bias in the data that it correctly or incorrectly identified? Common corrections we may make in response to these questions are increasing the text sent to the model, changing a prompt to be more specific, changing the learning examples, specifying additional cases to exclude from the output, or breaking up transform into separate steps.¹¹ After correcting our prompt and context, we can repeat our workflow until the model's output is sufficiently similar to our hand-coded data. Like any machine learning approach, LLMs will likely never perfectly correspond to our training data. However, as I highlight in later case studies, obtaining accuracy metrics between 80% and 95% is not unreasonable. If we desire to push the model's performance beyond these limits, we may risk effectively "overfitting" to our training set. This "overfitting" can be problematic if our set is not representative of our larger corpus or if we want to transform a new data set.¹²

Once our model reaches our desired level of accuracy, we can send our unlabeled text to the model. If any errors occur in this process, we may want to adjust our training set to incorporate problematic texts and reiterate through the workflow. After we process all unlabeled texts, we should record our chosen LLM, last prompt, context, and model accuracy and present them clearly within our final results.

3.4 Applications and Validation

In the following two applications, I demonstrate how LLMs can efficiently and accurately transform text-based sources into high-quality data. For both applications, I use OpenAI's GPT 3.5-turbo-1106 to identify, extract, and format meaningful information from text-based records across two starkly different areas of political science. I validate the model's performance with a comparison to its

¹¹If we are interested in many pieces of information, creating a different prompt for each piece may be more effective.

¹²To alleviate overfitting concerns, we may wish to hold out a portion of the training set as a validation set for our final model parameters.

predicted performance from the iterative workflow and human-coded data. Examples of each text-based source and how they were processed can be found in Section C.1 of the Appendix.

3.4.1 Extracting Commenters From Local Meeting Minutes

Understanding who and how individuals participate in local governmental meetings has become a key area of study for urban politics (Einstein, Palmer, and Glick 2019; Einstein et al. 2022; Yoder 2020; Nuamah 2021; Collins 2021). Much of this research relies on datasets of political participation constructed directly from publicly available meeting minutes. In many cases, these minutes record the names, comments, and, at times, addresses of those who participate in public meetings. However, extracting this fruitful data takes work. Meeting minutes, while by law include similar information, such as the names of public officials, votes on local policies, and public commenters, they do not follow similar formats.

The laws mandating meeting minutes frequently lack specificity, and it is up to the local government to determine the style and level of detail in these recorded minutes (O’connor and Baratz 2004). As a result, meeting minutes can vary substantially in detail and style between any two local governments.¹³ Formatting within governments can even change as new governmental clerks are appointed or the government adopts new strategies, formats, or technologies. These limitations make constructing a diverse and longitude dataset of meeting participation an extremely time-intensive task. As a compromise, research within this area often focuses only on a handful of similar local governments (e.g., Yoder 2020; Einstein, Palmer, and Glick 2019) or infers participation based on national survey responses (Nuamah and Ogorzalek 2021). These compromises, however, come at the cost of potentially noisy estimates and limited variation in the types of local governments examined.

A demand exists within the field to alleviate the burden of processing these minutes, and LLMs could effectively address this need. To demonstrate how LLMs can be applied to this problem, I

¹³Section C.1 within the Appendix provides a select sample of minutes from two similar sized cities within the St. Louis region.

apply the workflow above to select school board meeting minutes from the St. Louis region. The dataset consists of all the general meeting minutes from 2018 to 2022 for 16 school boards. The data includes 667 separate meetings ranging immensely in style and detail. This sample should provide a wide variety of meeting minutes between local governments and a perfect test of the capabilities of LLMs.

I start by hand coding each meeting's minutes and extracting the commenters, addresses, and comments for each meeting. In reality, most researchers will only label a small sample of the data, so to mimic that, I randomly select ten meetings from each government to represent my initial training set of 160 meetings. Next, I construct an initial prompt and display the final prompt in Figure 3.3. Given that public comments can occur at any place within a meeting, I feed the model the entirety of the meeting text. For longer meetings, I split the meetings into 3,000 token chunks. I set the model's temperature to 0 and then feed in each meeting with the associated prompt. To evaluate the model, I examine whether the model obtained the full names, addresses, and comments of each commenter within my training set of meetings.¹⁴ After approximately four iterations of adjusting the prompt, with the primary changes being additional specificity and modifications to the example outputs, the model displayed a precision of 0.934, recall of 0.984, and F1 of 0.957 for the training set. Given these promising estimates, I fed the remaining meeting through the model and evaluated its performance on the entire dataset.

¹⁴In the case of the comment's text, I only examined whether the model returned a comment and not necessarily whether the comment's text matched perfectly.

In the following text of recorded public meeting minutes can you identify all of the commenters who spoke during the public comment period(s), all of their addresses if available, and all of their complete comments. Return them in a JSON format. Please do not include public officials and use NA for any missing information.

Example Output:

```
{  
  "Name":["Shawn Spencer"],  
  "Address":["5005 North Fraser Way"],  
  "Comment":["Concerns about recent crime in the community"]  
}
```

If no one commented please return:

```
{  
  "Name":["No Commenters"],  
  "Address":[""],  
  "Comment":[""]  
}
```

Meeting Text:

[MEETING MINUTE TEXT]

Figure 3.3: Prompt to extract commenters from public meeting text using GPT-3

Overall, the model could accurately extract almost all commenters from the meeting minutes dataset with an average precision score of 0.935. Incredibly, the model exhibited a low false positive rate with an average recall of 0.977. To contextualize these findings, an untrained BERT NER model only achieved a precision of 0.780 and recall of 0.850, requiring extensively more coding and filtering of results.¹⁵ While the LLM successfully parsed most meeting minutes, it failed in a few select cases. The LLM tended to code meetings inaccurately, in which school boards only included the number of total commenters without any form of identifying information. If we would

¹⁵More details on the NER application can be found within the Appendix.

be interested in the count in addition to the commenter's details, we would need to adjust the prompt accordingly or construct a separate one. Another area in which the model struggled was when individuals made detailed and lengthy comments. In these cases, the model would occasionally truncate the comment down. Adjusting the context of the text remedied this problem, but not completely.

In terms of time to code, the model took approximately 3 hours to fully parse all the meeting minutes at a cost of roughly \$0.01 per meeting. Including the training, the model costs \$8.28 to process the dataset. In comparison, human coding the dataset required an estimated 15 hours to parse the meeting minutes. Beyond being faster and less costly, utilizing an LLM could allow us to parse meetings multiple times for other forms of information, such as the meeting format, the names of public officials, and other critical pieces of information, all only requiring us to hand code a small sample of the entire data.

3.4.2 Identifying Interest Groups in Amicus Curiae Briefs

Amicus briefs are a valuable tool to estimate interest group's political participation before the Supreme Court. Beyond simple participation, they can reveal underlying political networks (Box-Steffensmeier and Christenson 2015), track groups' political influence (Box-Steffensmeier, Christenson, and Hitt 2013), and provide measures of group ideology (Abi-Hassan et al. 2023). To utilize these records, previous researchers hired coders to go brief by brief and extract the names of interest groups and the legal direction of the brief in a time-intensive process. As highlighted previously in this article, the core difficulty with extracting this valuable information from these documents is that organizational names can come in various formats and can vary in their exact placement at the beginning or end of a brief. Identifying the legal direction of a brief is more straightforward as most briefs explicitly state the party they support; however, in a few cases, briefs may report support or opposition of a lower court decision rather than a case's litigants. The combination of varying formats and locations can make correctly identifying either the signers or legal direction difficult,

even for human coders. Expediting the processing of these briefs could bring numerous benefits, from easing the maintenance of existing datasets, expanding the coverage of included briefs, or even examining other novel information within the brief’s text. LLMs could be applied to these documents to achieve high-quality results at a fraction of the cost and time.

To demonstrate this application, I utilize the amicus curiae database from [Box-Steffensmeier and Christenson \(2012\)](#). I focus on all briefs sent to the Court in 2014. This selection narrows the universe of briefs to process but still provides a diverse sample of 190 individual briefs.¹⁶ Subsetting 40 briefs as my training set, I construct the initial prompt and set my initial context as the first and last five pages of each brief. Figure 3.4 displays the final prompt used to process the text. In each iteration, I evaluate whether the model correctly identified all signers and separately whether it correctly identified the direction of the brief. Given names may vary slightly between the handing coding and the original text of the brief, I allow for some difference between the two sets of names by utilizing a fuzzy string-matching algorithm to detect a match.¹⁷ After iteratively adjusting the prompt and enlarging the context to the first and last ten pages of each brief, the model identified brief singers with a recall of 0.922 and precision of 0.90 and the legal direction with an accuracy of 0.975.

¹⁶The Court received 351 total briefs in 2014, but only 190 were publicly available and hand-coded at the time of writing.

¹⁷Specifically, I measure the Damerau-Levenshtein distance between the vectors of names and consider a pair a match if their distance is less than five. Given the names were already close, this distance largely accounted for the inclusions/exclusion of words like “the” or “inc.”

In the following amicus curiae brief, which organizations signed the brief and was it for the petitioner, respondent or neither? Please record the organization in the form of a Python dictionary. An example of the output is below and then the text to parse. Do not include the law firm who prepared the brief.

Example Outputs:

```
{  
  "AMERICA'S ESSENTIAL HOSPITALS" : "Respondent",  
  "NATIONAL FEDERATION OF THE BLIND" : "Respondent",  
  "TYRELL CORPORATION" : "Respondent"  
}  
  
{  
  "EXXON MOBILE" : "Neither",  
  "WEYLAND-YUTANI CORPORATION" : "Neither"  
}
```

Brief Text:
[BRIEF TEXT]

Figure 3.4: Prompt to extract the name of signers from amicus curiae briefs using GPT-3

When applied to the remaining data, the model displayed a slightly lower average recall of 0.857 and a precision of 0.867.¹⁸ While the model did appear to extract the majority of information from briefs correctly, this drop in performance is concerning as, in real applications, it would be unknown to researchers. Upon closer inspection, two primary factors contributed to the performance dip. First, in some cases, the model would incorrectly format responses, leading to all names being considered incorrect. Often, the model would group two or more organizations rather than separate them in the response. Adjusting the prompt to specify this separation or further parsing the model's results would likely solve this issue. Second, the model would sometimes include individuals or agencies in addition to their parent organizations, whereas previous coders only had one or the other. For example, the model may include both the public finance authority for a city and the

¹⁸Processing the entire dataset cost \$1.18 and took approximately 30 minutes to complete.

city itself as signers. Whether this difference is substantively important is unclear but may be addressed by adding further specificity to the prompt. Regarding identifying a brief’s legal direction, the model performed similarly to the training set with an accuracy of 0.956. Overall, using an LLM to code briefs seems to be a feasible approach to accelerate the process while maintaining the quality of the final dataset. Nevertheless, the drop in performance between the training and full datasets underscores the potential vulnerability of relying on small training sets to evaluate an LLM’s performance.

3.5 Discussion

Traditionally, the conversion of text-based sources into meaningful datasets has been a laborious and time-consuming task. However, LLMs present a promising solution to this challenge. These models have the potential to combine the efficiency and speed of traditional automated approaches with the flexibility and adaptability of human coders. By iteratively guiding and validating these models, we can create high-quality datasets at a fraction of the cost and time investment. Furthermore, offering details on how we utilized these models, including the prompt, context, and other parameters, provides a unique opportunity to recreate datasets from their original sources. This form of replication, typically avoided due to its high coding costs, can enhance transparency in using text-based sources and facilitate innovation with existing data. Beyond text transformation, LLMs are a promising tool throughout the research process as they assist researchers in effectively handling various data needs, from formatting data to extracting information or even matching across data sources. However, it’s essential to acknowledge that despite these promises, my proposed framework and the broader application of LLMs come with their own caveats and challenges.

Without coding the entirety of the data, it is impossible to guarantee that coding errors and LLMs’ potential biases will not permeate into the datasets they help create. It may be fruitful for future work to explore how we could leverage consensus metrics to combine multiple models to minimize the adverse effects of any particular model. Researchers should also recognize that even

in the case of the training set for which we have outcomes to compare, it may be challenging to check for all forms of potential bias. For example, in the case of extracting public commenters, researchers may require additional data such as commenter demographics to gauge whether the misclassifies individuals along the lines of gender or race. Such information is often not widely available and can be difficult to match to the original dataset. Overall, the quality of data produced by the model is a product of the diligence and care researchers put into evaluating and validating the model's performance. Researchers should be transparent regarding their implementation of LLMs and aware of potential biases in their final datasets.

Additionally, while LLMs may alleviate the burden of transforming data, they do not address the initial burden of gathering it in the first place. A researchers' choice in which sources to include within their dataset can be just as impactful on their study as how they processed their data (Gaikwad and Herrera 2023). Hopefully by alleviating the pressures of coding each source, LLM can allow researchers to spend more time gathering novel data or better utilize pre-existing collections of data. The latter benefit may be especially promising given recent innovative data collection efforts such as (Barari and Simko 2023)'s collection of meeting transcripts. Each of which holds to potentially to vastly increase our understanding of each of their respective fields.

Overall, Large Language Models (LLMs) are a promising tool for future researchers. Like any tool, they should be implemented with transparency and acknowledgment of their potential limitations. If appropriately used, these models offer the potential to drastically increase the accessibility of political science research and expand the coverage and diversity of data used across research domains.

References

- Abi-Hassan, Sahar, Janet M. Box-Steffensmeier, Dino P. Christenson, Aaron R. Kaufman, and Brian Libgober. 2023. "The Ideologies of Organized Interests and Amicus Curiae Briefs: Large-Scale, Social Network Imputation of Ideal Points." *Political Analysis* 31, no. 3 (July): 396–413. ISSN: 1047-1987, 1476-4989, accessed August 14, 2023. <https://doi.org/10.1017/pan.2022.34>.
- Achen, Christopher H., and Larry M. Bartels. 2016. *Democracy for Realists: Why Elections Do Not Produce Responsive Government*. Princeton Studies in Political Behavior. Princeton: Princeton University Press. ISBN: 978-0-691-16944-6.
- Adams, Brian. 2004. "Public Meetings and the Democratic Process." *Public Administration Review* 64, no. 1 (January): 43–54. ISSN: 0033-3352, 1540-6210, accessed March 11, 2022. <https://doi.org/10.1111/j.1540-6210.2004.00345.x>.
- Adiwardana, Daniel, Minh-Thang Luong, David R. So, Jamie Hall, Noah Fiedel, Romal Thoppilan, Zi Yang, et al. 2020. *Towards a Human-like Open-Domain Chatbot*, arXiv:2001.09977, February. Accessed January 14, 2024. arXiv: [2001.09977](https://arxiv.org/abs/2001.09977) [cs, stat].
- Aldrich, John H., Jacob M. Montgomery, and Wendy Wood. 2011. "Turnout as a Habit." *Political Behavior* 33, no. 4 (December): 535–563. ISSN: 0190-9320, 1573-6687, accessed July 27, 2023. <https://doi.org/10.1007/s11109-010-9148-3>.
- Anzia, Sarah F. 2021. "Party and Ideology in American Local Government: An Appraisal." *Annual Review of Political Science* 24, no. 1 (May): 133–150. ISSN: 1094-2939, 1545-1577, accessed February 17, 2022. <https://doi.org/10.1146/annurev-polisci-041719-102131>.
- Anzia, Sarah F., and Terry M. Moe. 2014. "Collective Bargaining, Transfer Rights, and Disadvantaged Schools." *Educational Evaluation and Policy Analysis* 36, no. 1 (March): 83–111. ISSN: 0162-3737, accessed January 13, 2024. <https://doi.org/10.3102/0162373713500524>.
- Arceneaux, K. 2005. "Does Federalism Weaken Democratic Representation in the United States?" *Publius: The Journal of Federalism* 35, no. 2 (January): 297–311. ISSN: 0048-5950, 1747-7107, accessed January 26, 2021. <https://doi.org/10.1093/publius/pji015>.
- Arceneaux, Kevin. 2006. "The Federal Face of Voting: Are Elected Officials Held Accountable for the Functions Relevant to Their Office?" *Political Psychology* 27, no. 5 (October): 731–754. ISSN: 0162-895X, 1467-9221, accessed January 27, 2021. <https://doi.org/10.1111/j.1467-9221.2006.00530.x>.

- Argyle, Lisa P., Ethan C. Busby, Nancy Fulda, Joshua R. Gubler, Christopher Rytting, and David Wingate. 2023. “Out of One, Many: Using Language Models to Simulate Human Samples.” *Political Analysis* 31, no. 3 (July): 337–351. ISSN: 1047-1987, 1476-4989, accessed August 14, 2023. <https://doi.org/10.1017/pan.2023.2>.
- Barari, Soubhik, and Tyler Simko. 2023. “LocalView, a Database of Public Meetings for the Study of Local Politics and Policy-Making in the United States.” *Scientific Data* 10, no. 1 (March): 135. ISSN: 2052-4463. <https://doi.org/10.1038/s41597-023-02044-y>.
- Barber, Michael, and John B. Holbein. 2020. “The Participatory and Partisan Impacts of Mandatory Vote-by-Mail.” *Science Advances* 6, no. 35 (August): eabc7685. ISSN: 2375-2548, accessed August 23, 2023. <https://doi.org/10.1126/sciadv.abc7685>.
- Bélanger, Éric, and Bonnie M. Meguid. 2008. “Issue Salience, Issue Ownership, and Issue-Based Vote Choice.” *Electoral Studies* 27, no. 3 (September): 477–491. ISSN: 02613794, accessed January 13, 2024. <https://doi.org/10.1016/j.electstud.2008.01.001>.
- Bender, Emily M., Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. “On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? □.” In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623. Virtual Event Canada: ACM, March. ISBN: 978-1-4503-8309-7, accessed November 14, 2023. <https://doi.org/10.1145/3442188.3445922>.
- Benoit, Kenneth, Drew Conway, Benjamin E. Lauderdale, Michael Laver, and Slava Mikhaylov. 2016. “Crowd-Sourced Text Analysis: Reproducible and Agile Production of Political Data.” *American Political Science Review* 110, no. 2 (May): 278–295. ISSN: 0003-0554, 1537-5943, accessed December 30, 2023. <https://doi.org/10.1017/S0003055416000058>.
- Benton, J. Edwin. 2005. “An Assessment of Research on American Counties.” *Public Administration Review* 65 (4): 462–474. ISSN: 1540-6210, accessed June 28, 2023. <https://doi.org/10.1111/j.1540-6210.2005.00472.x>.
- Berry, Christopher R., and William G. Howell. 2007. “Accountability and Local Elections: Rethinking Retrospective Voting.” *The Journal of Politics* 69, no. 3 (August): 844–858. ISSN: 0022-3816, 1468-2508, accessed April 18, 2022. <https://doi.org/10.1111/j.1468-2508.2007.00579.x>.
- Bohannon, John. 2011. “Social Science for Pennies.” *Science* 334, no. 6054 (October): 307–307. ISSN: 0036-8075, 1095-9203, accessed December 29, 2023. <https://doi.org/10.1126/science.334.6054.307>.
- Bommasani, Rishi, Drew A. Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S. Bernstein, et al. 2022. *On the Opportunities and Risks of Foundation Models*, arXiv:2108.07258, July. Accessed December 30, 2023. arXiv: [2108.07258](https://arxiv.org/abs/2108.07258) [cs].

- Bonica, Adam, Jacob M. Grumbach, Charlotte Hill, and Hakeem Jefferson. 2021. "All-Mail Voting in Colorado Increases Turnout and Reduces Turnout Inequality." *Electoral Studies* 72 (August): 102363. ISSN: 0261-3794, accessed August 24, 2023. <https://doi.org/10.1016/j.electstud.2021.102363>.
- Box-Steffensmeier, Janet M., and Dino P. Christenson. 2015. "Comparing Membership Interest Group Networks across Space and Time, Size, Issue and Industry." *Network Science* 3 (1): 78–97.
- . 2012. *Database on Supreme Court Amicus Curiae Briefs. Version 1.0*. [Computer file].
- Box-Steffensmeier, Janet M., Dino P. Christenson, and Matthew P. Hitt. 2013. "Quality Over Quantity: Amici Influence and Judicial Decision Making." *American Political Science Review* 107, no. 3 (August): 446–460. ISSN: 0003-0554, 1537-5943, accessed September 7, 2022. <https://doi.org/10.1017/S000305541300021X>.
- Brown, Tom B., Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, et al. 2020. *Language Models Are Few-Shot Learners*, arXiv:2005.14165, July. Accessed November 15, 2023. arXiv: [2005.14165](https://arxiv.org/abs/2005.14165) [cs].
- Bryan, Frank M. 2010. *Real Democracy: The New England Town Meeting and How It Works*. University of Chicago Press, March. ISBN: 978-0-226-07798-7.
- Bucchianeri, Peter. 2020. "Party Competition and Coalitional Stability: Evidence from American Local Government." *American Political Science Review* 114, no. 4 (November): 1055–1070. ISSN: 0003-0554, 1537-5943, accessed December 5, 2020. <https://doi.org/10.1017/S0003055420000386>.
- Burden, Barry C., David T. Canon, Kenneth R. Mayer, and Donald P. Moynihan. 2014. "Election Laws, Mobilization, and Turnout: The Unanticipated Consequences of Election Reform." *American Journal of Political Science* 58 (1): 95–109. ISSN: 0092-5853, accessed August 8, 2023. JSTOR: [24363471](https://www.jstor.org/stable/24363471).
- Bureau of Justice Statistics. 2019. "Local Police Departments, 2016: Personnel" (August).
- Caliskan, Aylin, Joanna J. Bryson, and Arvind Narayanan. 2017. "Semantics Derived Automatically from Language Corpora Contain Human-like Biases." *Science* 356, no. 6334 (April): 183–186. ISSN: 0036-8075, 1095-9203, accessed January 7, 2024. <https://doi.org/10.1126/science.aal4230>.
- Callaway, Brantly, and Pedro H.C. Sant'Anna. 2021. "Difference-in-Differences with Multiple Time Periods." *Journal of Econometrics* 225, no. 2 (December): 200–230. ISSN: 03044076, accessed April 17, 2023. <https://doi.org/10.1016/j.jeconom.2020.12.001>.

- Campbell, Angus. 1980. *The American Voter*. Unabridged ed. Midway Reprints. Chicago [Ill.]: University of Chicago Press. ISBN: 978-0-226-09254-6.
- Carlson, David, and Jacob M. Montgomery. 2017. "A Pairwise Comparison Framework for Fast, Flexible, and Reliable Human Coding of Political Texts." *American Political Science Review* 111, no. 4 (November): 835–843. ISSN: 0003-0554, 1537-5943, accessed January 7, 2024. <https://doi.org/10.1017/S0003055417000302>.
- Carpini, Michael X. Delli, Fay Lomax Cook, and Lawrence R. Jacobs. 2004. "PUBLIC DELIBERATION, DISCURSIVE PARTICIPATION, AND CITIZEN ENGAGEMENT: A Review of the Empirical Literature." *Annual Review of Political Science* 7, no. 1 (May): 315–344. ISSN: 1094-2939, 1545-1577, accessed July 27, 2023. <https://doi.org/10.1146/annurev.polisci.7.121003.091630>.
- Carreras, Miguel, Sofia Vera, and Giancarlo Visconti. 2023. "Who Does the Caring? Gender Disparities in COVID-19 Attitudes and Behaviors." *Politics & Gender* 19, no. 1 (March): 5–33. ISSN: 1743-923X, 1743-9248, accessed April 18, 2023. <https://doi.org/10.1017/S1743923X21000386>.
- Chan, Nathan K. 2021. "Political Inequality in the Digital World: The Puzzle of Asian American Political Participation Online." *Political Research Quarterly* 74, no. 4 (December): 882–898. ISSN: 1065-9129, accessed August 24, 2023. <https://doi.org/10.1177/1065912920945391>.
- Cho, Seongkyung, Karen Mossberger, David Swindell, and John David Selby. 2021. "Experimenting with Public Engagement Platforms in Local Government." *Urban Affairs Review* 57, no. 3 (May): 763–793. ISSN: 1078-0874, 1552-8332, accessed June 22, 2022. <https://doi.org/10.1177/1078087419897821>.
- Coffé, Hilde, and Catherine Bolzendahl. 2010. "Same Game, Different Rules? Gender Differences in Political Participation." *Sex Roles* 62, no. 5 (March): 318–333. ISSN: 1573-2762, accessed April 18, 2023. <https://doi.org/10.1007/s11199-009-9729-y>.
- Collins, Jonathan E. 2021. "Does the Meeting Style Matter? The Effects of Exposure to Participatory and Deliberative School Board Meetings." *American Political Science Review* 115, no. 3 (August): 790–804. ISSN: 0003-0554, 1537-5943, accessed January 8, 2023. <https://doi.org/10.1017/S0003055421000320>.
- Craig, Stephen C., and Paulina S. Cossette. 2020. "Who Owns What, and Why? The Origins of Issue Ownership Beliefs." *Politics & Policy* 48 (1): 107–134. ISSN: 1747-1346, accessed January 13, 2024. <https://doi.org/10.1111/polp.12338>.
- Dalton, Russell J. 2017. *The Participation Gap: Social Status and Political Inequality*. Oxford University Press. ISBN: 978-0-19-179798-9.

- de Benedictis-Kessner, Justin. 2018. "How Attribution Inhibits Accountability: Evidence from Train Delays." *The Journal of Politics* 80, no. 4 (October): 1417–1422. ISSN: 0022-3816, 1468-2508, accessed January 27, 2021. <https://doi.org/10.1086/698754>.
- . 2022. "Strategic Government Communication about Performance." *Political Science Research and Methods* 10, no. 3 (July): 601–616. ISSN: 2049-8470, 2049-8489, accessed June 29, 2023. <https://doi.org/10.1017/psrm.2021.7>.
- de Benedictis-Kessner, Justin, Diana Da In Lee, Yamil R. Velez, and Christopher Warshaw. 2023. "American Local Government Elections Database." *Scientific Data* 10, no. 1 (December): 912. ISSN: 2052-4463, accessed December 30, 2023. <https://doi.org/10.1038/s41597-023-02792-x>.
- de Benedictis-Kessner, Justin, and Christopher Warshaw. 2020a. "Accountability for the Local Economy at All Levels of Government in United States Elections." *American Political Science Review* 114, no. 3 (August): 660–676. ISSN: 0003-0554, 1537-5943, accessed July 24, 2023. <https://doi.org/10.1017/S0003055420000027>.
- . 2016. "Mayoral Partisanship and Municipal Fiscal Policy." *The Journal of Politics* 78, no. 4 (October): 1124–1138. ISSN: 0022-3816, 1468-2508, accessed May 26, 2020. <https://doi.org/10.1086/686308>.
- . 2020b. "Politics in Forgotten Governments: The Partisan Composition of County Legislatures and County Fiscal Policies." *The Journal of Politics* 82, no. 2 (April): 460–475. ISSN: 0022-3816, 1468-2508, accessed June 28, 2023. <https://doi.org/10.1086/706458>.
- Druckman, James N., Erik Peterson, and Rune Slothuus. 2013. "How Elite Partisan Polarization Affects Public Opinion Formation." *American Political Science Review* 107, no. 01 (February): 57–79. ISSN: 0003-0554, 1537-5943, accessed March 31, 2019. <https://doi.org/10.1017/S0003055412000500>.
- Egan, P.J. 2013. *Partisan Priorities: How Issue Ownership Drives and Distorts American Politics*. Partisan Priorities: How Issue Ownership Drives and Distorts American Politics. Cambridge University Press. ISBN: 978-1-107-04258-2.
- Einstein, Katherine Levine, David Glick, Luisa Godinez Puig, and Maxwell Palmer. 2022. "Still Muted: The Limited Participatory Democracy of Zoom Public Meetings." *Urban Affairs Review* (February): 107808742110704. ISSN: 1078-0874, 1552-8332, accessed March 27, 2023. <https://doi.org/10.1177/10780874211070494>.
- Einstein, Katherine Levine, and David M. Glick. 2018. "Mayors, Partisanship, and Redistribution: Evidence Directly from U.S. Mayors." *Urban Affairs Review* 54, no. 1 (January): 74–106. ISSN: 1078-0874, 1552-8332, accessed April 19, 2022. <https://doi.org/10.1177/1078087416674829>.

- Einstein, Katherine Levine, Maxwell Palmer, and David M. Glick. 2019. "Who Participates in Local Government? Evidence from Meeting Minutes." *Perspectives on Politics* 17, no. 1 (March): 28–46. ISSN: 1537-5927, 1541-0986, accessed March 11, 2022. <https://doi.org/10.1017/S153759271800213X>.
- Enamorado, Ted, Benjamin Fifield, and Kosuke Imai. 2019. "Using a Probabilistic Model to Assist Merging of Large-Scale Administrative Records." *American Political Science Review* 113, no. 2 (May): 353–371. ISSN: 0003-0554, 1537-5943, accessed February 14, 2022. <https://doi.org/10.1017/S0003055418000783>.
- Fielding, Nigel, and Martin Innes. 2006. "Reassurance Policing, Community Policing and Measuring Police Performance." *Policing and Society* 16, no. 2 (June): 127–145. ISSN: 1043-9463. <https://doi.org/10.1080/10439460600662122>.
- Fiorina, Morris P., and Samuel J. Abrams. 2008. "Political Polarization in the American Public." *Annual Review of Political Science* 11, no. 1 (June): 563–588. ISSN: 1094-2939, accessed January 14, 2024. <https://doi.org/10.1146/annurev.polisci.11.053106.153836>.
- Fishkin, James. 2009. *When the People Speak: Deliberative Democracy and Public Consultation*. OUP Oxford, August. ISBN: 978-0-19-161015-8.
- Fung, Archon. 2006. "Varieties of Participation in Complex Governance." *Public Administration Review* 66:66–75. ISSN: 0033-3352, accessed April 18, 2022. JSTOR: 4096571.
- Gaikwad, Nikhar, and Veronica Herrera. 2023. "Transparency for Text-Based Sources: From Principles to Practice." *Perspectives on Politics* 21, no. 4 (December): 1240–1257. ISSN: 1537-5927, 1541-0986, accessed December 30, 2023. <https://doi.org/10.1017/S153759272200069X>.
- Garg, Nikhil, Londa Schiebinger, Dan Jurafsky, and James Zou. 2018. "Word Embeddings Quantify 100 Years of Gender and Ethnic Stereotypes." *Proceedings of the National Academy of Sciences* 115, no. 16 (April). ISSN: 0027-8424, 1091-6490, accessed January 7, 2024. <https://doi.org/10.1073/pnas.1720347115>.
- Gastil, John, E. Pierre Deess, and Phil Weiser. 2002. "Civic Awakening in the Jury Room: A Test of the Connection between Jury Deliberation and Political Participation." *The Journal of Politics* 64, no. 2 (May): 585–595. ISSN: 0022-3816, 1468-2508, accessed July 27, 2023. <https://doi.org/10.1111/1468-2508.00141>.
- Gimpel, James G., Nathan Lovin, Bryant Moy, and Andrew Reeves. 2020. "The Urban–Rural Gulf in American Political Behavior." *Political Behavior* 42, no. 4 (December): 1343–1368. ISSN: 0190-9320, 1573-6687, accessed July 26, 2023. <https://doi.org/10.1007/s11109-020-09601-w>.

- Gimpel, James G., and Andrew Reeves. 2023. "The Urban-Rural Divide and Residential Contentment as Antecedents of Political Ideology." *Cities* 146 (December): 104720. ISSN: 0264-2751. <https://doi.org/10.1016/j.cities.2023.104720>.
- Green, Donald P., Bradley Palmquist, and Eric Schickler. 2004. *Partisan Hearts and Minds: Political Parties and the Social Identities of Voters*. The Institution for Social and Policy Studies. Yale University Press. ISBN: 978-0-300-10156-0.
- Grimmer, Justin, Margaret E. Roberts, and Brandon M. Stewart. 2022. *Text as Data: A New Framework for Machine Learning and the Social Sciences*. Princeton University Press. ISBN: 978-0-691-20755-1.
- Grimmer, Justin, and Brandon M. Stewart. 2013. "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts." *Political Analysis* 21 (3): 267–297. ISSN: 1047-1987, 1476-4989, accessed June 21, 2023. <https://doi.org/10.1093/pan/mps028>.
- Häffner, Sonja, Martin Hofer, Maximilian Nagl, and Julian Walterskirchen. 2023. "Introducing an Interpretable Deep Learning Approach to Domain-Specific Dictionary Creation: A Use Case for Conflict Prediction." *Political Analysis* 31 (4): 481–499. ISSN: 1047-1987. <https://doi.org/10.1017/pan.2023.7>.
- Hartney, Michael T., and Leslie K. Finger. 2022. "Politics, Markets, and Pandemics: Public Education's Response to COVID-19." *Perspectives on Politics* 20, no. 2 (June): 457–473. ISSN: 1537-5927, 1541-0986, accessed June 22, 2022. <https://doi.org/10.1017/S1537592721000955>.
- Healy, Andrew, and Neil Malhotra. 2013. "Retrospective Voting Reconsidered." *Annual Review of Political Science* 16, no. 1 (May): 285–306. ISSN: 1094-2939, 1545-1577, accessed December 6, 2020. <https://doi.org/10.1146/annurev-polisci-032211-212920>.
- Hersh, Eitan. 2020. *Politics Is for Power: How to Move Beyond Political Hobbyism, Take Action, and Make Real Change*. Scribner, January. ISBN: 1-982116-78-1.
- Hodgkinson, Tarah, Tullio Caputo, and Michael L. McIntyre. 2019. "Beyond Crime Rates and Community Surveys: A New Approach to Police Accountability and Performance Measurement." *Crime Science* 8, no. 1 (December): 13. ISSN: 2193-7680, accessed December 6, 2020. <https://doi.org/10.1186/s40163-019-0108-x>.
- Hopkins, Daniel J. 2018. *The Increasingly United States, How and Why American Political Behavior Nationalized*. University of Chicago Press. ISBN: 0-226-53037-X.
- Huddy, Leonie, and Nayda Terkildsen. 1993. "The Consequences of Gender Stereotypes for Women Candidates at Different Levels and Types of Office." *Political Research Quarterly* 46 (3): 503–525. ISSN: 1065-9129, accessed August 22, 2023. <https://doi.org/10.2307/448945>. JSTOR: 448945.

- Hutto, C., and Eric Gilbert. 2014. “VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text.” *Proceedings of the International AAAI Conference on Web and Social Media* 8, no. 1 (May): 216–225. <https://doi.org/10.1609/icwsm.v8i1.14550>.
- ICMA. 2020. *Reflecting on Virtual Public Meetings*. <https://icma.org/articles/article/reflecting-virtual-public-meetings>, May. Accessed August 24, 2023.
- Imai, Kosuke, and In Song Kim. 2021. “On the Use of Two-Way Fixed Effects Regression Models for Causal Inference with Panel Data.” *Political Analysis* 29, no. 3 (July): 405–415. ISSN: 1047-1987, 1476-4989, accessed June 16, 2023. <https://doi.org/10.1017/pan.2020.33>.
- Imai, Kosuke, and Marc Ratkovic. 2014. “Covariate Balancing Propensity Score.” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 76, no. 1 (January): 243–263. ISSN: 13697412, accessed January 26, 2021. <https://doi.org/10.1111/rssb.12027>.
- Jalal, Mona, Kate K. Mays, Lei Guo, and Margrit Betke. 2020. “Performance Comparison of Crowdworkers and NLP Tools on Named-Entity Recognition and Sentiment Analysis of Political Tweets.” *arXiv:2002.04181 [cs]* (August). Accessed April 14, 2022. arXiv: [2002.04181 \[cs\]](https://arxiv.org/abs/2002.04181).
- Jensen, Amalie, William Marble, Kenneth Scheve, and Matthew J. Slaughter. 2021. “City Limits to Partisan Polarization in the American Public.” *Political Science Research and Methods* 9, no. 2 (April): 223–241. ISSN: 2049-8470, 2049-8489, accessed August 16, 2023. <https://doi.org/10.1017/psrm.2020.56>.
- Johnson, Carol, and Blair Williams. 2020. “Gender and Political Leadership in a Time of COVID.” *Politics & Gender* 16, no. 4 (December): 943–950. ISSN: 1743-923X, 1743-9248, accessed April 18, 2023. <https://doi.org/10.1017/S1743923X2000029X>.
- Karpinska, Marzena, and Mohit Iyyer. 2023. *Large Language Models Effectively Leverage Document-Level Context for Literary Translation, but Critical Errors Persist*, arXiv:2304.03245, May. Accessed November 14, 2023. arXiv: [2304.03245 \[cs\]](https://arxiv.org/abs/2304.03245).
- Kaufman, Aaron R. 2024. “Selecting More Informative Training Sets with Fewer Observations.” *Political Analysis* 32, no. 1 (January): 133–139. ISSN: 1047-1987, 1476-4989, accessed December 30, 2023. <https://doi.org/10.1017/pan.2023.19>.
- Kaufmann, Karen M. 2004. *The Urban Voter: Group Conflict and Mayoral Voting Behavior in American Cities*. University of Michigan Press, January. ISBN: 978-0-472-06857-9.
- Kerkvliet, Lennart, Jaap Kamps, and Maarten Marx. 2020. “Who Mentions Whom? Recognizing Political Actors in Proceedings.” In *Proceedings of the Second ParlaCLARIN Workshop*, edited by Darja Fišer, Maria Eskevich, and Franciska de Jong, 35–39. Marseille, France: European Language Resources Association, May. ISBN: 979-10-95546-47-4.

- Kim, Geewook, Teakgyu Hong, Moonbin Yim, Jeongyeon Nam, Jinyoung Park, Jinyeong Yim, Wonseok Hwang, Sangdoo Yun, Dongyoon Han, and Seunghyun Park. 2022. *OCR-free Document Understanding Transformer*, arXiv:2111.15664, October. Accessed November 14, 2023. arXiv: [2111.15664 \[cs\]](#).
- Kitchens, Karin E. 2021. “Exit or Invest: Segregation Increases Investment in Public Schools.” *The Journal of Politics* 83, no. 1 (January): 71–86. ISSN: 0022-3816, 1468-2508, accessed September 21, 2021. <https://doi.org/10.1086/708916>.
- Knox, Dean, Christopher Lucas, and Wendy K. Tam Cho. 2022. “Testing Causal Theories with Learned Proxies.” *Annual Review of Political Science* 25, no. 1 (May): annurev-polisci-051120–111443. ISSN: 1094-2939, 1545-1577, accessed February 17, 2022. <https://doi.org/10.1146/annurev-polisci-051120-111443>.
- Krueger, Brian S. 2002. “Assessing the Potential of Internet Political Participation in the United States: A Resource Approach.” *American Politics Research* 30, no. 5 (September): 476–498. ISSN: 1532-673X, accessed August 24, 2023. <https://doi.org/10.1177/1532673X02030005002>.
- Lassen, David Dreyer, and Søren Serritzlew. 2011. “Jurisdiction Size and Local Democracy: Evidence on Internal Political Efficacy from Large-scale Municipal Reform.” *American Political Science Review* 105, no. 2 (May): 238–258. ISSN: 0003-0554, 1537-5943, accessed August 8, 2023. <https://doi.org/10.1017/S000305541100013X>.
- Laurer, Moritz, Wouter van Atteveldt, Andreu Casas, and Kasper Welbers. 2023. “Less Annotating, More Classifying: Addressing the Data Scarcity Issue of Supervised Machine Learning with Deep Transfer Learning and BERT-NLI.” *Political Analysis* (June): 1–17. ISSN: 1047-1987, 1476-4989, accessed October 4, 2023. <https://doi.org/10.1017/pan.2023.20>.
- Lazaridou, Angeliki, Elena Gribovskaya, Wojciech Stokowiec, and Nikolai Grigorev. 2022. *Internet-Augmented Language Models through Few-Shot Prompting for Open-Domain Question Answering*, arXiv:2203.05115, May. Accessed November 14, 2023. arXiv: [2203.05115 \[cs\]](#).
- Lenz, Gabriel S. 2009. “Learning and Opinion Change, Not Priming: Reconsidering the Priming Hypothesis.” *American Journal of Political Science* 53, no. 4 (October): 821–837. ISSN: 00925853, 15405907, accessed March 31, 2019. <https://doi.org/10.1111/j.1540-5907.2009.00403.x>.
- Levendusky, Matthew. 2010. *The Partisan Sort: How Liberals Became Democrats and Conservatives Became Republicans*. Chicago: University of Chicago Press. ISBN: 978-0-226-47364-2 978-0-226-47365-9.
- Lin, Stephanie, Jacob Hilton, and Owain Evans. 2022. *TruthfulQA: Measuring How Models Mimic Human Falsehoods*, arXiv:2109.07958, May. Accessed January 15, 2024. arXiv: [2109.07958 \[cs\]](#).

- Lynch, Julia, Michael Bernhard, and Daniel O'Neill. 2022. "Pandemic Politics." *Perspectives on Politics* 20, no. 2 (June): 389–394. ISSN: 1537-5927, 1541-0986, accessed June 22, 2022. <https://doi.org/10.1017/S1537592722000676>.
- Malhotra, Neil, and Alexander G. Kuo. 2008. "Attributing Blame: The Public's Response to Hurricane Katrina." *The Journal of Politics* 70, no. 1 (January): 120–135. ISSN: 0022-3816, 1468-2508, accessed April 13, 2020. <https://doi.org/10.1017/S0022381607080097>.
- Marschall, Melissa, Paru Shah, and Anirudh Ruhil. 2011. "The Study of Local Elections: Editors' Introduction: A Looking Glass into the Future." *PS: Political Science & Politics* 44, no. 1 (January): 97–100. ISSN: 1049-0965, 1537-5935, accessed January 14, 2024. <https://doi.org/10.1017/S1049096510001940>.
- Miller, Blake, Fridolin Linder, and Walter R. Mebane. 2020. "Active Learning Approaches for Labeling Text: Review and Assessment of the Performance of Active Learning Approaches." *Political Analysis* 28, no. 4 (October): 532–551. ISSN: 1047-1987, 1476-4989, accessed June 16, 2023. <https://doi.org/10.1017/pan.2020.4>.
- Minozzi, William, Ryan Kennedy, Kevin M. Esterling, Michael A. Neblo, and Ryan Jewell. 2023. "Testing the Benefits of Public Deliberation." *American Journal of Political Science* (March): ajps.12775. ISSN: 0092-5853, 1540-5907, accessed July 27, 2023. <https://doi.org/10.1111/ajps.12775>.
- Mohr, Zachary, JoEllen V. Pope, Martha E. Kropf, and Mary Jo Shepherd. 2019. "Strategic Spending: Does Politics Influence Election Administration Expenditure?" *American Journal of Political Science* 63, no. 2 (April): 427–438. ISSN: 0092-5853, 1540-5907, accessed February 27, 2020. <https://doi.org/10.1111/ajps.12422>.
- Moore, Aaron A. 2022. "Motivations for Mobilization: Comparing Urban and Suburban Residents' Participation in the Politics of Planning and Development." *Urban Affairs Review* 58, no. 4 (July): 1124–1151. ISSN: 1078-0874, 1552-8332, accessed June 22, 2022. <https://doi.org/10.1177/10780874211016939>.
- Morel, Domingo. 2021. "How Community Engagement Helped Transform the Union City Public Schools." *Peabody Journal of Education* 96, no. 3 (May): 330–340. ISSN: 0161-956X, 1532-7930, accessed June 30, 2023. <https://doi.org/10.1080/0161956X.2021.1943238>.
- . 2018. *Takeover: Race, Education, and American Democracy*. New York, NY: Oxford University Press. ISBN: 978-0-19-067898-2 978-0-19-067897-5.

- Mozer, Reagan, Luke Miratrix, Aaron Russell Kaufman, and L. Jason Anastasopoulos. 2019. “Matching with Text Data: An Experimental Evaluation of Methods for Matching Documents and of Measuring Match Quality.” *arXiv:1801.00644 [cs, stat]* (March). Accessed September 21, 2021. arXiv: [1801.00644 \[cs, stat\]](https://arxiv.org/abs/1801.00644).
- Nabatchi, Tina, and Lisa Blomgren Amsler. 2014. “Direct Public Engagement in Local Government.” *The American Review of Public Administration* 44, no. 4_suppl (July): 63S–88S. ISSN: 0275-0740, 1552-3357, accessed July 4, 2023. <https://doi.org/10.1177/0275074013519702>.
- Nemerever, Zoe. 2021. “Contentious Federalism: Sheriffs, State Legislatures, and Political Violence in the American West.” *Political Behavior* 43, no. 1 (March): 247–270. ISSN: 0190-9320, 1573-6687, accessed September 21, 2021. <https://doi.org/10.1007/s11109-019-09553-w>.
- Nuamah, Sally A. 2021. “The Cost of Participating While Poor and Black: Toward a Theory of Collective Participatory Debt.” *Perspectives on Politics* 19, no. 4 (December): 1115–1130. ISSN: 1537-5927, 1541-0986, accessed June 30, 2023. <https://doi.org/10.1017/S1537592720003576>.
- Nuamah, Sally A., and Thomas Ogorzalek. 2021. “Close to Home: Place-Based Mobilization in Racialized Contexts.” *American Political Science Review* 115, no. 3 (August): 757–774. ISSN: 0003-0554, 1537-5943, accessed March 25, 2022. <https://doi.org/10.1017/S0003055421000307>.
- O’connor, John, and Michael Baratz. 2004. “Some Assembly Required: The Application of State Open Meeting Laws to Email Correspondence.” *George Mason Law Review* 12 (January).
- Oliver, J. Eric. 2001. *Democracy in Suburbia*. Princeton University Press, August. ISBN: 978-0-691-08880-8.
- Oliver, J. Eric, Shang E. Ha, and Zachary Callen. 2012. *Local Elections and the Politics of Small-Scale Democracy*. PRINCETON UNIVERSITY PRESS. ISBN: 978-0-691-14355-2.
- Ornstein, Joseph T, Elise N Blasingame, and Jake S Truscott. n.d. “How to Train Your Stochastic Parrot: Large Language Models for Political Texts.”
- Ouyang, Long, Jeff Wu, Xu Jiang, Diogo Almeida, Carroll L. Wainwright, Pamela Mishkin, Chong Zhang, et al. 2022. *Training Language Models to Follow Instructions with Human Feedback*, arXiv:2203.02155, March. Accessed January 7, 2024. arXiv: [2203.02155 \[cs\]](https://arxiv.org/abs/2203.02155).
- Payson, Julia, Alexander Fourinaies, and Andrew B. Hall. 2023. “Decomposing the Source of the Gender Gap in Legislative Committee Service: Evidence from US States.” *Political Science Research and Methods* 11, no. 1 (January): 191–197. ISSN: 2049-8470, 2049-8489, accessed August 16, 2023. <https://doi.org/10.1017/psrm.2021.72>.

- Payson, Julia A. 2017. "When Are Local Incumbents Held Accountable for Government Performance? Evidence from US School Districts: Accountability and School District Performance." *Legislative Studies Quarterly* 42, no. 3 (August): 421–448. ISSN: 03629805, accessed December 5, 2020. <https://doi.org/10.1111/lsq.12159>.
- Permaloff, Anne, and Carl Grafton. 1992. "Optical Character Recognition." *PS: Political Science and Politics* 25 (3): 523–531. ISSN: 10490965, 15375935, accessed January 14, 2024. <https://doi.org/10.2307/419444>. JSTOR: 419444.
- Peterson, Paul E. 1981. *City Limits*. University of Chicago Press, July. ISBN: 0-226-66293-4.
- Petrocik, John R. 1996. "Issue Ownership in Presidential Elections, with a 1980 Case Study." *American Journal of Political Science* 40 (3): 825–850. ISSN: 0092-5853, accessed September 25, 2018. <https://doi.org/10.2307/2111797>. JSTOR: 2111797.
- Pew Research Center. 2017. "Partisans Differ Widely in Views of Police Officers, College Professors," 16.
- Pin, Laura. 2022. "Bridging the Gap between Electoral and Participatory Democracy: The Electoral Motivations behind Participatory Budgeting in Chicago." *Urban Affairs Review* 58, no. 2 (March): 388–418. ISSN: 1078-0874, 1552-8332, accessed June 22, 2022. <https://doi.org/10.1177/1078087420964871>.
- Prior, Markus. 2013. "Media and Political Polarization." *Annual Review of Political Science* 16 (1): 101–127. <https://doi.org/10.1146/annurev-polisci-100711-135242>. eprint: <https://doi.org/10.1146/annurev-polisci-100711-135242>.
- Radford, Alec, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. 2019. "Language Models Are Unsupervised Multitask Learners."
- Roberts, Nancy. 1997. "Public Deliberation: An Alternative Approach to Crafting Policy and Setting Direction." *Public Administration Review* 57 (2): 124–132. ISSN: 0033-3352, accessed July 27, 2023. <https://doi.org/10.2307/977060>. JSTOR: 977060.
- Sances, Michael W. 2021. "When Voters Matter: The Limits of Local Government Responsiveness." *Urban Affairs Review* 57, no. 2 (March): 402–427. ISSN: 1078-0874, 1552-8332, accessed June 22, 2022. <https://doi.org/10.1177/1078087419878812>.
- Scala, Dante J., and Kenneth M. Johnson. 2017. "Political Polarization along the Rural-Urban Continuum? The Geography of the Presidential Vote, 2000–2016." *The ANNALS of the American Academy of Political and Social Science* 672, no. 1 (July): 162–184. ISSN: 0002-7162, accessed January 13, 2024. <https://doi.org/10.1177/0002716217712696>.

- Scao, Teven Le, BigScience Workshop, Angela Fan, Christopher Akiki, Ellie Pavlick, Suzana Ilić, Daniel Hesslow, et al. 2023. *BLOOM: A 176B-Parameter Open-Access Multilingual Language Model*, arXiv:2211.05100, June. Accessed January 14, 2024. arXiv: [2211.05100](https://arxiv.org/abs/2211.05100) [cs].
- Schlozman, Kay Lehman, Sidney Verba, and Henry E. Brady. 2012. *The Unheavenly Chorus: Unequal Political Voice and the Broken Promise of American Democracy*. Princeton ; Oxford: Princeton University Press. ISBN: 978-0-691-15484-8.
- Shapiro, Robert Y., Anja Kilibarda, Sofi Sinozich, and Oliver McClellan. 2021. “American Public Opinion and Partisan Conflict: Education’s Exceptionalism?” In *Public Opinion and the Political Economy of Education Policy around the World*, edited by Martin R. West and Ludger Woessmann, 0. The MIT Press, April. ISBN: 978-0-262-36348-8, accessed January 14, 2024. <https://doi.org/10.7551/mitpress/13712.003.0009>.
- Sinclair, B. 2006. *Party Wars: Polarization and the Politics of National Policy Making*. The Julian J. Rothbaum Distinguished Lecture Series. University of Oklahoma Press. ISBN: 978-0-8061-3779-7.
- Stone, Clarence N. 2001. “Civic Capacity and Urban Education.” *Urban Affairs Review* 36, no. 5 (May): 595–619. ISSN: 1078-0874, 1552-8332, accessed June 30, 2023. <https://doi.org/10.1177/10780870122185019>.
- Sumner, Jane Lawrence, Emily M. Farris, and Mirya R. Holman. 2020. “Crowdsourcing Reliable Local Data.” *Political Analysis* 28, no. 2 (April): 244–262. ISSN: 1047-1987, 1476-4989, accessed February 17, 2022. <https://doi.org/10.1017/pan.2019.32>.
- Sun, Liyang, and Sarah Abraham. 2021. “Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects.” *Journal of Econometrics*, Themed Issue: Treatment Effect 1, 225, no. 2 (December): 175–199. ISSN: 0304-4076, accessed August 22, 2023. <https://doi.org/10.1016/j.jeconom.2020.09.006>.
- Thompson, Daniel M. 2020. “How Partisan Is Local Law Enforcement? Evidence from Sheriff Cooperation with Immigration Authorities.” *American Political Science Review* 114, no. 1 (February): 222–236. ISSN: 0003-0554, 1537-5943, accessed September 8, 2020. <https://doi.org/10.1017/S0003055419000613>.
- Thoppilan, Romal, Daniel De Freitas, Jamie Hall, Noam Shazeer, Apoorv Kulshreshtha, Heng-Tze Cheng, Alicia Jin, et al. 2022. *LaMDA: Language Models for Dialog Applications*, arXiv:2201.08239, February. Accessed January 7, 2024. arXiv: [2201.08239](https://arxiv.org/abs/2201.08239) [cs].
- Touvron, Hugo, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, et al. 2023. *LLaMA: Open and Efficient Foundation Language Models*, arXiv:2302.13971, February. Accessed January 7, 2024. arXiv: [2302.13971](https://arxiv.org/abs/2302.13971) [cs].

- Tracy, Karen, and Margaret Durfy. 2007. "Speaking out in Public: Citizen Participation in Contentious School Board Meetings." *Discourse & Communication* 1, no. 2 (May): 223–249. ISSN: 1750-4813, 1750-4821, accessed March 25, 2022. <https://doi.org/10.1177/1750481307076008>.
- Trounstone, Jessica. 2010. "Representation and Accountability in Cities." *Annual Review of Political Science* 13, no. 1 (May): 407–423. ISSN: 1094-2939, 1545-1577, accessed April 13, 2020. <https://doi.org/10.1146/annurev.polisci.032808.150414>.
- Urban Institute. 2020. "State and Local Finance Initiative: Elementary and Secondary Education Expenditures."
- Van Strien, Daniel, Kaspar Beelen, Mariona Ardanuy, Kasra Hosseini, Barbara McGillivray, and Giovanni Colavizza. 2020. "Assessing the Impact of OCR Quality on Downstream NLP Tasks:" in *Proceedings of the 12th International Conference on Agents and Artificial Intelligence*, 484–496. Valletta, Malta: SCITEPRESS - Science and Technology Publications. ISBN: 978-989-758-395-7, accessed December 30, 2023. <https://doi.org/10.5220/0009169004840496>.
- Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. *Attention Is All You Need*, arXiv:1706.03762. Accessed January 7, 2024. arXiv: [1706.03762 \[cs\]](https://arxiv.org/abs/1706.03762).
- Verba, Sidney, Kay Lehman Schlozman, and Henry E. Brady. 1995. *Voice and Equality: Civic Voluntarism in American Politics*. Harvard University Press, September. ISBN: 978-0-674-94293-6.
- Volden, Craig, Alan E. Wiseman, and Dana E. Wittmer. 2018. "Women's Issues and Their Fates in the US Congress." *Political Science Research and Methods* 6, no. 4 (October): 679–696. ISSN: 2049-8470, 2049-8489, accessed August 22, 2023. <https://doi.org/10.1017/psrm.2016.32>.
- Walgrave, Stefaan, Jonas Lefevere, and Michiel Nuytemans. 2009. "Issue Ownership Stability and Change: How Political Parties Claim and Maintain Issues Through Media Appearances." *Political Communication* 26, no. 2 (May): 153–172. ISSN: 1058-4609. <https://doi.org/10.1080/10584600902850718>.
- Walgrave, Stefaan, Jonas Lefevere, and Anke Tresch. 2012. "The Associative Dimension of Issue Ownership." *Public Opinion Quarterly* 76, no. 4 (June): 771–782. ISSN: 0033-362X. <https://doi.org/10.1093/poq/nfs023>. eprint: <https://academic.oup.com/poq/article-pdf/76/4/771/5386699/nfs023.pdf>.
- Walsh, Katherine Cramer. 2012. "Putting Inequality in Its Place: Rural Consciousness and the Power of Perspective." *American Political Science Review* 106, no. 3 (August): 517–532. ISSN: 0003-0554, 1537-5943, accessed September 16, 2019. <https://doi.org/10.1017/S0003055412000305>.

- Wang, Yu. 2023. "On Finetuning Large Language Models." *Political Analysis* (November): 1–5. ISSN: 1047-1987, 1476-4989, accessed December 30, 2023. <https://doi.org/10.1017/pan.2023.36>.
- Wichowsky, Amber, Paru Shah, and Amanda Heideman. 2022. "Call and Response? Neighborhood Inequality and Political Voice." *Urban Affairs Review* 58, no. 4 (July): 1182–1197. ISSN: 1078-0874, 1552-8332, accessed June 22, 2022. <https://doi.org/10.1177/10780874211021327>.
- Wiedemann, Gregor. 2019. "Proportional Classification Revisited: Automatic Content Analysis of Political Manifestos Using Active Learning." *Social Science Computer Review* 37, no. 2 (April): 135–159. ISSN: 0894-4393, 1552-8286, accessed January 7, 2024. <https://doi.org/10.1177/0894439318758389>.
- Wilkerson, John, and Andreu Casas. 2017. "Large-Scale Computerized Text Analysis in Political Science: Opportunities and Challenges." *Annual Review of Political Science* 20, no. 1 (May): 529–544. ISSN: 1094-2939, 1545-1577, accessed March 25, 2022. <https://doi.org/10.1146/annurev-polisci-052615-025542>.
- Wilson, Anne E, Victoria A Parker, and Matthew Feinberg. 2020. "Polarization in the Contemporary Political and Media Landscape." *Political Ideologies* 34 (August): 223–228. ISSN: 2352-1546. <https://doi.org/10.1016/j.cobeha.2020.07.005>.
- Yoder, Jesse. 2020. "Does Property Ownership Lead to Participation in Local Politics? Evidence from Property Records and Meeting Minutes." *American Political Science Review* 114, no. 4 (November): 1213–1229. ISSN: 0003-0554, 1537-5943, accessed September 11, 2022. <https://doi.org/10.1017/S0003055420000556>.
- Zaller, John. 1992. *The Nature and Origins of Mass Opinion*.
- Zhou, Kaitlyn, Kawin Ethayarajh, and Dan Jurafsky. 2021. *Frequency-Based Distortions in Contextualized Word Embeddings*, arXiv:2104.08465, April. Accessed January 7, 2024. arXiv: [2104.08465](https://arxiv.org/abs/2104.08465) [cs].

Appendix A:

Appendix For Is a More Accessible Local Government a More Representative One? The Effects of Online Public Meetings

A.1 Additional Sample Details:

Within the section below, I include details of my primary sample of local governments, including their names, socio-economic makeup, and meeting access over the course of my study.

A.1.1 Included Local Governments:

The subsection below details the local governments included within my sample. Meeting minutes were gathered from each government's official website or data repository.

County:

- St. Louis County Government

Cities:

- | | |
|--------------------|--------------|
| • St. Ann | • Maplewood |
| • Richmond Heights | • Ballwin |
| • Kirkwood | • Bridgeton |
| • Ellisville | • Hazelwood |
| • Creve Coeur | • Des Peres |
| • Brentwood | • Manchester |
| • Chesterfield | |

School Districts:

- | | |
|----------------|--------------------------|
| • Parkway SD | • Ferguson-Florissant SD |
| • Kirkwood SD | • Pattonville SD |
| • Brentwood SD | • Hazelwood SD |

- Ritenour SD
- Riverview Gardens SD
- Ladue SD
- Affton SD
- University City SD
- Clayton SD
- Rockwood SD
- Webster Groves SD
- Maplewood-Richmond Heights SD

A.1.2 Demographic Characteristics:

All of the following demographic estimates come from the 2020 American Community Survey.

Table A.1: Average Demographic Characteristics of Sample Governments

Variable	County	City	School
Population	1001982.00	21051.55	52118.20
Gender (F)	47.77	47.51	47.68
White	65.25	65.70	65.30
Black	24.18	22.59	22.80
Asian	4.55	4.87	5.12
Other	1.36	1.79	1.67
Age (45+)	44.41	43.38	42.07
Education (College+)	45.29	48.10	51.42
Median Income	72562.00	81979.77	82670.27
Per Capita Income	45307.00	46662.45	49951.33
Gini Index	0.50	0.44	0.46
Median Structure Year	1968	1966	1961

A.1.3 Access By Level of Government

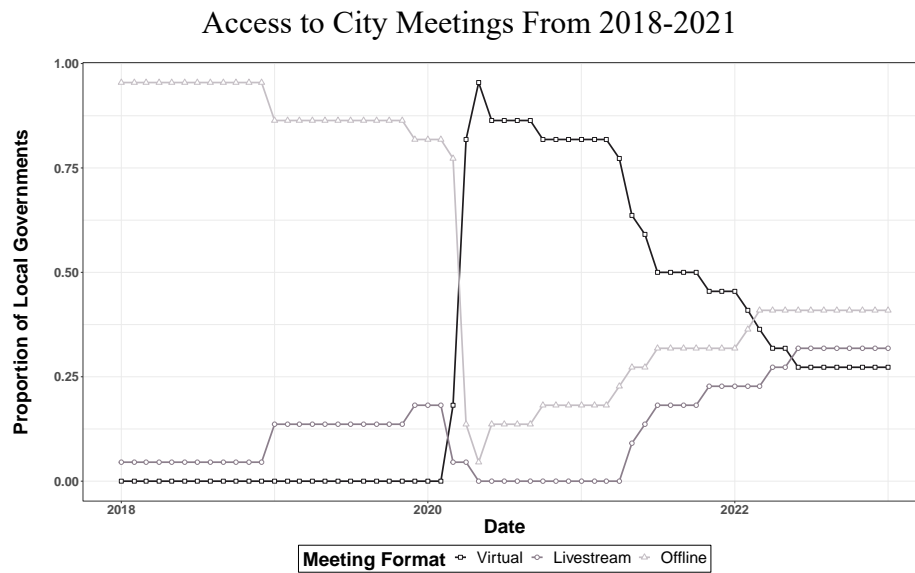


Figure A.1: Access to City Meetings From 2018-2021

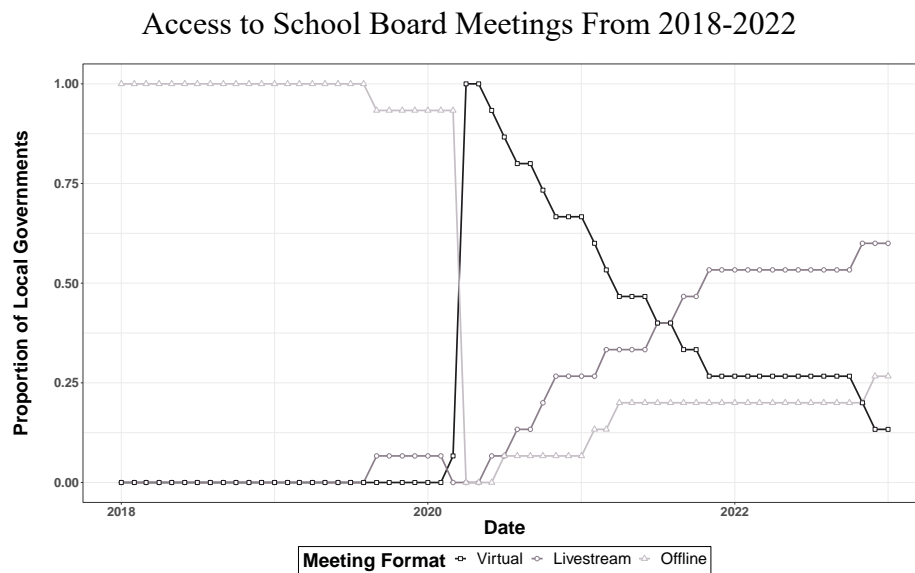


Figure A.2:

Access to County Council Meetings From 2018-2022

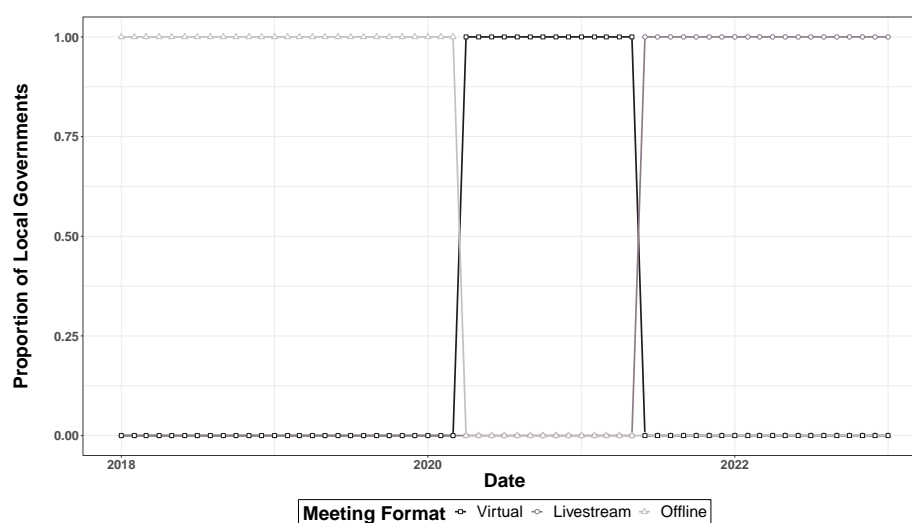


Figure A.3:

A.2 Full Tables For Figures:

A.2.1 Figure 2 Regression Tables:

Table A.2: Full Regression Results for Figure 2: Virtual

	Pooled	Number of Commenters		
		County	School District	City
Virtual	0.368* (0.176)	-0.129 (0.363)	0.330 (0.255)	0.306 (0.260)
Constant	-0.968* (0.415)	3.707* (0.240)	-1.904* (0.673)	-0.019 (0.449)
Year FE		✓		
Month FE	✓		✓	✓
Government FE	✓		✓	✓
Num. Gov	38	1	22	15
N	1800	52	780	968
AIC	5421.063	550.832	2614.811	2236.291

*p<0.05

Table A.3: Full Regression Results for Figure 2: Livestreamed

	Number of Commenters			
	Pooled	County	School District	City
Livestreamed	−0.137 (0.178)	1.139* (0.528)	−0.363 (0.256)	0.261 (0.290)
Constant	−0.998* (0.419)	3.707* (0.261)	−1.882* (0.674)	−0.017 (0.449)
Year FE		✓		
Month FE	✓		✓	✓
Government FE	✓		✓	✓
Num. Gov	38	1	22	15
N	1800	52	780	968
AIC	5434.779	564.507	2614.505	2236.835

*p<0.05

Table A.4: Full Regression Results for Figure 2: Online

	Number of Commenters			
	Pooled	County	School District	City
Online	0.246 (0.181)	2.483* (0.566)	−0.050 (0.346)	0.293 (0.203)
Constant	−1.048* (0.419)	3.707* (0.247)	−1.941* (0.675)	−0.004 (0.449)
Year FE		✓		
Month FE	✓		✓	✓
Government FE	✓		✓	✓
Num. Gov	38	1	22	15
N	1800	52	780	968
AIC	5433.647	557.457	2616.083	2235.560

*p<0.05

A.2.2 Figure 3 Regression Tables

Table A.5: Full Regression Results of Switching Away from Virtual Meetings

Group	Estimate	Std. Error	[95% Pointwise Conf. Band]
2	-6.0815	1.8841	[-9.7743, -2.3886]*
3	4.4265	7.9697	[-11.1939, 20.0469]
4	-8.5765	2.5928	[-13.6583, -3.4946]*
6	-29.9583	3.5539	[-36.9239, -22.9927]*
7	1.4810	5.0879	[-8.4911, 11.4530]
10	-2.0547	1.6289	[-5.2472, 1.1378]
11	1.4188	2.4016	[-3.2883, 6.1259]
13	-0.7655	1.3506	[-3.4126, 1.8816]
15	-0.4200	1.4075	[-3.1786, 2.3386]
17	4.6250	2.8540	[-0.9688, 10.2188]
19	-11.5000	0.7413	[-12.9529, -10.0471]*

Table A.6: Full Regression Results of Switching Away from Online Meetings

Group	Estimate	Std. Error	[95% Pointwise Conf. Band]
3	1.2063	4.3231	[-7.2669, 9.6794]
10	-2.6389	1.8715	[-6.3070, 1.0292]
13	-2.5486	2.1956	[-6.8519, 1.7547]

A.2.3 Figure 4 Regression Tables

Table A.7: Full Regression Results of Switching Away from Virtual Meetings

Group	Estimate	Std. Error	[95% Pointwise Conf. Band]
2	-0.4765	0.9735	[-2.4581, 1.5051]
6	1.1591	0.2685	[0.6124, 1.7057]*
12	1.1461	0.3533	[0.4270, 1.8652]*
13	-0.1097	0.4969	[-1.1213, 0.9018]
14	-0.7552	0.7571	[-2.2964, 0.7859]
15	1.0909	1.4397	[-1.8398, 4.0216]

Table A.8: Full Regession Results of Switching Away from Online Meetings

Group	Estimate	Std. Error	[95% Pointwise Conf. Band]
2	-0.4508	1.0045	[-2.4292, 1.5276]
6	1.2019	0.2565	[0.6967, 1.7071]*
12	1.2030	0.3007	[0.6107, 1.7953]*
13	-0.2583	0.3845	[-1.0156, 0.4990]
15	0.7000	1.1169	[-1.4998, 2.8998]

A.3 Supplementary Analyses:

A.3.1 Simultaneous Comparisons of Commenters to Voters

Table A.9: Pooled Commenters Compared to Voters

	<i>Dependent variable:</i>			
	Pooled	County	City	School
	(1)	(2)	(3)	(4)
White	0.177* (0.062)	0.154 (0.081)	0.259* (0.128)	0.106 (0.149)
Black	−0.089 (0.121)	−0.028 (0.164)	−0.493 (0.378)	−0.144 (0.219)
Gender (F)	0.492* (0.046)	0.760* (0.064)	−0.171 (0.088)	0.805* (0.120)
Income (100K+)	0.247* (0.047)	0.305* (0.062)	0.029 (0.094)	0.443* (0.117)
Education (HS+)	−0.048 (0.058)	−0.068 (0.077)	0.157 (0.123)	−0.172 (0.136)
Homeowner	0.530* (0.068)	0.500* (0.091)	0.682* (0.142)	0.636* (0.149)
National Voter	0.785* (0.078)	0.837* (0.100)	0.825* (0.174)	0.622* (0.193)
Local Voter	1.104* (0.048)	0.860* (0.062)	1.403* (0.104)	1.439* (0.123)
Constant	−7.778* (0.189)	−8.083* (0.229)	−8.639* (0.303)	−9.214* (0.374)
Zipcode FE	✓	✓	-	-
City FE	-	-	✓	-
School FE	-	-	-	✓
Observations	632,187	632,187	171,772	327,165
Akaike Inf. Crit.	25,908.220	16,388.060	6,302.015	4,858.056

Note:

*p<0.05

Table A.10: Online and Offline Commenters Compared to All Voters

	<i>Dependent variable:</i>							
	Pooled		County		City		School	
	(Online)	(Offline)	(Online)	(Offline)	(Online)	(Offline)	(Online)	(Offline)
White	0.149 (0.077)	0.253* (0.109)	0.097 (0.087)	0.474* (0.220)	0.409 (0.246)	0.230 (0.159)	0.319 (0.218)	-0.106 (0.209)
Black	-0.189 (0.159)	0.105 (0.193)	-0.128 (0.201)	0.333 (0.315)	-0.046 (0.558)	-0.701 (0.527)	-0.218 (0.305)	0.018 (0.339)
Gender (F)	0.792* (0.061)	0.059 (0.075)	0.942* (0.074)	0.043 (0.137)	-0.224 (0.159)	-0.076 (0.111)	1.073* (0.174)	0.557* (0.175)
Income (100K+)	0.325* (0.059)	0.130 (0.082)	0.383* (0.068)	-0.100 (0.158)	-0.080 (0.172)	0.133 (0.119)	0.351* (0.158)	0.531* (0.177)
Education (HS+)	-0.092 (0.072)	0.056 (0.101)	-0.110 (0.084)	0.137 (0.192)	0.208 (0.223)	0.146 (0.155)	-0.236 (0.180)	-0.017 (0.217)
Homeowner	0.577* (0.087)	0.411* (0.109)	0.590* (0.105)	0.181 (0.181)	0.630* (0.242)	0.658* (0.180)	0.624* (0.200)	0.667* (0.233)
National Voter	0.755* (0.094)	0.849* (0.144)	0.778* (0.108)	1.145* (0.259)	0.778* (0.296)	0.907* (0.228)	0.650* (0.246)	0.538 (0.324)
Local Voter	0.903* (0.060)	1.443* (0.087)	0.801* (0.069)	1.127* (0.149)	1.228* (0.182)	1.407* (0.132)	1.150* (0.158)	1.908* (0.209)
Constant	-8.088* (0.222)	-9.405* (0.391)	-8.183* (0.241)	-24.595 (410.557)	-10.272* (0.600)	-9.290* (0.391)	-9.580* (0.469)	-10.911* (0.717)
Zipcode FE	✓	✓	✓	✓	-	-	-	-
City FE	-	-	-	-	✓	✓	-	-
School FE	-	-	-	-	-	-	✓	✓
Observations	632,187	632,187	632,187	632,187	171,772	171,772	327,165	327,165
Akaike Inf. Crit.	17,646.340	10,154.080	13,628.920	3,802.738	2,261.061	4,212.815	2,921.446	2,248.457

Note:

*p<0.05

A.4 Text-Analysis of Comments

A.4.1 Examining the Content of Public Meeting Comments

In the following section I construct a separate STM model (K=20) for each level government within my sample (Roberts et al. 2019). Each model includes government and month level controls. For

the county model, I replace government controls with zipcode controls. Table A.11 presents the top 5 topics for each level of government.

Table A.11: STM: Comment Topics By Level of Government

County Council	Municipal Council	School Board
mask, mandat, counti, wear, vaccin	develop, light, traffic, concern	comment, school, chromebook, instruct
sport, counti, youth, page, support	spoke, propos, opposit, favor, rezon	earn, virtual, person, graduat, shorten
school, kid, play, children, team	contract, renew, trash, waste	propos, oppos, rezon, concern, spoke
prosecuti, judge, attorney, file, process	concern, construct, express, sidewalk	return, school, staff, online, learn
project, develop, tax, propos, resid	maryland, trail, thank, event	spoke, regard, favor, concern, kirkwood

Next I follow the same procedure but subset each level of my sample by whether a meeting was accessible online or not. I then reconstruct a model for each subset across each level of government. Table A.12, Table A.13, and Table A.14 display the top four topics for online and offline comments at the county, city, and school district levels. While the exact topics change between online and offline meetings, most topics appear jurisdictionally relevant to their respective level of government.

Table A.12: County Topics of Online Versus Offline Comments

Online	Offline
mask, mandat, vaccin, counti, pleas	employe, counti, year, work, servic
sport, play, counti, loui, school	anim, volunt, dog, shelter, counti
counti, bill, council, page, power	counti, loui, council, execut, meet
counti, loui, tax, propos, properti	tax, million, counti, loui, zoo

Table A.13: City Topics of Online Versus Offline Comments

Online	Offline
council, address, street, highland, terrac	spoke, propos, opposit, comment, concern
state, propos, oppos, develop, favor	develop, concern, oppos, state, traffic
van, mark, way, space, rezon	ask, council, citi, servic, note, approach
food, unhous, support, mini, pantri	polic, protest, charg, galleria, peac

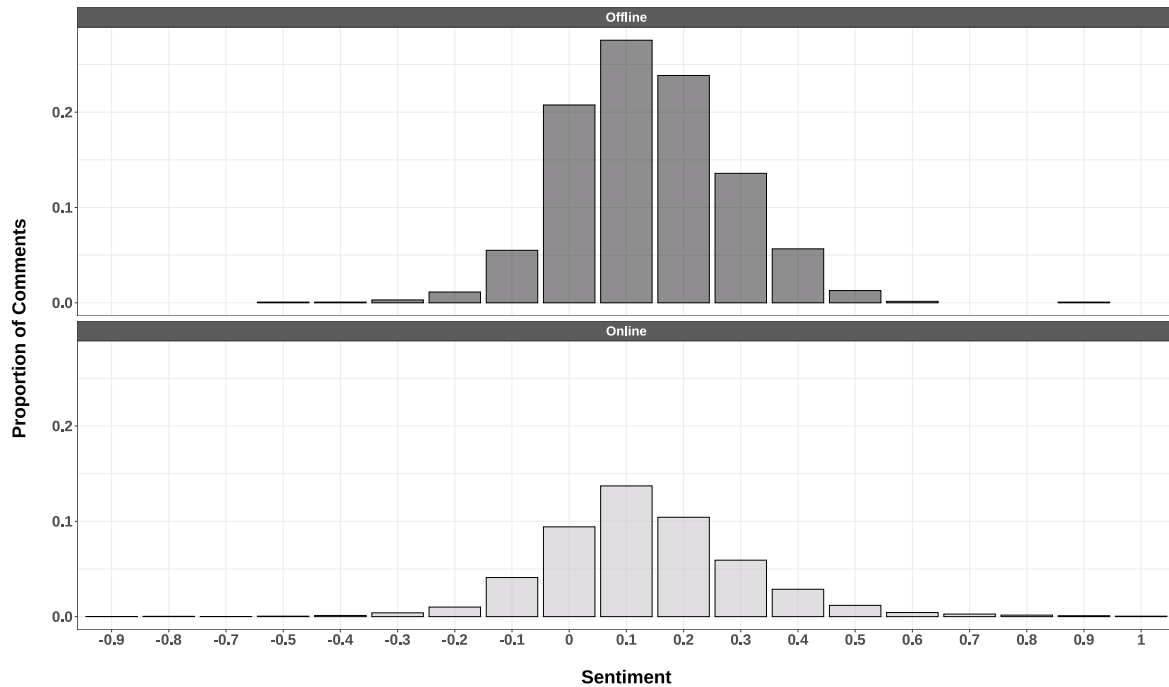
Table A.14: School Board Topics of Online Versus Offline Comments

Online	Offline
patron, comment, total, various, read	gender, ident, ave, geyer, kepling
graduat, ask, student, parent, ceremoni	comment, polici, kelli, meet, advoc
learn, virtual, comment, return, -person	school, spoke, high, pattonvill, student
school, year, plan, involv, back-school	board, educ, spoke, member, regard

A.4.2 Do Online Meetings Elicit More Negative Comments?

Beyond the topics of the comments, I also examine whether online meetings elicit more negative comments. To analyze this, I calculate each comment's average sentence level VADER sentiment scores and scale it between -1 and 1 (Hutto and Gilbert 2014). Overall, average sentiment for Offline (0.100) and Online (0.099) meetings are both positive, but not statistically different.

Figure A.4: Average Comment Sentiment By Meeting Format



A.5 Survey Details:

A.5.1 Question Wording:

Non-Meeting Attendees: What best describes why you haven't attended a local political meeting (such as school district or city council) this past year?

1. Inconvenient meeting times
2. No compelling reason or issue
3. Unable to join the meeting (either in-person or virtually)
4. Didn't know how to attend meetings
5. Didn't want to attend meetings alone
6. Voiced concerns in an alternative way (email, phone call, etc.)
7. Other (please specify)

Meeting Attendees: What best describes why you attended a local political meeting (such as school district or city council) this past year?

1. Support a specific policy issue.
2. Oppose a specific policy issue.
3. Voice an opinion or concern.
4. Thank a local public official.

5. Critique a local public official.
6. Observe the meeting
7. Other (please specify)

A.5.2 Survey Responses:

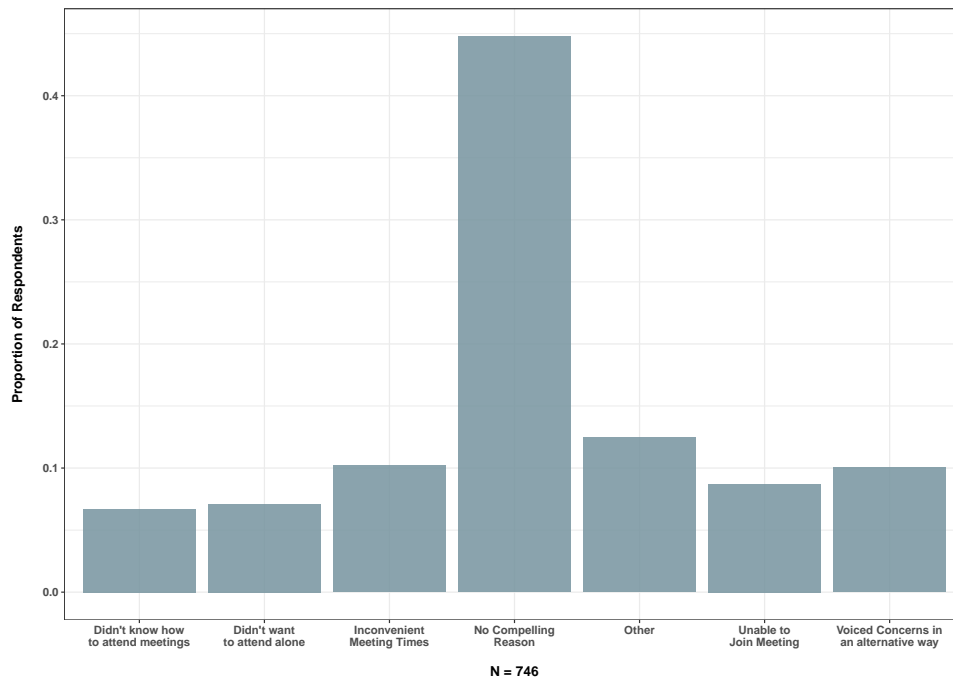


Figure A.5: Response to Non-Meeting Attendees Question

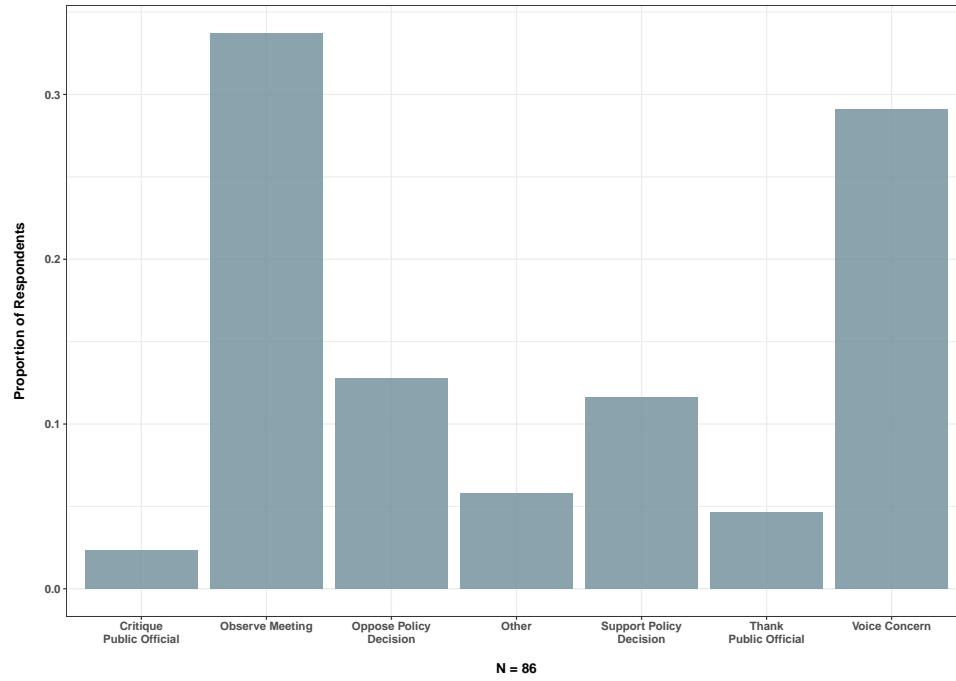


Figure A.6: Response to Meeting Attendees Question

Appendix B:

Appendix For Is it All Good in the Neighborhood?

How Partisanship May Shape Evaluations of Municipal Services

B.1 Survey Details

B.1.1 Variable Distributions

Table B.1: Descriptive Statistics of Independent and Dependent Variables: Full CCES

	Mean	Median	SD	Min	Max
Individual Covariates:					
Age	48.680	49	17.608	18	93
Gender (M)	0.445	0	0.497	0	1
Parent	0.658	1	0.475	0	1
Republican	0.377	0	0.485	0	1
Independent	0.139	0	0.346	0	1
Black	0.079	0	0.270	0	1
Hispanic	0.087	0	0.282	0	1
Asian	0.029	0	0.169	0	1
Homeowner	0.622	1	0.485	0	1
Income	6.412	6	3.306	1	16
Rural	0.150	0	0.357	0	1
Objective Measures:					
School Rating	2.651	2.605	0.826	0	5
Crime Rate (2018)	482.325	300.771	534.689	0	3,063.617
Crime Rate (2017)	576.387	355.356	658.426	0	13,124.290
Crime Rate Difference	-97.506	-27.720	259.851	-2,013.473	1,195.273
Service Evaluations:					
School	0.335	0	0.998	-2	2
Police	0.458	0	0.931	-2	2

Table B.2: Difference in Covariate Means Between CCES and Large Cities Subset

	CCES	Large City Subset	Δ	P-Value
Demographics:				
Age	48.667	46.537	2.130	0
Gender (M)	0.451	0.459	-0.008	0.156
Parent	0.654	0.572	0.082	0
Black	0.091	0.145	-0.054	0
Hispanic	0.080	0.124	-0.044	0
Asian	0.028	0.041	-0.013	0
Homeowner	0.630	0.536	0.094	0
Rural Res.	0.191	0.030	0.160	0
Income	6.520	6.530	-0.010	0.786
Partisanship:				
Republican	0.385	0.279	0.106	0
Independent	0.122	0.111	0.011	0.001
Objective Measures:				
School Rating	2.628	2.271	0.357	0
Violent Crime Rate	82.940	69.789	13.151	0
Property Crime Rate	462.579	395.580	66.999	0

Note: The above table presents results of a two sample t-test between the 2018 CCES and the new merged data set. Covariates such as Republican, Independent, and Black are coded as a binary choice. Income (1-10) is measured on its own categorical scale. Crime rates are standardized to incidents per 100,000 citizens.

B.2 Supplemental Analyses

B.2.1 Full Regression Results For Table 1

Table B.3: OLS Regression of Local School and Police Evaluations on Partisan Identity: Full Results

	<i>Service Evaluation:</i>	
	Police	School
Republican	0.264* (0.018)	-0.035 (0.028)
Property Crime Rate	0.0000 (0.00003)	
Violent Crime Rate	-0.0004* (0.0001)	
School Rating		0.340* (0.014)
Age	0.007* (0.0005)	0.0002 (0.001)
Gender (M)	0.004 (0.010)	0.034* (0.011)
Parent	0.018 (0.012)	0.087* (0.015)
Education	0.013* (0.005)	0.006 (0.004)
Rural Res.	-0.175* (0.018)	-0.028 (0.017)
Homeowner	0.053* (0.012)	0.032* (0.015)
Black	-0.282* (0.033)	-0.016 (0.035)
Hispanic	-0.038* (0.015)	0.040 (0.027)
Asian	0.060 (0.046)	0.118* (0.048)
Income	0.025* (0.002)	0.016* (0.002)
State Fixed Effects	✓	✓
Observations	29,020	34,812
R^2	0.078	0.103
Adjusted R^2	0.076	0.101
F Statistic	187.880 (df = 13; 28958)	332.247 (df = 12; 34749)

Note: The above table presents the results of a OLS regression with state fixed effects and clustered standard errors. Crime rates are standardized to incident per 100,000 citizens. Standard errors are presented in parentheses. *p<0.05

Table B.4: Hierarchical Linear Regression of Local Roads Evaluation on Partisan Identity: Full Results

	<i>Service Evaluation:</i>
	Roads
Republican	0.030 (0.026)
Traffic Index	−0.003 (0.003)
Age	−0.005* (0.001)
Gender (M)	0.020 (0.022)
Parent	0.049* (0.024)
Education	0.016 (0.008)
Rural	−0.102 (0.095)
Homeowner	0.075* (0.025)
Black	−0.175* (0.033)
Hispanic	−0.019 (0.036)
Asian	−0.036 (0.056)
Income	0.016* (0.004)
Constant	−0.003 (0.077)
Observations	6,928
State Random Effects	✓
Municipality Random Effects	✓
AIC	18,191.980

Note: The above table presents the results of a hierarchical linear regression with random effects. All respondents are nested within municipalities within states. Standard errors are presented in parentheses.*p<0.05

B.2.2 Alternative Modeling Specifications

Table B.5: OLS Regression of Local School and Police Evaluations on Partisan Identity: Full Results and Independents

	<i>Service Evaluation:</i>	
	Police	School
Republican	0.261* (0.018)	-0.037 (0.028)
Independent	-0.048* (0.018)	-0.185* (0.021)
Property Crime Rate	-0.00000 (0.00003)	
Violent Crime Rate	-0.0004* (0.0001)	
School Rating		0.341* (0.013)
Age	0.007* (0.0004)	0.0003 (0.001)
Gender (M)	-0.001 (0.009)	0.029* (0.011)
Parent	0.021 (0.011)	0.089* (0.014)
Education	0.012* (0.005)	0.004 (0.004)
Rural Res.	-0.165* (0.017)	-0.020 (0.016)
Homeowner	0.055* (0.013)	0.034* (0.014)
Black	-0.289* (0.033)	-0.022 (0.032)
Hispanic	-0.046* (0.014)	0.048 (0.026)
Asian	0.073 (0.040)	0.115* (0.044)
Income	0.026* (0.002)	0.017* (0.002)
Observations	32,994	39,548
R^2	0.078	0.106
Adjusted R^2	0.077	0.104

Note: The above table presents the results of a OLS regression with state fixed effects and clustered standard errors. Crime rates are standardized to incident per 100,000 citizens. Standard errors are presented in parentheses. *p<0.05

B.2.3 Alternative Model Specifications

Table B.6: Hierarchical Linear Regression of 7-Point Party ID on Local School and Police Evaluations

	<i>Service Evaluation:</i>	
	Police	School
Party ID	0.050* (0.002)	−0.010* (0.002)
Crime Rate	−0.0001* (0.00001)	
School Rating		0.321* (0.007)
Age	0.007* (0.0003)	0.0005 (0.0003)
Gender (M)	−0.006 (0.010)	0.023* (0.010)
Parent	0.013 (0.011)	0.086* (0.011)
Education	0.013* (0.004)	0.004 (0.003)
Income	0.025* (0.002)	0.017* (0.002)
Rural	−0.179* (0.013)	−0.045* (0.013)
Homeowner	0.051* (0.012)	0.030* (0.011)
Black	−0.251* (0.020)	0.002 (0.018)
Hispanic	−0.032 (0.020)	0.057* (0.019)
Asian	0.069* (0.032)	0.118* (0.030)
Constant	−0.254* (0.031)	−0.682* (0.044)
State Fixed Effects	✓	✓
Municipality Fixed Effects	✓	✓
Observations	32,994	39,548
AIC	86,091.060	106,191.300

Note: The above table presents the results of a hierarchical linear regression with random effects. All respondents are nested within municipalities within states. Standard errors are presented in parentheses. *p<0.05

Table B.7: Hierarchical Regression Results With Alternative Measures of Police Performance

	(1)	(2)
2017 Crime Rate	0.00001 (0.00003)	
2018 Crime Rate	−0.0001* (0.00004)	
Crime Rate Difference		0.00000 (0.00003)
Constant	−0.138* (0.031)	−0.186* (0.030)
State Fixed Effects	✓	✓
Municipality Fixed Effects	✓	✓
Additional Controls	✓	✓
Observations	32,115	32,115
AIC	83,722.270	83,730.040

Note: The above table presents the results of a hierarchical linear regression with random effects. Model (1) includes both current and lagged crime rates. Model (2) includes only the year change in crime rates. All respondents are nested within municipalities within states. Standard errors are presented in parentheses. * $p < 0.05$

Appendix C:

Appendix For Utilizing Large Language Models To Transform Messy Text-Based Sources In Useful Data

C.1 Example Applications

C.1.1 Local Meeting Minutes:

The following section displays a selection of meeting minute excerpts and their resulting output from OpenAI's GPT 3.5 model. For each example, I prompted the model with the prompt displayed within the body of the main paper and then fed the entirety of the meeting text from which the excerpt originated into the model.

Meeting Text Excerpt #1

Middle school math teacher Megan Grossbauer, received the district's Excellence in Education Award for leadership and contributions to the learning environment for students in the district. Those in attendance viewed a video of Mrs. Grossbauer filmed by high school teacher Don Goble.

Turning to the superintendent's remarks, Dr. Wipke congratulated Megan on her well-deserved recognition and talked about his recent podcast recording which included Megan. Dr. Wipke went on to share brief updates related to the transition back to in-person learning for students and plans to remain aligned with St. Louis County Health Department recommendations as COVID-19 trends appear to be increasing. Dr. Wipke also highlighted the recent design meeting hosted by several architect team members from Ittner and Perkins & Will. Dr. Wipke commended Sue Blumm and her foodservice team for their quick planning to ensure all meals served at Old Bonhomme were uninterrupted as they were faced with a boil order last week.

President Jaeger called for Board committee reports. Dr. Buckner reported on the Finance and Facilities Board Advisory Committees.

President Jaeger called for superintendent committee reports. Dr. Buckner reported on the Superintendent's Finance and Facilities Committee.

One patron, Andy Cohen, addressed the Board during the time period set aside for public comments regarding agenda and non-agenda items.

Turning to the action agenda, Mrs. Jaeger called for a motion to approve the evening's consent agenda. On a motion by Mr. Johnson, seconded by Dr. Goldstein, the following resolution passed on a vote of 7-0:

Figure C.1: Sample School District Meeting Text From Ladue School District

Model Output:

```
{  
  "Name": ["Andy Cohen"],  
  "Address": ["NA"],  
  "Comment": ["Comment regarding agenda and non-agenda items"]  
}
```

Figure C.2: Formatted output from the LLM

Meeting Text Excerpt #2

```
3. ANNOUNCEMENTS  
3.01 Student Representative Announcements  
Ms. Jaiden Smith was not present.  
3.02 Superintendent's Announcements  
Dr. Hardin-Bartley read superintendent announcements.  
4. CITIZEN COMMENTS  
4.01 Citizen Comments  
e Stacy Brown- surplus property  
e Maureen Axelbaum- surplus property  
e Nicole Adewale- surplus property  
e Jacqueline Stack- surplus property  
e Colleen Dille- surplus property  
e Roma Trousdale- surplus property  
e Aaren Ginsburg- surplus property  
e Lori Yudovich- surplus property  
e Yvette Joy Liebesman, 7570 Cornell Avenue- surplus property  
e Steven McMahon- surplus property  
e Mark Tubbs- surplus property  
e Meg A Ullman- surplus property  
e Kim Field- surplus property  
5. CONSENT AGENDA  
5.01 Approval of the Consent Agenda
```

Figure C.3: Sample of Meeting Text From University City School District

Model Output:

```
[
{ "Name": ["Stacy Brown"],
  "Address": ["NA"],
  "Comment": ["Surplus property"]},
{ "Name": ["Maureen Axelbaum"],
  "Address": ["NA"],
  "Comment": ["Surplus property"]},
{ "Name": ["Nicole Adewale"],
  "Address": ["NA"],
  "Comment": ["Surplus property"]},
{ "Name": ["Jacqueline Stack"],
  "Address": ["NA"],
  "Comment": ["Surplus property"]},
{ "Name": ["Colleen Dille"],
  "Address": ["NA"],
  "Comment": ["Surplus property"]},
{ "Name": ["Roma Trousdale"],
  "Address": ["NA"],
  "Comment": ["Surplus property"]},
{ "Name": ["Aaren Ginsburg"],
  "Address": ["NA"],
  "Comment": ["Surplus property"]},
{ "Name": ["Lori Yudovich"],
  "Address": ["NA"],
  "Comment": ["Surplus property"]},
{ "Name": ["Yvette Joy Liebesman"],
  "Address": ["7570 Cornell Avenue"],
  "Comment": ["Surplus property"]},
{ "Name": ["Steven McMahon"],
  "Address": ["NA"],
  "Comment": ["Surplus property"]},
{ "Name": ["Mark Tubbs"],
  "Address": ["NA"],
  "Comment": ["Surplus property"]},
{ "Name": ["Meg A Ullman"],
  "Address": ["NA"],
  "Comment": ["Surplus property"]},
{ "Name": ["Kim Field"],
  "Address": ["NA"],
  "Comment": ["Surplus property"]}
]
```

Figure C.4: Formatted output from the LLM

C.1.2 Amicus Curiae Briefs:

The following section displays an example amicus curiae brief excerpt and its resulting output from OpenAI's GPT 3.5 model. For this example, I prompted the model with the prompt displayed within the body of the main paper and then fed the first and last 10 pages of the brief from which the excerpt originates into the model.

Brief Text Excerpt #1

No. 14-15
In the supreme Court of the United states

RICHARD ARMSTRONG, et al.,
Petitioners,
V.
EXCEPTIONAL CHILD CENTER, INC., et al.,
Respondents.

On Writ of Certiorari to the United States
Court of Appeals for the Ninth Circuit
Brief of Amici Curiae National Health Law Program,
AARP, Service Employees International Union,
National Respondent Legal Aid and Defender Association,
Sargent Shriver National Center on Poverty Law,
National Disability Rights Network,
DisAbility Rights Idaho, Center for Medicare Advocacy,
National Senior Citizens Law Center, First Focus,
National Center for Youth Law, National Center for
Law and Economic Justice, National Housing Law Project,
National Women's Law Center,
Planned Parenthood Federation of America,
National Family Planning & Reproductive Health Association,
National Latina Institute for Reproductive Health,
Asian Americans Advancing Justice Los Angeles,
Southern Poverty Law Center in Support of Respondents

Kelly Bagby Jane Perkins
AARP Foundation Counsel of Record
Litigation National Health
Counsel for Amici Curiae
Becker Gallagher Cincinnati,
OH Washington, D.C. 800.890.5001

Figure C.5: Sample of amicus curiae brief text

Model Output:

```
{
"National Health Law Program":
Respondent,
"AARP":
Respondent,
"Service Employees International Union":
Respondent,
"National Legal Aid and Defender Association":
Respondent,
"Sargent Shriver National Center on Poverty Law":
Respondent,
"National Disability Rights Network":
Respondent,
"DisAbility Rights Idaho":
Respondent,
"Center for Medicare Advocacy":
Respondent,
"National Senior Citizens Law Center":
Respondent,
"First Focus" :
Respondent,
"National Center for Youth Law":
Respondent,
"National Center for Law and Economic Justice":
Respondent,
"National Housing Law Project":
Respondent,
"National Women's Law Center":
Respondent,
"Planned Parenthood Federation of America":
Respondent,
"National Family Planning&Reproductive Health Association":
Respondent,
"National Latina Institute for Reproductive Health":
Respondent,
"Asian Americans Advancing Justice Los Angeles":
Respondent,
"Southern Poverty Law Center":
Respondent
}
```

Figure C.6: Formatted output from the LLM

C.2 Alternative BERT-based Approach

To highlight how Large Language Models (LLMs) compare to other modern machine learning solutions. I process all local meeting minutes within my sample using BERT-based Named Entity Recognition (NER). These models are trained to identify concepts such as people, locations, organizations, dates, and other entities from text and provide the next best solution to the problem of extracting public commenters from meeting minutes. For this application, I utilize an open-source implementation from the spaCy natural language processing toolkit. Importantly, I do not retrain the model to identify commenters specifically; instead, I rely on its ability to identify individuals’ names with texts. To compensate for this lack of additional training, I preprocess each set of meeting minutes only to include the rough text surrounding their public comment sections, which I identify using typical section titles or specific references or calls for comments. I then utilize the model to identify all individuals within these sections and assume the proceeding text represents their address and comment. Given that the model can not distinguish between public officials and regular commenters, I remove all commenters whose names include formal governmental titles such as “counselor,” “chair,” or “alderperson.” Table C.1 below presents the model’s performance relative to that of the LLM from the main paper across both the training and unlabeled data sets. Overall, the NER model performed markedly worse than the LLM and required significantly more implementation and coding to transform the texts successfully. While a trained model could likely close this performance gap, the accessibility and adaptability of LLMs make them a far more attractive tool for transforming text-based sources.

Table C.1: LLMs Compared to Alternative Name Entity Recognition Approach

Model	Precision	Recall	F1
<i>NER</i>	0.78	0.85	0.81
<i>LLM_{Training}</i>	0.934	0.984	0.957
<i>LLM_{Unlabeled}</i>	0.935	0.977	0.950