The Costs of Housing Regulation: Evidence From Generative Regulatory Measurement

Alexander W. Bartik, Arpit Gupta, and Daniel Milo*

October 27, 2023

Abstract

We introduce a new approach to decode and interpret statutes and administrative documents employing Large Language Models (LLMs) for data collection and analysis that we call *generative regulatory measurement*. We use this tool to construct a detailed assessment of U.S. zoning regulations. We estimate the correlation of these housing regulations with housing costs and construction. Our work highlights the efficacy and reliability of LLMs in measuring and interpreting complex regulatory datasets.

JEL-Classification: R52, R58, K11, O38, R31, C81
Keywords: housing regulation, zoning codes, large language models, natural language processing, artificial intelligence, municipal ordinances

*Bartik: Department of Economics, University of Illinois at Urbana-Champaign, abartik@illinois.edu. Gupta: New York University, Stern School of Business, ag5808@stern.nyu.edu. Milo: New York University, Stern School of Business, dm4766@stern.nyu.edu This paper has benefited from conversations with Theresa Kuchler and comments from seminar participants at the NYU Stern Corporate Governance luncheon. Grace Getman, Karin Hobelsberger, Nagharjun Mariappan, and Alok Ranjan provided excellent research assistance. Any errors or omissions are the responsibility of the authors.

1 Introduction

From the early 2000s, housing production in the United States has consistently declined. Shortages in housing construction are associated with increased house prices, a rising housing capital share of income, and sluggish productivity growth in the real estate construction sector (Rognlie, 2016; Goolsbee and Syverson, 2023; Glaeser and Gyourko, 2018), and many studies link these changes to stringent housing regulations (Glaeser and Ward, 2009; Gyourko et al., 2008). However, current evidence remains limited in measuring housing regulations and identifying their role in affecting housing supply. A critical challenge to answering these questions lies in simply understanding what current housing regulations are, given the fact that every municipality has adopted a distinct set of zoning and building code regulations, and no comprehensive dataset exists to simply evaluate the diversity in housing regulations across the country.

Our paper argues that advances in Large language Models (LLMs) have advanced to enable the systematic analysis of local regulations, a task which we refer to as *generative regulatory measurement*. We use state-of-the-art Artificial Intelligence (AI) methodologies to estimate zoning regulations across a large fraction of municipalities in the United States. Through the use of LLMs (such as Llama 2 and ChatGPT) applied to the full-text of local regulations, we aim to offer a solution to the knowledge gap in understanding the nature of zoning codes across the country, as well as in assessing their impact on broader economic outcomes. Our focus is on zoning codes specifically, due to their importance in shaping housing markets. However, challenges in appropriately interpreting and analyzing textual databases are common across multiple domains (in building codes, regulations, court cases, earnings call transcripts, newspapers, etc.) and so our approach also has broader applicability in suggesting possible approaches towards the classification of legal and regulatory texts more broadly. Developing such approaches has become increasingly important as the quantity and complexity of regulation has risen over time (Singla, 2023).

To do so, our study answers three main questions. First, we assess how well can LLMs perform in accurately creating datasets which summarize complicated laws and regulations. Assessing the quality of this task is crucial to establishing whether these AI approaches can aid in systematically organizing texts for research and policy. Existing research on AI models emphasizes both their promise in analyzing textual data (Zhao et al., 2023), as well as challenges with undesirable AI features such as "hallucination" and manufactured model output (Azamfirei et al., 2023). Verifying whether LLMs can accurately parse large legal documents—and for which questions— is therefore a crucial step towards our understanding of the capacities of these models, with the promise of either opening up the large-scale use of textual documents for quantitative research, or cautioning about important pitfalls in the interpretation of such research.

We conduct a direct comparison of answers to specific questions on zoning codes conducted on human-coded fields taken from the Pioneer Institute (see Glaeser and Ward (2009)) in the state of Massachusetts, which provides an effective training dataset for our analysis. Our results point to substantial heterogeneity in the ability of LLMs to accurately answer questions about housing regulations, depending on the specific question and model. Using the best models (i.e., ChatGPT 4 by Open AI), we find an accuracy rate around 80%, varying between 70%–94% depending on the precise question. These results suggest that these models are currently able to reach high-levels of ability to accurately assess the information contained in zoning codes on at least some questions. Importantly, these accuracy statistics vary substantially based on the nature of the question and the vintage of the model, indicating that model and question specifics matter substantially in assessing the ability of AI models to perform interpretative tasks on regulatory documents. Accuracy is lower in estimating questions with categorical or continuous answers, rather than binary responses. These accuracy statistics are also likely to change with newer versions of AI models scheduled to appear, suggesting that regulatory interpretative questions are likely to become more feasible in the future.

Second, we assess the variation in zoning regulation across the United States. As a result of the knowledge gap in the literature resulting from the complexity of accurately classifying and understanding zoning and building code data, we lack a clear understanding of how precisely these housing regulations vary across the United States. Importantly, this analysis is conducted at the municipal level, the relevant unit of local government responsible both for the construction of zoning codes, as well as in providing public goods. While some existing research has explored proxies for housing regulations at more aggregate levels (Gyourko et al., 2008), we provide both more detailed as well as granular data across a large sample of municipalities in the United States. Our data covers between 30–40% of all residents in the country, and we are working to scale this up to cover a much larger fraction of the country.

Third, we also use our dataset to make progress on the question of whether housing regulations are associated with housing costs and construction. We find that affordable housing requirements and minimum lot sizes are strongly related with measures of housing prices and rents, while measures of frontage requirements, the legality of cluster developments, unit caps, and building conversion rules are associated with permits per capita. Obviously, these correlations could reflect selective adoption or a causal impact of these zoning rules. In future work we will attempt to disentangle these explanations.

Our results serve as an initial proof of concept towards the use of LLMs in the systematic generation of content in regulatory and legal documents, with widespread applications across domains. We focus on assessing the accuracy of these models in a limited domain with existing, high-quality trained data, as well as exploring scalability of these tools across much larger samples. Our estimates point to substantial heterogeneity in the accuracy of LLMs in accuracy assessing regulatory content across questions and specific models, with the best models achieving high accuracy on some questions. These results suggest that LLMs, currently, cannot substitute for human interpretation in contexts in which extremely high accuracy is required: for instance, considering establishing the specific regulatory environment in a jurisdiction for a developer or city planner. However, they suggest that LLMs are reaching the level of accuracy for which they may be useful to researchers who wish to understanding underlying variation in regulation and its impacts on other variables, in statistical contexts which tolerate some amount of measurement error.

Importantly, our results should be seen as illustrating a base level of performance using widely accessible tools, and have considerable scope for improvement along several dimensions. We use standard open source models for both our embedding algorithm as well as to scale our results across the United States; but will likely achieve higher accuracy using the highest-end proprietary models. Our results use "zero shot learning," rather than fine-tuning, but training LLMs on existing regulatory documents will likely improve performance further. Further pre-processing of documents to focus LLMs on relevant text is also likely able to improve model accuracy. Finally, we use the highest quality LLM available at the time of writing (ChatGPT 4), but these models are likely to improve over time. We plan to expand the scope of this work to examine changes zoning codes over time, in analyzing housing regulations across countries (including in other languages), as well as in analyzing building codes in conjunction with zoning codes. Combined, the promise of these efforts suggest that LLMs are likely to fundamentally reshape our ability to understand the content and impact of housing regulation. We will make available our entire data workflow: including underlying text, code used for LLM classification, as well as the final regulatory dataset.

Contributions to Literature The central contribution of our project is the creation of a standardized, comprehensive dataset of zoning across the United States. Much of the existing literature on housing regulations has used either indirect proxies for zoning regulation, or carefully analyzed specific regulations only in certain regions. The first strand of this literature has focused on survey-based approaches to measuring housing regulations. One of the most heavily used such nationwide measures of housing regulation includes the Wharton Regulatory Index (Gyourko et al., 2008, 2021; Huang and Tang, 2012). This pioneering approach to measuring housing regulations was based on surveys sent to 2,649 distinct municipalities (there are 19,522 municipalities in the United States in total), asking for information on the regulatory process, details of local land use regulations, and outcomes of the permitting and regulatory process. The survey itself builds on earlier work which surveyed a smaller number of municipalities (Mayer and Somerville, 2000), and other research has focused on surveys given to local officials and planners (Saks, 2008). While this survey approach provides invaluable information on the perceived nature of regulatory burdens, we hope to build on this literature to provide an even more comprehensive and direct measure of housing regulations. We will do so both by expanding the range of municipalities covered to span the entire United States, as well as by taking our information directly from the textual source in actual zoning documents.

The second strand of this literature includes wedge-based approaches, which instead aim to impute housing regulations by examining the expected spatial macroeconomic distortions resulting from zoning. Examples in this literature Hsieh and Moretti (2019), Glaeser et al. (2005), Herkenhoff et al. (2018), and Duranton and Puga (2019). Babalievsky et al. (2021) apply a similar production function based approach to impute the impact of commercial zoning impacts. While this literature provides a measure of the spatial distortions in economic activity, it is unclear how to link these distortions to specific housing regulations, which poses a challenge for policymakers hoping to address housing costs and housing undersupply through targeted policy actions and to researchers who seek to better understand the microeconomics of housing production

Third, other national approaches have examined textual data, but in more limited ways. Ganong and Shoag (2017) focus a scaled count of judicial decisions on "land use." While this is surely a proxy for regulatory strictness, it leaves open the question of precisely which housing regulations are driving housing litigation. In a similar spirit, Stacy et al. (2023) use machine learning tools to identify newspaper articles discussing changes to zoning restrictions in eight metropolitan areas and classify them as either loosening and tightening zoning restrictions and then analyze the effects of these changes in regulation on housing supply and rents. Our approach, by contrast, is able to establish more cleanly the precise nature of housing regulations across a broad sample of jurisdictions in the United States.

Another strand of this literature has attempted to address the limitations in national-level approaches through more detailed analysis of specific regulations at the state level. Most prominent is the approach by the Pioneer Institute, which has engaged in explicit classification of zoning rules for 186 municipalities in the state of Massachusetts. Prior work by Glaeser and Ward (2009) establishes that regulatory intensity measured in this dataset does indeed associate with higher costs and lower construction. Gyourko et al. (2008) mention both the importance of this kind of detailed local analysis, as well as the challenges in scaling this approach to the national level:

"The proliferation of barriers and hurdles to development has made the local regulatory environment so complex that it is now virtually impossible to describe or map in its entirety. Glaeser et al. (2006) come closest to doing so. For a subset of the Boston metropolitan area, they conducted a detailed analysis of local zoning codes, permitting precise calculations of potential housing supply across communities. However, the enormity of that effort prevents it from being replicated in other markets by a single research team."

We argue that the practical difficulties behind the scaling up of this approach may now be addressed through the development of modern AI LLMs, providing both the granularity of the state-based approaches along with the scale of the national regulatory studies. Indeed, the Pioneer Institute data—the most comprehensive of these state based approaches—will be a crucial test for our approach. We begin our analysis by first analyzing data in Massachusetts using the same data source for municipal documents identified by the Pioneer Institute team, which allows for a cross-validation of the accuracy of our AI-led approach against the existing housing regulation classification. This serves as an important validation check of our approach. Other detailed state-level analyses of housing regulation include Shanks (2021) which also focuses on Massachusetts and uses Machine Learning tools (Latent Dirichlet Allocation). California has also been the subject of detailed and specific analysis, focusing in particular on growth limitations (Quigley and Raphael, 2005; Jackson, 2016), as has Florida (Ihlanfeldt, 2007).

These studies leave important gaps in our understanding of housing regulations under both the national and state-level analyses. While the national approaches establish that housing regulations appear to drive important variation across the country in housing costs and construction activity, they have less to say about which specific regulations are the key drivers. Isolating specific regulatory impacts is essential for policy seeking to remedy possible impacts of regulatory driven housing cost increases. Alternatively, more detailed state-level data offers the potential to isolate the specific aspects of ho using regulation that are most binding. These approaches, however, are limited in their geographic scope outside the unique states of Massachusetts, California, and Florida. Consequently, the extent to which specific housing regulations drive costs and construction activity across the country are unclear. Both line of research are also not able to contrast costs with potential benefits or amenities, making it impossible to disentangle supply and demand side effects which are crucial to establishing the cost-benefit tradeoffs of housing regulation.

Relative to this literature, our contribution is to construct a more comprehensive and detailed measure of how zoning regulations and building codes vary across the United States. We provide the most detailed assessment to date of all relevant housing regulations (i.e., minimum lot sizes, whether multifamily apartments can be constructed, inclusionary zoning mandates, setback rules, etc.) that apply to construction in local areas.

Additionally, we also contribute to the literature by testing the accuracy and usefulness of LLMs in creating novel regulatory and policy datasets. A broader contribution of our project will be a large-scale application of large language models to a complex regulatory and policy dataset generation task. This initiative will serve as a critical test case for the efficacy and reliability of LLMs in not only understanding and processing complex legal and regulatory language but also in discovering and extracting novel, actionable insights from a vast array of documents. While prior literature has used textual data to extract information, particularly sentiment, from text (Hassan et al., 2019; Romer and Romer, 2004; Tetlock, 2007; Lopez-Lira and Tang, 2023); a few papers have begun to use LLMs to examine existing textual or regulatory documents (Jha et al., 2023; Yang, 2023; Hansen and Kazinnik, 2023). Hoffman and Arbel (2023) argues for the use of LLMs in "generative interpretation" in estimating the meaning of legal contracts.

To ensure the accurate translation of legal and policy documents into comprehensive datasets, it is crucial to validate the LLM's ability to comprehend the intricacies and nuances of legal language, which often varies by jurisdiction and context. By leveraging LLMs' contextual understanding and pattern recognition capabilities, we aim to create datasets that accurately represent the regulatory landscape across multiple domains and timeframes.

In addition to the creation of these datasets, a significant component of our project will involve evaluating the usefulness of the information that LLMs can extract and structure. We plan to assess the practical applicability and relevance of the datasets for different users—policymakers, regulators, researchers, and industry practitioners.

This project's broader contribution is to provide a robust foundation for the continued use and development of LLMs in the domain of legal and regulatory research. By showcasing how these sophisticated models can streamline data creation, enhance accessibility of information, and support predictive analysis, we can make a compelling case for the value of integrating LLMs into the broader regulatory and policy landscape.

2 Data and Background

2.1 Municipal Codes and Zoning

In the United States, local governments are "creatures of the state" subordinate to state control. Municipal corporations are authorized, subject to state law, to organize local government, and refer to cities, towns, villages, and other government units which function in that capacity. This concept largely overlaps with the Census definition of "incorporated place" which we use to organize our analysis.[1]

Municipalities have rights over local zoning decisions; indeed, the desire to control local zoning is a common reason to incorporate in the first place. Zoning, broadly, consists of two key sets of regulations: land use regulations, which partition local land into distinct use classes, and bulk regulations, which restrict the density of buildings in different land use classes. Examples of bulk regulations include: coverage, setbacks, height restrictions, and floor area ratio caps. Other mandates and requirements, such as parking minimums, further constrain both commercial and residential development in different areas.

Municipalities enforce laws by issuing municipal codes which outline local regulation in different domains. Zoning codes outline permitted uses for different classes of land, and outline housing regulations. Some regulations apply broadly to all land within a jurisdiction; other regulations (such as minimum lot sizes) typically vary depending on the specific use class (i.e., single-family zoning, commonly referred to as R1, or commercial or industrial). These ordinances are typically updated over time to reflect changes in local regulations, and are aggregated by different companies. Table 1 illustrates a range of underlying sources for these sets of municipal codes, which are a matter of public record. American Legal Publishing provides significant numbers of records outside of the south, Municode has coverage in all regions outside the Northeast, while we only

[1]In several states the "Township" form of government also has jurisdiction in zoning which aligns with the Census County Subdivision definition.

use Ordinance.com data for Massachusetts[2]. Combining these sources, our data covers municipalities covering 38 percent of the US population overall and 39 percent of the US population that lives in metropolitan areas.

The primary dataset for our analysis consists of the full-text of zoning documents. At the municipality-level, we also draw on information on building permits data from the Census Building Permits Survey. We also connect to rent and price data drawn from the American Community Survey (ACS) at the municipality level.

2.2 Large Language Models

Large Language Models (LLMs) are a form of artificial intelligence that primarily handle sequential data such as sequences of words in textual data. LLMs are based on the deep learning "transformer" architecture as introduced in Vaswani et al. (2017). The key innovation is the “attention mechanism," enabling the model to focus on multiple words of the input text at once. This helps the model understand words in context, such as sentences or paragraphs. Transformers also represents a significant advancement in terms of both accuracy and runtime over previous models like Recursive Neural Networks, which processed sequences linearly. LLMs are trained with semi-supervised learning, first pretraining the model on a large corpus of text and subsequently fine-tuning the model with human feedback. After training, LLMs can generate human-like text, answer questions, summarize text, and generalize from their training to perform tasks they were never explicitly trained for, a concept known as zero-shot learning. This means the model does not need as an input explicit examples of additional training to perform well in an out-of-sample exercise, a key advantage we use in our analysis.

LLMs have several advantages and disadvantages relevant for our setting in applying

[2] We currently only have access to Massachusetts data for Ordinance.com but the provider also has data for California, Connecticut, New Jersey, Rhode Island, and Washington State as well as the metro areas for Chicago, New York City, Philadelphia, and Washington D.C.

to housing regulatory textual analysis. The central advantage is scalability: we are able to load large quantities of municipal code data for classification and analysis, which far exceeds the capacity of any human team to analyze. Other advantages include the prospect for additional training, allowing for increased accuracy over time as LLMs improve in accuracy and additional training data is incorporated into the analysis.

Potential drawbacks in using LLMs for this purpose center on the inaccuracy of measurement and classification. This can happen either through limitations in the context window used to identify relevant text from the sample corpus, or the content and lack thereof of similar questions and related texts in the underlying training sample. Current state-of-the-art LLMs may inadvertently produce incorrect information, produce information with an incorrect degree of certitude, and potentially manufacture data output ("hallucination"). Possible biases in the responses are linked to the quality of training data, and so measurement error may or may not be classical depending on the explanatory variable of interest. Finally, relevant information to answer zoning regulation questions may be outside the domain (i.e., in the form of state regulation not contained within our ordinance sample). We attempt to measure these drawbacks through comparison of LLM-generated output against human defined categorizations of regulation.

2.3 Processing Municipal Codes Using LLMs

To conduct our analysis, we use a standard framework known as "retrieval-augmented generation" (Lewis et al., 2020). The basic objective of this approach is to combine a large pre-trained language model with external information retrieval, in order to give the LLM the ability to "look up" information from a vast corpus of documents during the generation process. We outline our general procedure in Figure 1.

The first step of our process is to download and scrape the sources of municipal codes listed in Table 1, which provides us with a large corpus of zoning documents relevant for our analysis. These municipal codes contain detailed housing and zoning regulations relevant for our study. We then take advantage of text embedding approaches, which are representations of this textual data encoded in vectors. This enables the efficient processing of this large textual data. The basic intuition behind embedding is to represent words with vectors which represent a dimension in embedding space, such that words with similar semantic meaning are closer in this space. For our zoning document, this ensures that we are able to retrieve components of the document relevant for our specific questions. Different embedding algorithms conduct this task in distinct ways; we use the `multi-qa-mpnet-base-dot-v1` algorithm, which is fine-tuned for the purposes of question and answering.[3] We similarly embed the questions we want answered from the documents, which for ease of comparison we limit to the question base already developed by the Pioneer Institute (i.e., "Is multifamily zoning allowed in this area as-of-right?").

With two separate embedded vectors in hand, the zoning documents from a particular municipality and a question we would like answered, we then isolate the component of the document most relevant to answer the question. The length of typical zoning documents far exceeds the context windows currently usable by LLMs, so we need to select specific blocks of text relevant to answer the question. We use cosine similarity, a standard measure of distance between two multi-dimensional objects, to find the textual component of zoning documents most relevant to answer each question, and pick the five chunks of text which are relevant in answering the question. We chose five chunks based on initial experimentation of how the choice of the number of chunks impacted the accuracy of results and computational costs.

With the relevant text in hand, we then feed the relevant information to a LLM in order to perform the task of generative regulatory measurement: having the AI model estimate the content of the regulation. Here, we experiment with several LLMs, including both open source and proprietary models, and show the relative accuracy of different approaches. In the future, many additional models are likely to become available, further

[3]See https://huggingface.co/sentence-transformers/multi-qa-mpnet-base-dot-v1.

enhancing the accuracy of the procedure.

3 Results

3.1 Validation using Pioneer Data

As an initial stage of validation, Table 3 shows the accuracy of three LLM approaches, cross-validated using the Pioneer Institute data. The rows highlight each zoning-related question drawn from the Pioneer Institute's zoning question, and we show the relative squared error for each question across Llama-13b, Llama-70b, Chat GPT 3.5, and Chat GPT 4. These are two sets of LLMs made available by Meta (Llama) and Open AI (Chat GPT), which differ primarily in the number of parameters used in the training process. For these cases, we compare the response recorded in the Pioneer Institute data against three LLM approaches, and classify the answer as correct if the LLM provides an identical text answer.

In Panel A, we examine questions for which the desired answer is a continuous variable. We measure model accuracy both using squared error relative to the sample mean (i.e., the mean squared error divided by the square of the difference between the sample mean and observed value). This measure has the advantage of being easy to compare across different data types (continuous, categorical, and binary) as well as being robust to differences in the base rate. Heuristically, this metric compares the performance of the LLMs to someone who just always guessed the sample mean (Panel A) or sample mode (Panel B). This is fairly conservative because even the sample average already contains a non-trivial amount of information. We also show correlations as an alternative, easier to interpret, measure. We find variation across LLMs, with the highest quality (Chat GPT 4) model giving correlations between 0.46–0.64 across question types.

In Panel B, we examine binary variables, i.e., those with a yes or now answer. Here, instead of correlation, we report raw accuracy as an easy to interpret measure of model fit. We find accuracy rates across questions varying from 55–94% across questions for the highest quality GPT 4 model, averaging at 79%. We also consider categorical questions in Panel C, finding the lowest accuracy in this section.

In Figure 3, we visualize the average results across questions in Table 3. In dark blue, we plot the percent correct for each model using the percent accuracy for binary variables, the correlation for continuous variables, and adjusted percent correct for categorical questions. The smaller model Llama 2-13B underperforms the other larger models. We also plot the frequency each model says "I don't know" in medium blue, which varies across each model and question type. Finally, we attribute the remainder as the incorrect percent for each model (shown in grey).

3.2 Nationwide Variation in Zoning Codes

Figures 6 and 7 show maps of minimum lot sizes and affordable housing mandates, respectively, for jurisdictions within the metropolitan areas surrounding four select cities in the United States, San Francisco, Chicago, Atlanta, and Boston. We chose these metro-areas to span all major regions and to capture a variety of policy and legal environments. Our nationwide results were produced based on the Llama 2-13b model; as previously discussed, these are not the most accurate models, but are free and open source options. As a result, we interpret the results with caution and seek to improve the accuracy rate over time. These graphs document substantial variation in both minimum lot sizes and affordable housing mandates and incentives within metropolitan areas across municipalities, with the central city and inner suburbs having lower minimum lot sizes and higher rates of affordable housing mandates than in jurisdictions farther from the central city. This figure illustrates a key advantage of our approach: the ability to construct measured of zoning ordinances at the level of the municipality across a wide variety of municipalities and regions in the United States.

Figure 2 shows the distribution of four different housing regulations across the US:

number of zoning districts, largest frontage requirement, mean minimum lot size (across all zoning districts), and minimum minimum lot sizes (across all zoning districts). The figure shows that these regulations vary substantially. For example, a large mass of municipalities has no minimum lot size requirement at all, while a non-trivial share of municipalities have minimum minimum lot sizes in excess of 10 thousand square feet.

Table 2 shows the association of housing regulations across income and urban categories across the United States. We observe, for instance, that affordable housing mandates are found much more often in higher income and urban areas. Lot sizes appear much higher in higher income areas, but lower in urban areas—consistent with their role in suburbs as as a form of "exclusionary zoning." Other categories of regulation appear surprisingly balanced across regional attributes. For instance most municipalities do not allow multi-family housing, by right or special permit, even in the most urban areas. Appendix Table A1 shows the distribution of responses, by model and question, for which the model response was "I don't know."

3.3 Housing Regulation and Broader Outcomes

In Table A5, we perform initial analysis of housing regulatory fields we measure across the country using our Llama 2-13b model, correlating these different measures of housing supply measured as levels and changes of rents, house values, and building permits. Due to the limitations in the accuracy of the Llama 2 accuracy and the obvious potential for selection bias, we view these results as highly preliminary, and include them only to illustrate the scope of analysis possible through this procedure, which we intend to further corroborate using higher-quality LLMs, as well as through quasi-experimental methods to establish causation.

Nonetheless, our analysis reveals some interesting patterns of associations. Areas with affordable housing mandates are associated with regions with substantially higher rents and prices, consistent with these regulations being clustered in more expensive housing markets, but are associated with less construction. Allowing certain housing types by right (cluster developments, planned unit developments, open space residential designs, or other types of flexible zoning) are associated with higher development, though also higher rents and prices. Median house values and gross rents are associated with higher lot sizes. Caps on residential permits are also associated with less construction. To be sure, our correlations consistent either with a causal impact on supply, or are the product of selection. Both possibilities are potentially interesting, highlighting either the impacts of housing regulation on other outcomes, or the differential adoption of housing regulation by different areas. Future work will work to better tease out these implications using more accurate models and empirical designs.

4 Conclusion

Our paper is the first to develop the use of LLMs to extract the content of regulation from administrative documents, which we refer to as *generative regulatory measurement*, and validate the accuracy of this procedure. Our project serves as a critical examination of the utility and accuracy of LLMs in processing and understanding complex legal and regulatory language. By creating comprehensive datasets that accurately reflect the regulatory landscape, we explore the potential of LLMs as invaluable tools for various stakeholders, including policymakers, researchers, and practitioners in the field.

In addition to generating actionable insights, the research is committed to making all collected data and the associated replication code publicly available. This open-access approach ensures that the wider research community can benefit from the project's findings, use the datasets for various analytical purposes, and even apply the LLM-based classification methodology to explore other regulatory domains.

Ultimately, this project stands as a significant step forward in the integration of advanced technological tools in legal and regulatory research. By demonstrating the effectiveness of LLMs in data creation, analysis, and predictive tasks, we argue for the broader application of LLMs in understanding and navigating the complex tapestry of regulations that shape the housing and construction sectors in the United States and beyond.

References

Azamfirei, R, SR Kudchadkar, and J Fackler, "Large language models and the perils of their hallucinations," *Critical Care*, 2023, *27* (1), 1–2. 1

Babalievsky, F, K Herkenhoff, LE Ohanian, and EC Prescott, "The Sky is Not the Limit: The Macroeconomic Consequences of Commercial Zoning Laws," Technical Report 2021. 1

Duranton, Gilles and Diego Puga, "Urban growth and its aggregate implications," Technical Report, National Bureau of Economic Research 2019. 1

Ganong, Peter and Daniel Shoag, "Why has regional income convergence in the US declined?," *Journal of Urban Economics*, 2017, *102*, 76–90. 1

Glaeser, Edward and Joseph Gyourko, "The economic implications of housing supply," *Journal of Economic Perspectives*, 2018, *32* (1), 3–30. 1

Glaeser, Edward L and Bryce A Ward, "The causes and consequences of land use regulation: Evidence from Greater Boston," *Journal of Urban Economics*, 2009, *65* (3), 265–278. 1, 1

_ , Joseph Gyourko, and Raven Saks, "Why is Manhattan so expensive? Regulation and the rise in housing prices," *The Journal of Law and Economics*, 2005, *48* (2), 331–369. 1

Goolsbee, Austan and Chad Syverson, "The Strange and Awful Path of Productivity in the US Construction Sector," Technical Report, National Bureau of Economic Research 2023. 1

Gyourko, Joseph, Albert Saiz, and Anita Summers, "A new measure of the local regulatory environment for housing markets: The Wharton Residential Land Use Regulatory Index," *Urban Studies*, 2008, *45* (3), 693–729. 1, 1

—, **Jonathan S Hartley, and Jacob Krimmel**, "The local residential land use regulatory environment across US housing markets: Evidence from a new Wharton index," *Journal of Urban Economics*, 2021, *124*, 103337. 1

Hansen, Anne Lundgaard and Sophia Kazinnik, "Can ChatGPT Decipher Fedspeak?," *Available at SSRN*, 2023. 1

Hassan, Tarek A, Stephan Hollander, Laurence Van Lent, and Ahmed Tahoun, "Firm-level political risk: Measurement and effects," *The Quarterly Journal of Economics*, 2019, *134* (4), 2135–2202. 1

Herkenhoff, Kyle F, Lee E Ohanian, and Edward C Prescott, "Tarnishing the golden and empire states: Land-use restrictions and the US economic slowdown," *Journal of Monetary Economics*, 2018, *93*, 89–109. 1

Hoffman, David A and Yonathan A Arbel, "Generative interpretation," *Available at SSRN*, 2023. 1

Hsieh, Chang-Tai and Enrico Moretti, "Housing constraints and spatial misallocation," *American Economic Journal: Macroeconomics*, 2019, *11* (2), 1–39. 1

Huang, Haifang and Yao Tang, "Residential land use regulation and the US housing price cycle between 2000 and 2009," *Journal of Urban Economics*, 2012, *71* (1), 93–99. 1

Ihlanfeldt, Keith R, "The effect of land use regulation on housing and land prices," *Journal of Urban Economics*, 2007, *61* (3), 420–435. 1

Jackson, Kristoffer, "Do land use regulations stifle residential development? Evidence from California cities," *Journal of Urban Economics*, 2016, *91*, 45–56. 1

Jha, Manish, Jialin Qian, Michael Weber, and Baozhong Yang, "Chatgpt and corporate policies," *Chicago Booth Research Paper*, 2023, (23-15). 1

Lewis, Patrick, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen tau Yih, Tim Rocktäschel et al., "Retrieval-augmented generation for knowledge-intensive nlp tasks," *Advances in Neural Information Processing Systems*, 2020, *33*, 9459–9474. 2.3

Lopez-Lira, Alejandro and Yuehua Tang, "Can chatgpt forecast stock price movements? return predictability and large language models," *arXiv preprint arXiv:2304.07619*, 2023. 1

Mayer, Christopher J and C Tsuriel Somerville, "Land use regulation and new construction," *Regional Science and Urban Economics*, 2000, *30* (6), 639–662. 1

Quigley, John M and Steven Raphael, "Regulation and the high cost of housing in California," *American Economic Review*, 2005, *95* (2), 323–328. 1

Rognlie, Matthew, "Deciphering the fall and rise in the net capital share: accumulation or scarcity?," *Brookings Papers on Economic Activity*, 2016, *2015* (1), 1–69. 1

Romer, Christina D and David H Romer, "A new measure of monetary shocks: Derivation and implications," *American economic review*, 2004, *94* (4), 1055–1084. 1

Saks, Raven E, "Job creation and housing construction: Constraints on metropolitan area employment growth," *Journal of Urban Economics*, 2008, *64* (1), 178–195. 1

Shanks, Brendan, "Land Use Regulations and Housing Development," 2021. manuscript. 1

Singla, Shikhar, "Regulatory costs and market power," Technical Report 2023. 1

Stacy, C, C Davis, YS Freemark, L Lo, G MacDonald, V Zheng, and R Pendall, "Land-use reforms and housing costs: Does allowing for increased density lead to greater affordability?," *Urban Studies*, 2023. 00420980231159500. 1

Tetlock, Paul C, "Giving content to investor sentiment: The role of media in the stock market," *The Journal of finance*, 2007, *62* (3), 1139–1168. 1

Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, 2017, *30.* 2.2

Yang, Stephen, "Predictive Patentomics: Forecasting Innovation Success and Valuation with ChatGPT," *Available at SSRN 4482536*, 2023. 1

Zhao, Wayne Xin, Kun Zhou, Junyi Li, Tianyi Tang, Xiaolei Wang, Yupeng Hou, Yingqian Min, Beichen Zhang, Junjie Zhang, Zican Dong et al., "A survey of large language models," *arXiv preprint arXiv:2303.18223*, 2023. 1

Tables

Table 1: Percent of Population Coverage by Source

Sample	National	Midwest	Northeast	South	West	MSA
American Legal Publishing	12.5	20.7	18.3	4.0	15.0	13.4
Municode	21.8	21.2	4.4	29.2	22.8	23.6
Ordinance.com	2.1	0.0	12.2	0.0	0.0	2.4
Total	36.4	41.9	34.9	33.1	37.8	39.4

Note: We use population estimates from the 2022 Census of Governments for municipality population, 2022 State-Level Census Population Data for census region and national population, and 2022 MSA-Level Census Population for MSA population.
Links to data sources are American Legal Publishing, Municode, and Ordinance.com.

Table 2: National Sample Question Means

Panel A: Continuous Questions

	National			Income Tercile			Urban/Rural		
Question	Mean	Weight	Count	Low	Mid	High	Rural	Mix	Urban
How many zoning districts, including overlays, are in the municipality?	9	11	4070	9	10	9	7	10	9
What is the longest frontage requirement for single family residential development in any district?	345	177	3889	199	222	612	491	285	631
Mean of Mean Lot Sizes (Square Feet)	87060	61753	3191	60992	66699	132937	84776	96088	41350
Mean of Min Lot Sizes (Square Feet)	7665	3684	3191	7708	7263	8014	10029	7028	6804

Note: We define the count (sample size) as the number of municipalities where the primary model (Llama-2 13B) does not say "I don't know" as the answer. The 'Weight' column weights each municipality by its population in the 2022 census of governments. We designate Urban/Rural using the percent overlap of the 2022 shape file for the municipality with the 2020 shape file for urban areas. Specifically, we define Urban as a municipality being 100% in an urban area, Mix as a municipality being partially in an urban area, and Rural as a municipality being 0% in an urban area. From the 2021 Five-Year American Community Survey we use median household income (B19013_001E). We have valid income data for 99.7% of our 4090 municipalities.

Panel B: Binary Questions

Question	National			Income Tercile			Urban/Rural		
	Mean	Weight	Count	Low	Mid	High	Rural	Mix	Urban
Is multi-family housing allowed, either by right or special permit (including through overlays or cluster zoning)?	36%	45%	2208	34%	37%	37%	32%	36%	44%
Are apartments above commercial (mixed use) allowed in any district?	18%	28%	3701	18%	18%	19%	11%	20%	25%
Is multi-family housing listed as allowed through conversion (of either single family houses or non residential buildings)?	5%	16%	2869	3%	5%	6%	3%	5%	8%
Are attached single family houses (townhouses, 3+ units) listed as an allowed use (by right or special permit)?	11%	10%	2778	9%	12%	13%	8%	13%	13%
Does zoning include any provisions for housing that is restricted by age?	6%	7%	3379	5%	7%	7%	5%	7%	9%
Are accessory or in-law apartments allowed (by right or special permit) in any district?	14%	20%	2335	13%	17%	13%	12%	15%	10%
Is cluster development, planned unit development, open space residential design, or another type of flexible zoning allowed by right?	46%	47%	2341	43%	47%	50%	39%	49%	42%
Is cluster development, planned unit development, open space residential design, or another type of flexible zoning allowed by special permit?	89%	76%	2134	88%	88%	90%	86%	90%	91%
Does the zoning bylaw/ordinance include any mandates or incentives for development of affordable units?	17%	43%	3054	9%	15%	26%	5%	21%	18%
Is there a town-wide annual or biannual cap on residential permits issued, and/or is project phasing required?	3%	3%	3006	2%	3%	5%	3%	4%	5%
Are there restrictions on counting wetlands, sloped land or easements in lot size calculations?	26%	35%	3728	26%	25%	28%	23%	27%	25%

Table 3: Model Performance Metrics on Pioneer Institute Massachusetts Study

Panel A: Continuous Questions - Relative Squared Error (RSE) and Correlation (Corr.)

Question	Llama-13B		Llama-70B		Chat GPT 3.5		Chat GPT 4	
	RSE	**Corr.**	**RSE**	**Corr.**	**RSE**	**Corr.**	**RSE**	**Corr.**
How many zoning districts, including overlays, are in the municipality?	2.2	0.36	1.6	0.48	1.1	0.60	0.9	0.64
What is the longest frontage requirement for single family residential development in any district?	4.5	0.24	2.7	0.26	2.2	0.41	1.8	0.40
Average Minimum Lot Size	0.8	0.28	0.5	0.48	1.2	0.52	1.6	0.47
Minimum Minimum Lot Size	0.6	0.28	0.5	0.39	1.2	0.47	0.7	0.63
Cumulative Average	2.0	0.29	1.3	0.40	1.4	0.50	1.2	0.53
Cumulative Median	2.0	0.28	1.3	0.40	1.2	0.50	1.2	0.53

Note: We calculate performance metrics and sample means (for RSE) only on the set of question municipality pairs that a model has valid answers for (does not say "I don't know"). Note that this set varies based on the model (see Appendix Table A1). For Relative Squared Error we compare each model's results to the naive model that guesses the sample mean. The correlation column is simply the correlation between the model answer and the Pioneer Institute answer. We winsorize data from our models at the 5% level but do not winsorize data from the Pioneer Institute. The Cumulative Average and Cumulative Median are calculated across questions giving equal weight to each question.

Panel B: Binary Questions - Relative Squared Error (RSE) and Percent Accuracy (Acc.)

Question	Llama-13B		Llama-70B		Chat GPT 3.5		Chat GPT 4	
	RSE	Acc.	RSE	Acc.	RSE	Acc.	RSE	Acc.
Is multi-family housing allowed, either by right or special permit (including through overlays or cluster zoning)?	8.5	51%	1.3	91%	2.0	88%	3.5	85%
Are apartments above commercial (mixed use) allowed in any district?	1.0	56%	0.7	67%	0.9	56%	0.5	76%
Is multi-family housing listed as allowed through conversion (of either single family houses or non residential buildings)?	1.1	57%	1.0	59%	0.8	68%	0.6	76%
Are attached single family houses (townhouses, 3+ units) listed as an allowed use (by right or special permit)?	1.9	40%	1.6	50%	1.6	51%	1.5	55%
Does zoning include any provisions for housing that is restricted by age?	0.8	60%	0.8	59%	0.7	66%	0.6	71%
Are accessory or in-law apartments allowed (by right or special permit) in any district?	1.2	47%	0.5	76%	0.7	81%	0.7	85%
Is cluster development, planned unit development, open space residential design, or another type of flexible zoning allowed by right?	27.0	38%	16.5	37%	3.7	87%	1.2	94%
Is cluster development, planned unit development, open space residential design, or another type of flexible zoning allowed by special permit?	1.3	79%	1.0	76%	0.8	83%	0.7	85%
Does the zoning bylaw/ordinance include any mandates or incentives for development of affordable units?	0.7	69%	0.6	73%	0.5	74%	0.4	79%
Is there a town-wide annual or biannual cap on residential permits issued, and/or is project phasing required?	0.8	75%	0.5	87%	0.3	91%	0.2	93%
Are there restrictions on counting wetlands, sloped land or easements in lot size calculations?	1.3	45%	1.0	57%	1.0	61%	0.8	70%
Cumulative Average	4.1	56%	2.3	67%	1.2	73%	1.0	79%
Cumulative Median	1.2	56%	1.0	67%	0.8	74%	0.7	79%

Note: For Relative Squared Error we compare each model's results to the naive model that guesses the sample mode. The accuracy column is calculated as the percent of municipalities that the model matches the Pioneer Institute answer for each question.

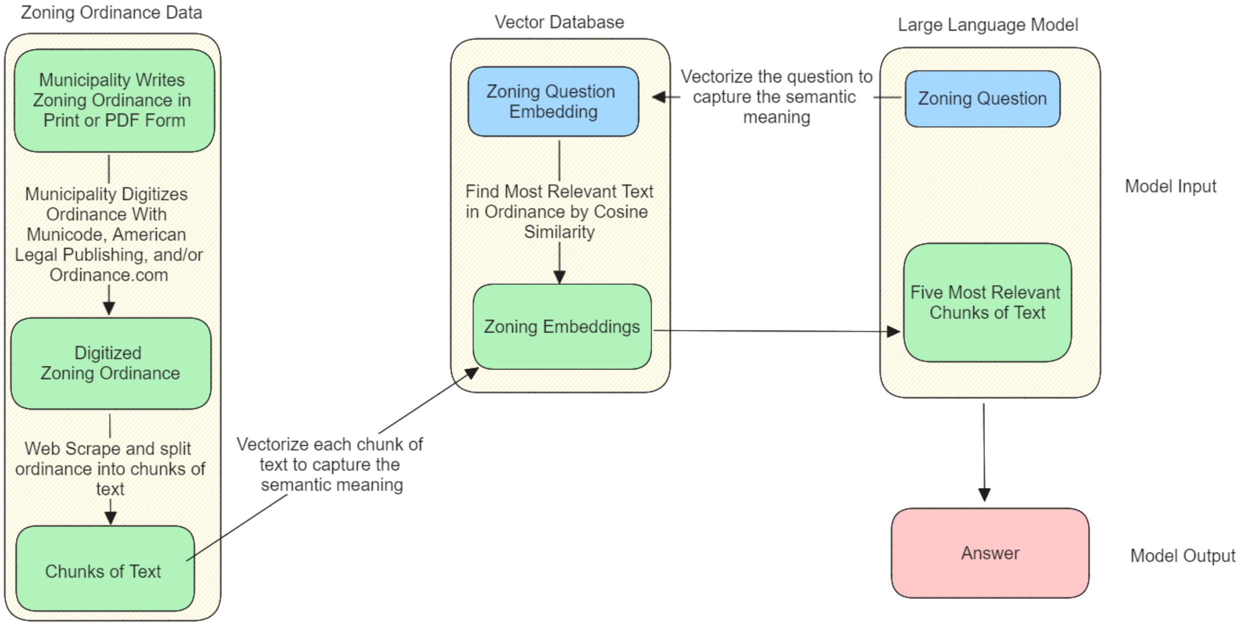
Panel C: Categorical Questions - Adjusted Percent Accuracy

Question	Llama-13B	Llama-70B	Chat GPT 3.5	Chat GPT 4
Which entity acts as the special permit granting authority for multi-family housing?	23%	23%	24%	29%
If the municipality requires special permits for accessory apartments, which entity is the special permit granting authority?	26%	26%	45%	43%
Which entity is the special permit granting authority for cluster/flexible zoning?	21%	18%	32%	36%
Cumulative Average	23%	22%	34%	36%
Cumulative Median	23%	23%	33%	36%

Note: Categorical answers are a list of entities. The Adjusted Percent Accuracy rewards the model for correctly finding entities in the Pioneer Institute answer while also penalizing it for guessing entities that are not in the Pioneer Institute answer. Specifically, we define the Adjusted Percent Accuracy as the ratio of the number of entities the model found correctly divided by the number of entities in the union of the answers from both the model and the Pioneer Institute. The Cumulative Average and Cumulative Median are calculated across questions giving equal weight to each question.

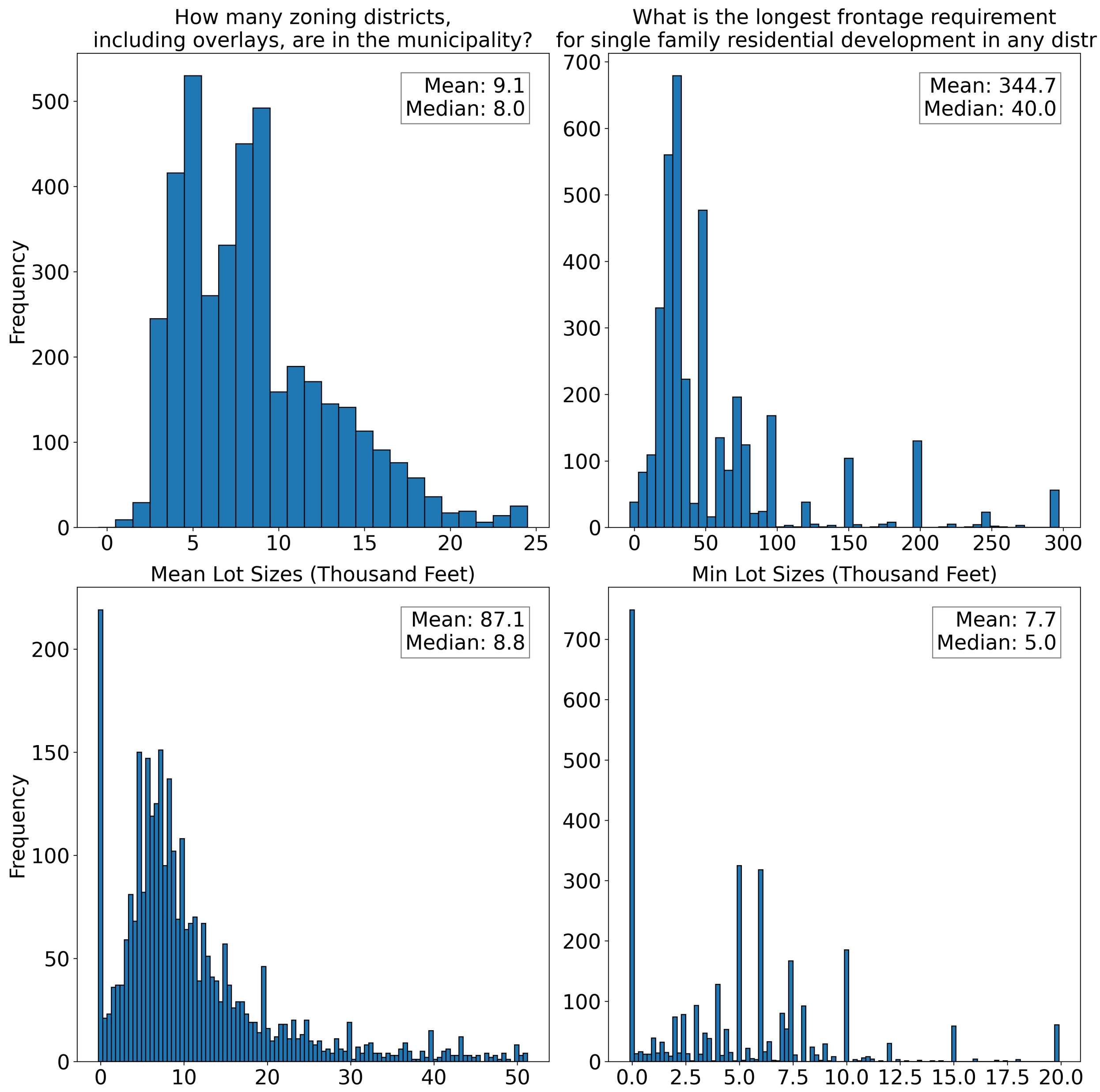
Figures

Figure 1: Model Overview



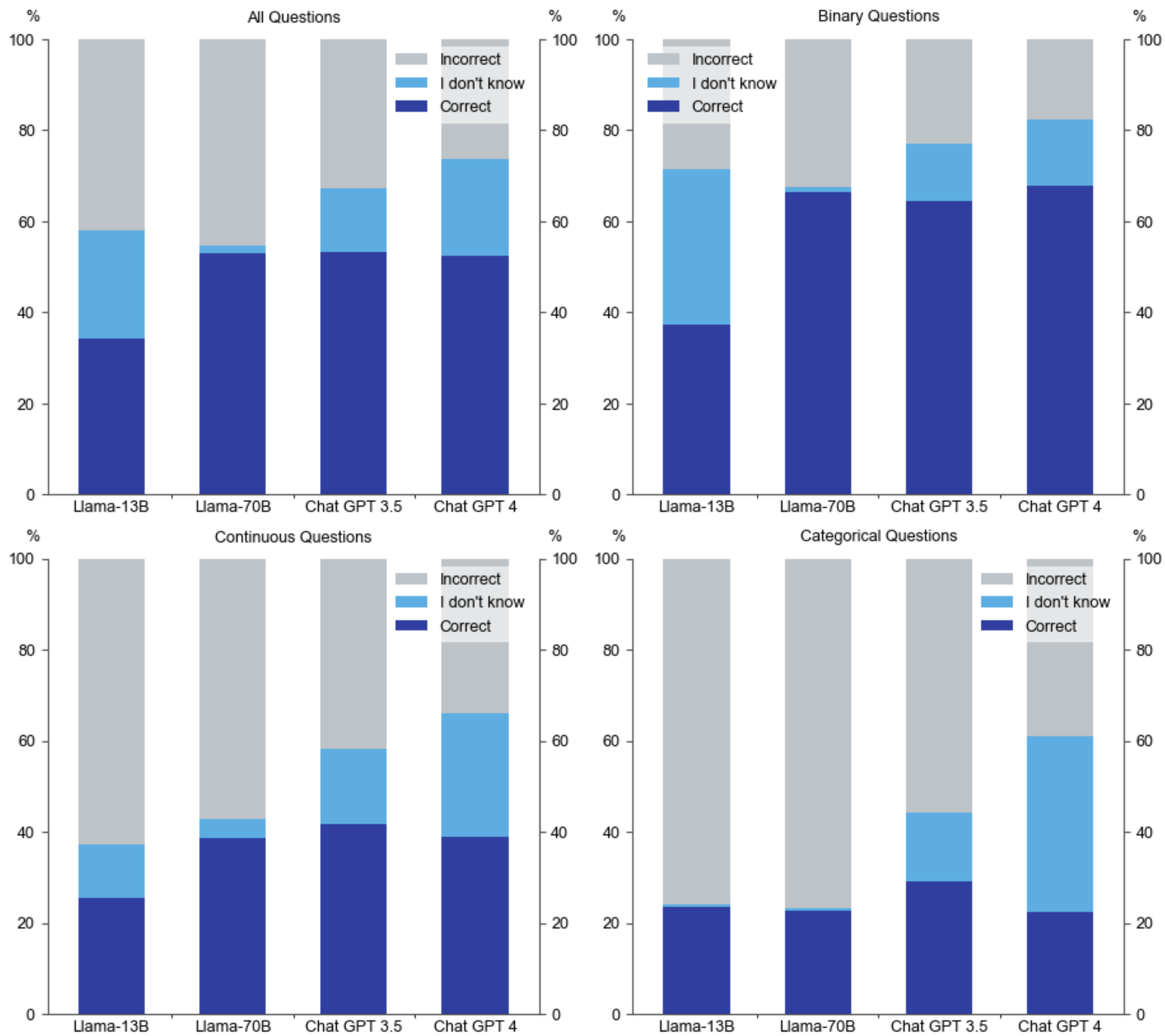
Note: Ordinances from digital aggregators (Municode, American Legal Publishing, and Ordinance.com) are either entirely about zoning, partially about zoning (i.e. have one or more sections about zoning), or not about zoning at all. We filter out ordinances not at all about zoning by searching through key phrases, table headers, and zoning district names (i.e. R-1 for the first residential zoning district). Sometimes digital aggregators leave tables in image form, especially the aggregator Ordinance.com. So that the model can still read the table, we transcribe images of tables using Amazon Textract. We split the text file for each ordinance into chunks of 2100 characters with 200 characters of overlap between chunks using a recursive character text splitter. We vectorized each chunk of text using the multi-qa-mpnet-base-dot-v1 embedding algorithm. The format of the model output, or 'Answer', depends on which type of model we use. For Llama-2 based models we elicit an open-ended response to each question and then use Kor to parse out a structured answer (i.e. to ascertain whether an answer is "Yes", "No", or "I don't know" to a binary question). For Chat GPT based models we simultaneously elicit structured answers along with an explanation for them in one shot.

Figure 2: Distribution of Housing Regulations



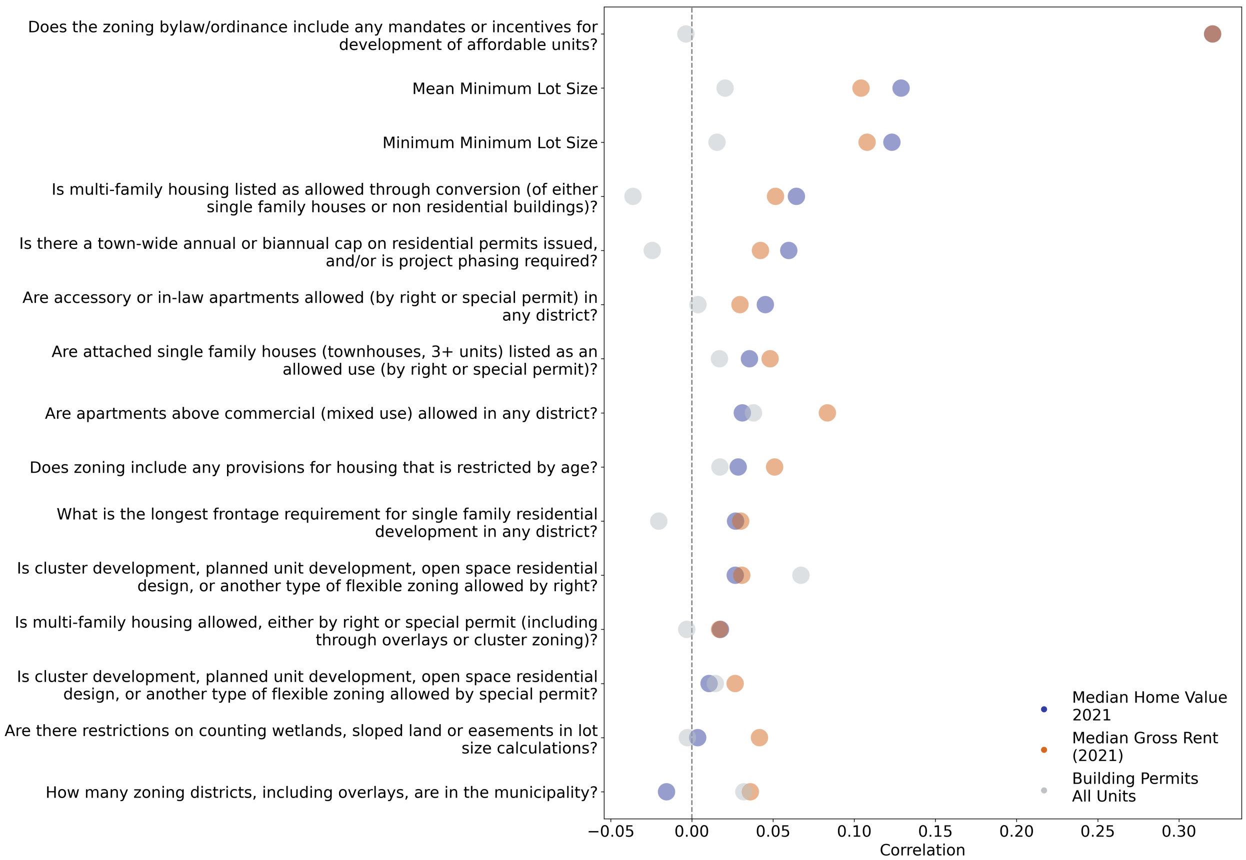
Note: See table 2 footnote for details on the sample. X-axis stops at 99th percentile for upper left chart and 95th percentile for the other charts. Mean and median include all data.

Figure 3: Comparison of Cumulative Performance Across Models



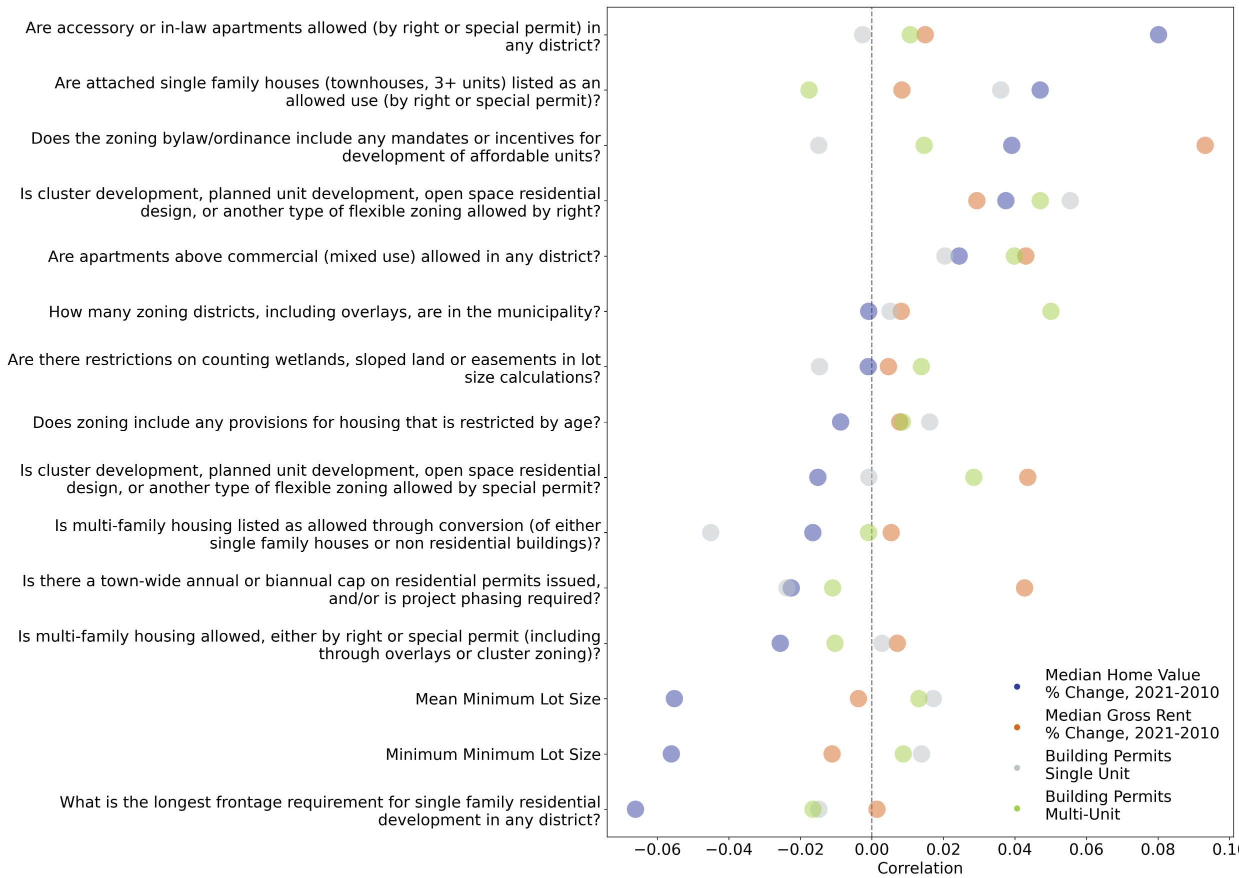
Note: For binary questions we use the percent accuracy, for continuous questions we use the correlation, and for categorical questions we use the adjusted percent accuracy metric. For detailed definitions of these performance metrics please see footnotes of table 3.

Figure 4: Correlation Between Median Gross Rents, Median Home Values, Building Permits Per Capita and Zoning Regulations



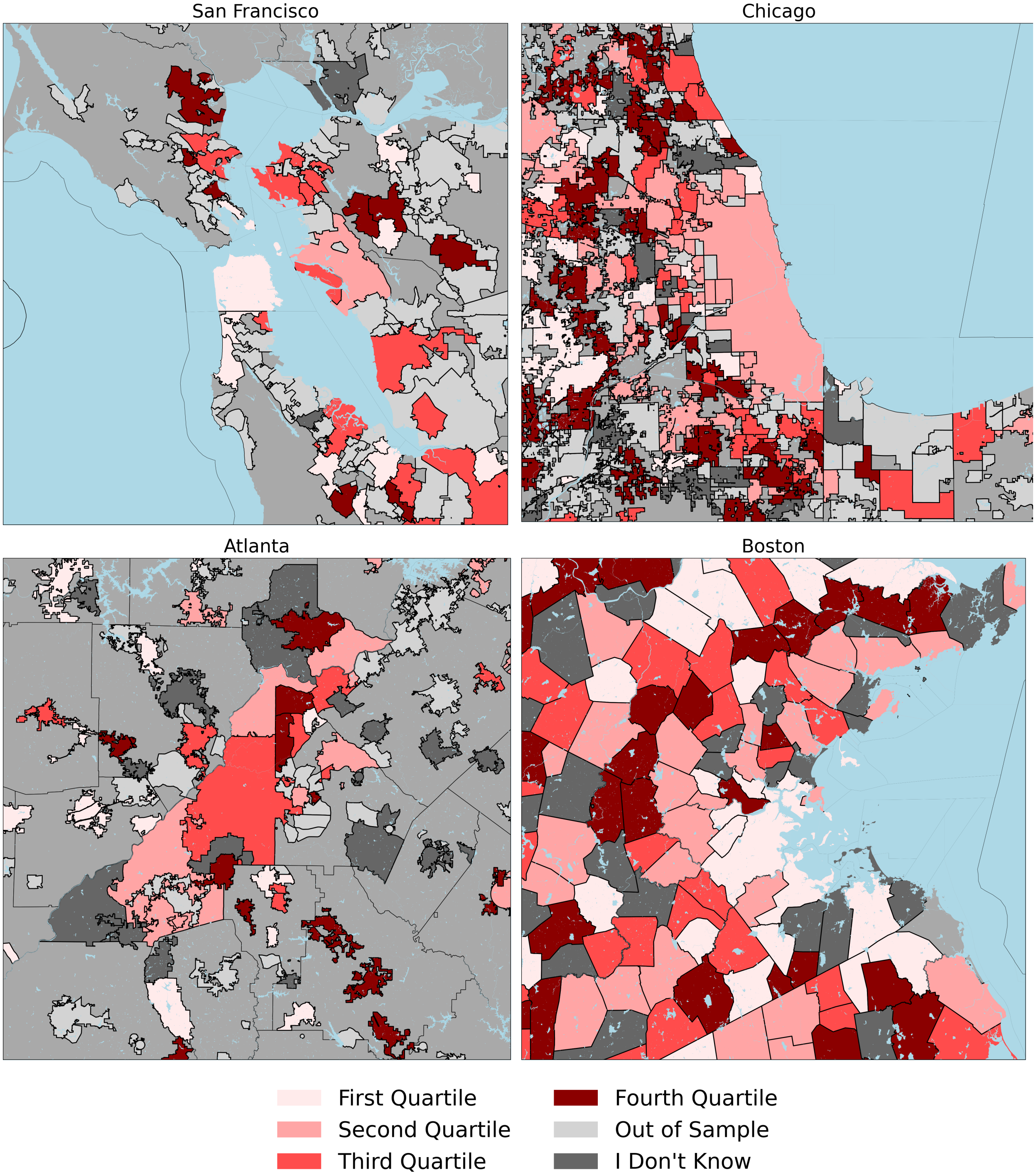
Note: Univariate correlations are calculated over all valid municipality question pairs (i.e. where the model does not say "I don't know") with a valid outcome variable (i.e. not missing) over our national sample with Llama-2 13B. We winsorize continuous variable answers from our model at the 5% level, but do not winsorize housing outcomes data. Median Gross Rent data comes from both the 2021 and 2010 Five-Year American Community Surveys we use median gross rent (B25064_001E). We have valid rent data for 98.3% of our 4090 municipalities in 2021 and in both 2010 and 2021 for 97.6% of our 4090 municipalities. Median Home Value data comes from both the 2021 and 2010 Five-Year American Community Surveys we use median home value (B25077_001E). We have valid home value data for 99.7% of our 4090 municipalities in 2021 and in both 2010 and 2021 for 99.4% of our 4090 municipalities. Building permits data comes from the 2022 Census Building Permits Survey we use the estimated number of units permitted in 2022. Multi-Unit covers any building with 2-units or more. We have valid building permits data for 81.2% of our 4090 municipalities.

Figure 5: Correlations Between Changes in Median Gross Rents, Changes in Median Home Value, Single-Family Building Permits, Multi-Family Building Permit Units and Zoning Regulations



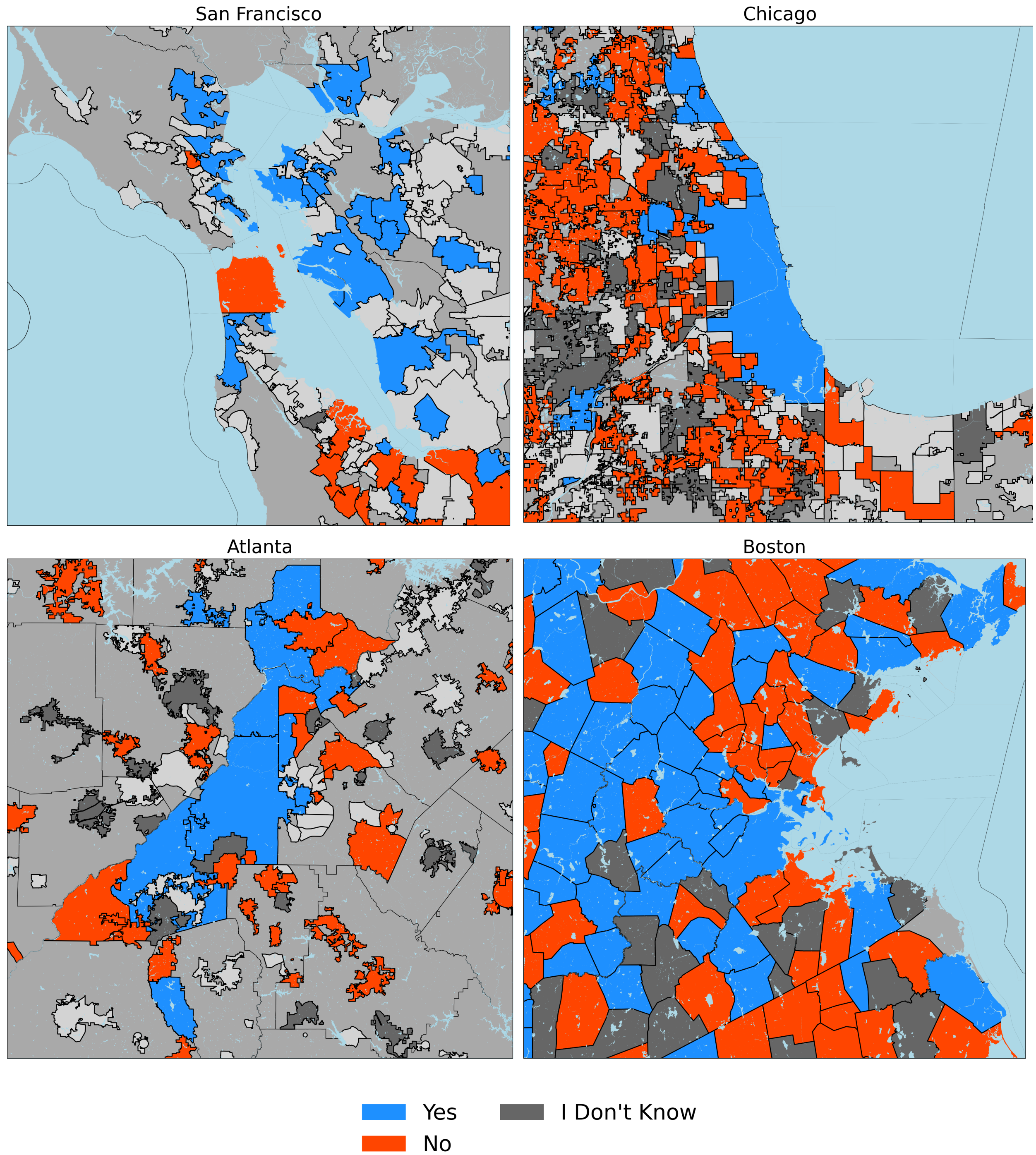
Note: Univariate correlations are calculated over all valid municipality question pairs (i.e. where the model does not say "I don't know") with a valid outcome variable (i.e. not missing) over our national sample with Llama-2 13B. We winsorize continuous variable answers from our model at the 5% level, but do not winsorize housing outcomes data. Median Gross Rent data comes from both the 2021 and 2010 Five-Year American Community Surveys we use median gross rent (B25064_001E). We have valid rent data for 98.3% of our 4090 municipalities in 2021 and in both 2010 and 2021 for 97.6% of our 4090 municipalities. Median Home Value data comes from both the 2021 and 2010 Five-Year American Community Surveys we use median home value (B25077_001E). We have valid home value data for 99.7% of our 4090 municipalities in 2021 and in both 2010 and 2021 for 99.4% of our 4090 municipalities. Building permits data comes from the 2022 Census Building Permits Survey we use the estimated number of units permitted in 2022. Multi-Unit covers any building with 2-units or more. We have valid building permits data for 81.2% of our 4090 municipalities.

Figure 6: Minimum Minimum Lot Size Quartiles For Select Metropolitan Areas



Note: Each map shows roughly a 100km × 100km square area, except for Boston where we show a 75km × 75km square area. Within each map we plot all Census-designated places, except for Boston where we also plot Census county subdivisions that correspond with townships. Both Census-designated place and Census county subdivisions data comes from the 2022 Census TIGER/Line shape files.

Figure 7: Whether There Are Mandates or Incentives For The Development of Affordable Units in Select Metropolitan Areas



Note: Each map shows roughly a 100km x 100km square area, except for Boston where we show a 75km x 75km square area. Within each map we plot all Census-designated places, except for Boston where we also plot Census county subdivisions that correspond with townships. Both Census-designated place and Census county subdivisions data comes from the 2022 Census TIGER/Line shape files.

A Data Appendix

Table A1: Percentage of 'I don't know' Responses in Training Data

Panel A: Continuous Questions

Question	Llama-13B	Llama-70B	Chat GPT 3.5	Chat GPT 4
How many zoning districts, including overlays, are in the municipality?	0.9	3.7	15.0	13.1
What is the longest frontage requirement for single family residential development in any district?	2.8	1.9	10.3	36.4
What is the Minimum Lot Size For Each District?	21.5	5.6	20.6	29.0
Cumulative Average	8.4	3.7	15.3	26.2

Panel B: Categorical Questions

Question	Llama-13B	Llama-70B	Chat GPT 3.5	Chat GPT 4
If the municipality requires special permits for accessory apartments, which entity is the special permit granting authority?	0.9	0.9	3.7	23.4
Which entity is the special permit granting authority for cluster/flexible zoning?	0.0	0.0	31.8	61.7
Which entity acts as the special permit granting authority for multi-family housing?	0.0	0.9	10.3	30.8

Panel C: Binary Questions

Question	Llama-13B	Llama-70B	Chat GPT 3.5	Chat GPT 4
Are accessory or in-law apartments allowed (by right or special permit) in any district?	60.7	0.0	40.2	46.7
Is cluster development, planned unit development, open space residential design, or another type of flexible zoning allowed by right?	58.9	0.9	17.8	17.8
Is cluster development, planned unit development, open space residential design, or another type of flexible zoning allowed by special permit?	59.8	0.0	5.6	18.7
Does the zoning bylaw/ordinance include any mandates or incentives for development of affordable units?	23.4	2.8	0.0	0.0
Is there a town-wide annual or biannual cap on residential permits issued, and/or is project phasing required?	35.5	1.9	0.9	0.0
Are there restrictions on counting wetlands, sloped land or easements in lot size calculations?	9.3	0.0	5.6	11.2
Is multi-family housing allowed, either by right or special permit (including through overlays or cluster zoning)?	34.6	1.9	20.6	7.5
Are apartments above commercial (mixed use) allowed in any district?	11.2	0.0	30.8	47.7
Is multi-family housing listed as allowed through conversion (of either single family houses or non residential buildings)?	33.6	0.9	1.9	0.9
Are attached single family houses (townhouses, 3+ units) listed as an allowed use (by right or special permit)?	38.3	0.9	15.0	10.3
Does zoning include any provisions for housing that is restricted by age?	12.1	1.9	0.0	0.0

Table A5: Univariate Correlation Between Zoning Regulations and Housing Outcomes

Panel A: Continuous Questions

	Median Gross Rent		**Median Home Value**		**Building Permits per Capita**		
Question	2021	% Chg	2021	% Chg	Single	Multi	Total
How many zoning districts, including overlays, are in the municipality?	0.04	0.01	-0.02	-0.00	0.01	0.05	0.03
What is the longest frontage requirement for single family residential development in any district?	0.03	0.00	0.03	-0.07	-0.01	-0.02	-0.02
Minimum Minimum Lot Size	0.11	-0.01	0.12	-0.06	0.01	0.01	0.02
Mean Minimum Lot Size	0.10	-0.00	0.13	-0.06	0.02	0.01	0.02

Note: Univariate correlations are calculated over all valid municipality question pairs (i.e. where the model does not say "I don't know") with a valid outcome variable (i.e. not missing) over our national sample with Llama-2 13B. We winsorize continuous variable answers from our model at the 5% level, but do not winsorize housing outcomes data. Median Gross Rent data comes from both the 2021 and 2010 Five-Year American Community Surveys we use median gross rent (B25064_001E). We have valid rent data for 98.3% of our 4090 municipalities in 2021 and in both 2010 and 2021 for 97.6% of our 4090 municipalities. Median Home Value data comes from both the 2021 and 2010 Five-Year American Community Surveys we use median home value (B25077_001E). We have valid home value data for 99.7% of our 4090 municipalities in 2021 and in both 2010 and 2021 for 99.4% of our 4090 municipalities. Building permits data comes from the 2022 Census Building Permits Survey we use the estimated number of units permitted in 2022. Multi-Unit covers any building with 2-units or more. We have valid building permits data for 81.2% of our 4090 municipalities.

Panel B: Binary Questions

Question	Median Gross Rent 2021	Median Gross Rent % Chg	Median Home Value 2021	Median Home Value % Chg	Building Permits per Capita Single	Building Permits per Capita Multi	Building Permits per Capita Total
Is multi-family housing allowed, either by right or special permit (including through overlays or cluster zoning)?	0.02	0.01	0.02	-0.03	0.00	-0.01	-0.00
Are apartments above commercial (mixed use) allowed in any district?	0.08	0.04	0.03	0.02	0.02	0.04	0.04
Is multi-family housing listed as allowed through conversion (of either single family houses or non residential buildings)?	0.05	0.01	0.06	-0.02	-0.05	-0.00	-0.04
Are attached single family houses (townhouses, 3+ units) listed as an allowed use (by right or special permit)?	0.05	0.01	0.04	0.05	0.04	-0.02	0.02
Does zoning include any provisions for housing that is restricted by age?	0.05	0.01	0.03	-0.01	0.02	0.01	0.02
Are accessory or in-law apartments allowed (by right or special permit) in any district?	0.03	0.01	0.05	0.08	-0.00	0.01	0.00
Is cluster development, planned unit development, open space residential design, or another type of flexible zoning allowed by right?	0.03	0.03	0.03	0.04	0.06	0.05	0.07
Is cluster development, planned unit development, open space residential design, or another type of flexible zoning allowed by special permit?	0.03	0.04	0.01	-0.02	-0.00	0.03	0.01
Does the zoning bylaw/ordinance include any mandates or incentives for development of affordable units?	0.32	0.09	0.32	0.04	-0.01	0.01	-0.00
Is there a town-wide annual or biannual cap on residential permits issued, and/or is project phasing required?	0.04	0.04	0.06	-0.02	-0.02	-0.01	-0.02
Are there restrictions on counting wetlands, sloped land or easements in lot size calculations?	0.04	0.00	0.00	-0.00	-0.01	0.01	-0.00