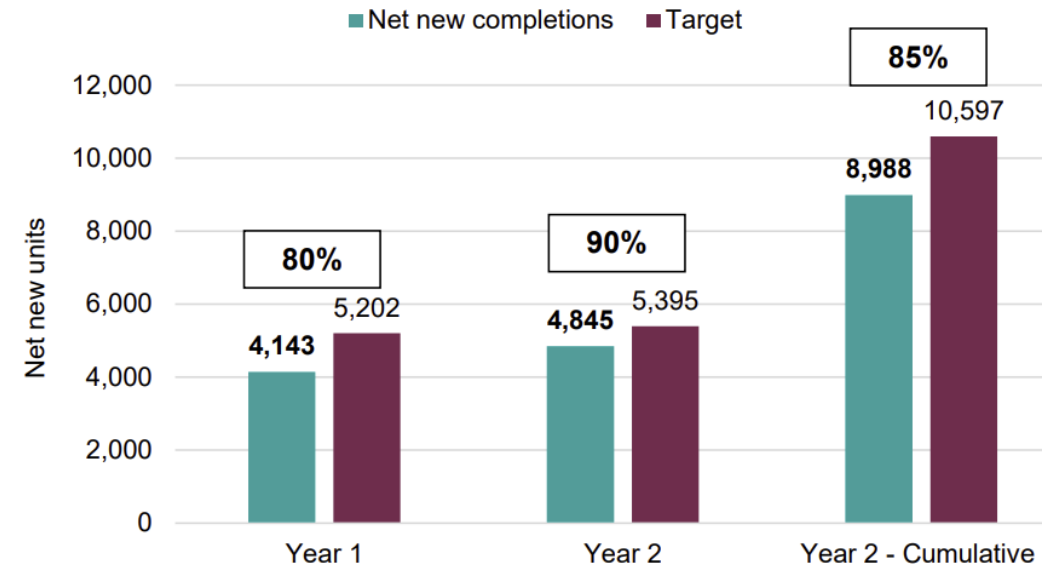


# How do People Talk about New Housing? Leveraging Large Language Models and City Council Records to Analyze Housing Supply Narratives in Vancouver, CA

Julia Harten, Assistant Professor  
Wenzheng Li, Post-doc Research Fellow  
The University of British Columbia

# Background – Why Vancouver? Housing Pressure and Public Engagement

- **Housing pressure:** The City faces strong housing supply pressure and a provincial target of 28,900 net new homes between Oct. 2023 and Sept. 2028 ([Provincial housing legislation](#))
- **Policy change and approval reform:** Provincial legislation is reshaping local approval processes, including limits on public hearings for policy-conforming rezonings (via, e.g., Bill 44, 18).
- **Rich public records:** Rezoning applications generate various public engagement records, including online feedback, staff referral reports, public hearing transcripts, and Council decisions.



Progress Towards Provincial Target Net New Units for Year 1 (Oct 1, 2023 – Sept 30, 2024) and year 2 (Oct 1, 2024 – Sept 30, 2025).

Source: [Vancouver City Council](#)

# Research Gaps

## ***Gap 1: Public engagement as more than “who speaks”***

- literature on public engagement seeks to resolve the gap between procedural inclusion and substantive influence (Arnstein, 1969; Forester, 2006; Corburn, 2006). The key point is not only about who participates, but how to understanding, synthesizing, and integrating these viewpoints into actionable policies and plans (Fu et al., 2024; Innes & Booher, 2004).

## ***Gap 2: Housing Opposition is more than support vs. opposition***

- Existing studies examined whether people support or oppose housing and what factors shape these attitudes (e.g., Manville & Monkkonen 2021; Tighe, 2010). But less is known about how specific arguments are mobilized and how they vary across project types and locations.

## ***Gap 3: Data and method limitations***

- Classic studies on public engagement often rely on surveys, samples, and manual coding, which are valuable, but limited in analyzing public data at scale, and compare arguments across projects (Fu et al., 2024; Rong et al., 2026).

# Research Objective - Mapping Housing Narratives at Scale

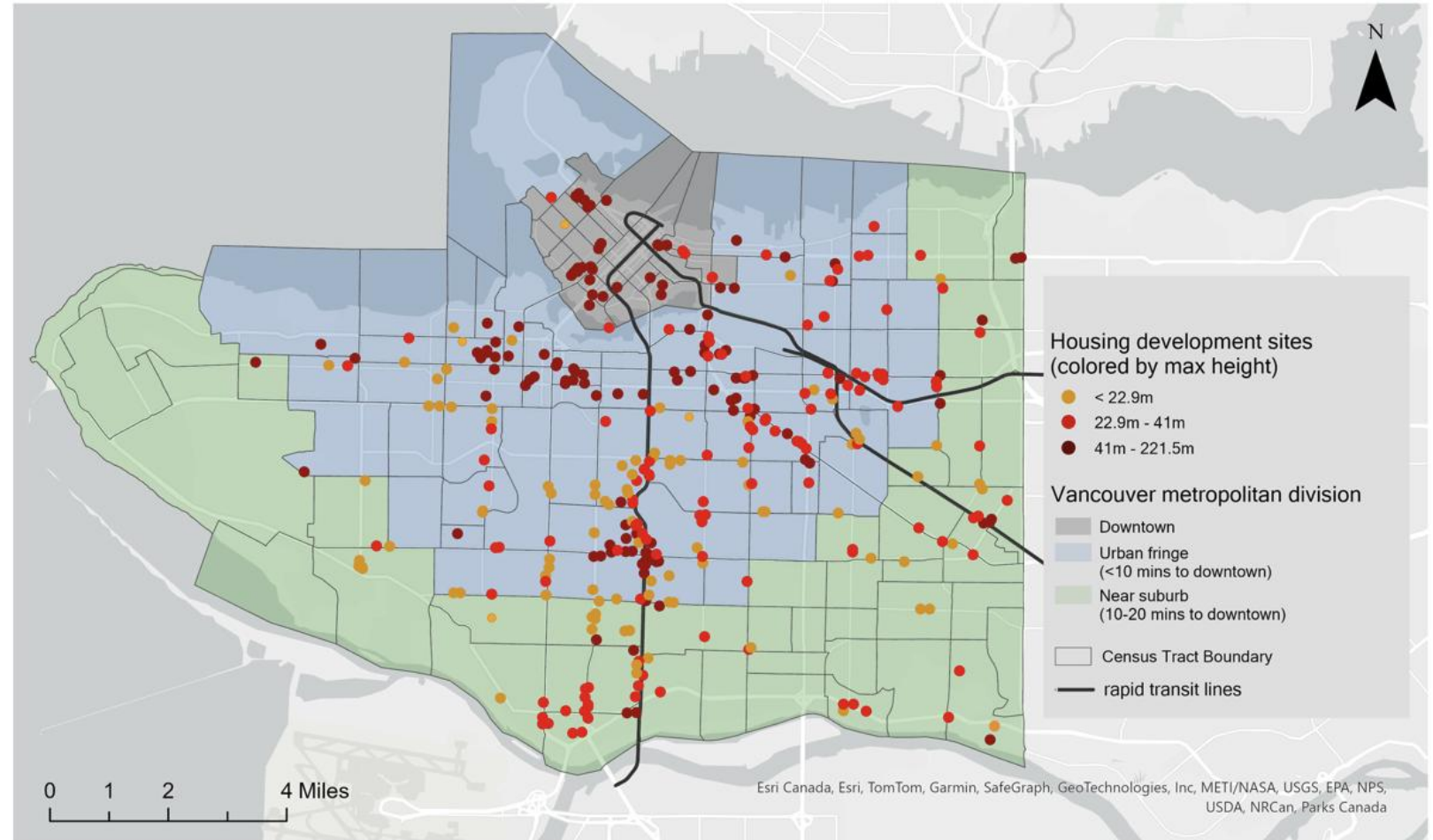
- Leveraging online feedback and LLM-based methods, we aim to better understand how communities think about new housing development, moving beyond “how much support or opposition” to examine with more nuance how different arguments are mobilized across project types and locations.
- Specifically:
  - Q1. What topics do Vancouverites raise when commenting on new housing proposals?
    - LLM-based topic classification and codebook design
  - Q2. Which topics are most frequently mentioned in support and opposition?
    - Descriptive statistics of topic frequency by comment position
  - Q3. How do topic-attitude patterns vary by project characteristics?
    - Multinomial logit models

# Study Area

**350** housing development projects in the City of Vancouver, covering **Jan 2018 - Oct 2025**

## ***Project sample***

- Building height:
  - 95 low-rise,
  - 124 mid-rise,
  - 130 high-rise
- location:
  - 38 at downtown,
  - 228 at urban fringe,
  - 84 at near suburban
- Social housing:
  - 55 social housing projects
  - 295 non-social housing projects



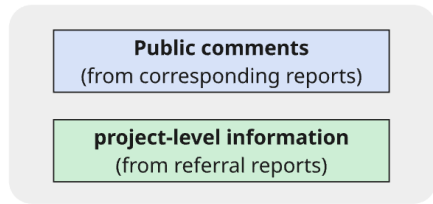
Spatial distribution of 350 housing development projects, categorized by maximum proposed building height.

# Analytical Workflow

1. Data collection

2. codebook + prompt design:

3. prompt refinement (using human labeling sets)



4. Final datasets preparation and cleaning

5. Formal analysis

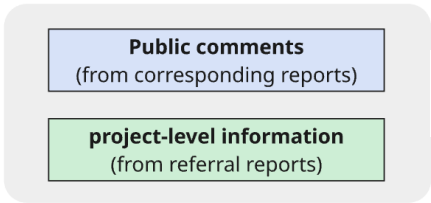
# Analytical Workflow

## 1. Data collection

 VANCOUVER.CA

Council meetings, agendas, minutes, and decisions

*"As a local resident it is already almost impossible to enter and exit the alleyway on this block around the hospital shift change. Adding this new development will mean even more traffic, with no more parking or infrastructure to support the number of people it will bring to the area."*



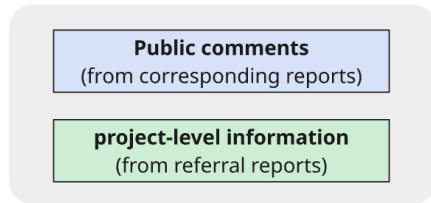
### Purpose and Executive Summary

This report evaluates an application to rezone the site at 816-860 West 13th Avenue and 2915-2925 Willow Street from R3-3 to CD-1 (Comprehensive Development) District. The proposal is for two 21-storey mixed-use buildings with 507 rental units (100% of the dwelling units to be rented at rates no greater than the average rents set for the Class B Development Cost Levy (DCL) waiver) and a 49-space private childcare facility accessible from the ground floor.

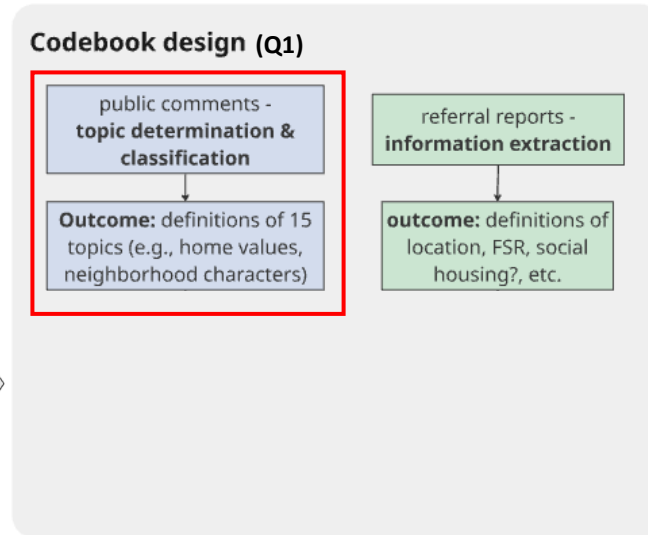
The application meets the intent of the *Broadway Plan* (the "Plan") and associated policies. The General Manager of Planning, Urban Design and Sustainability recommends approval in principle subject to conditions contained in Appendix B.

# Analytical Workflow

## 1. Data collection



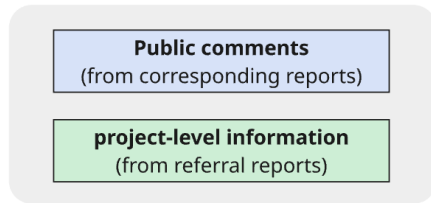
## 2. codebook design:



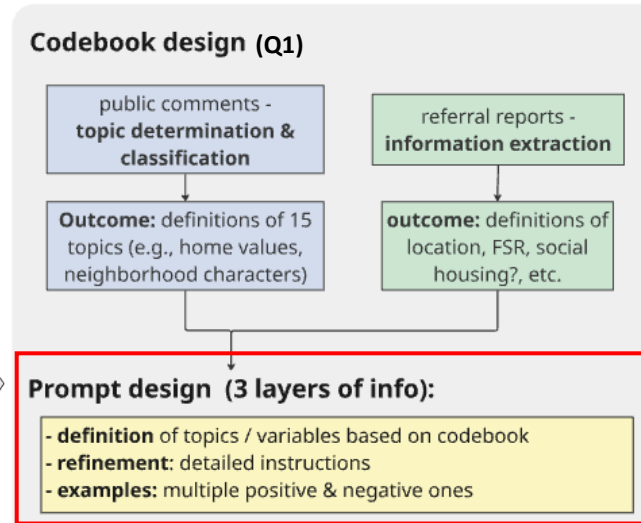
Q1. What **topics** do Vancouverites raise when commenting on new housing proposals?

# Analytical Workflow

## 1. Data collection



## 2. codebook + prompt design:



### "noise & privacy":

#### Definition:

- Concerns focus on disruptions either during construction phase, or concerns from ...

#### Refined definition:

- Label Noise / Privacy = 1 ONLY if the comment explicitly mentions: ...
- Do NOT label Noise / Privacy if the comment only talks about: ...

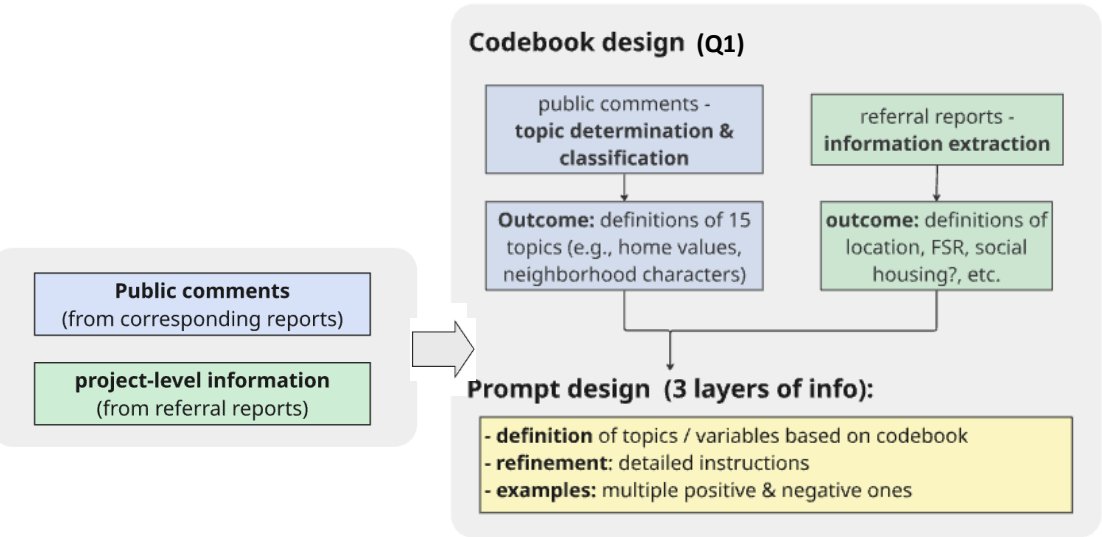
#### Positive & negative examples:

- The noise will make it impossible for us to sleep (positive).
- The tower will block my view of the mountains (negative).

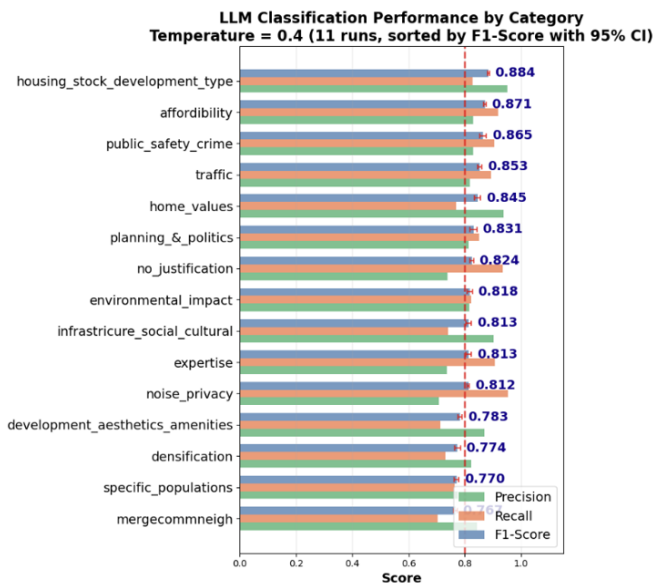
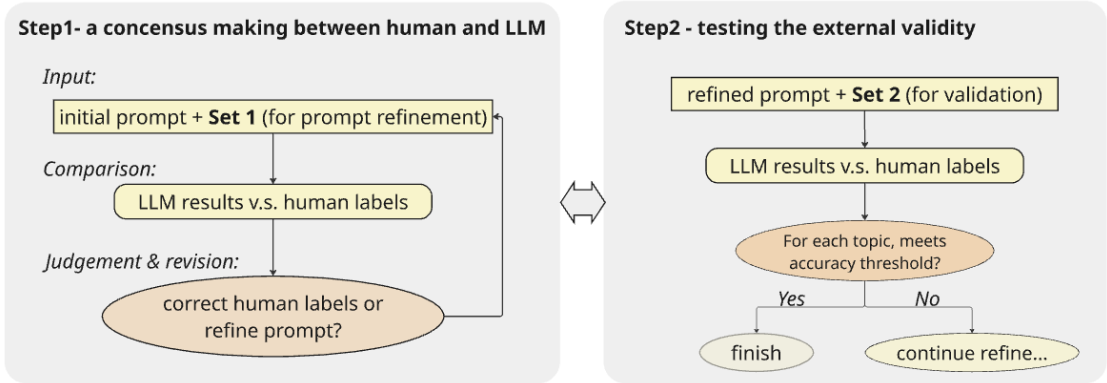
# Analytical Workflow

## 1. Data collection

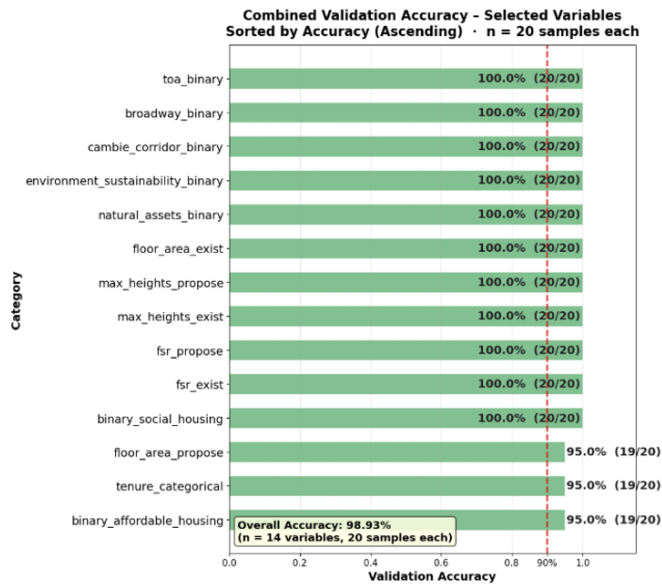
## 2. codebook + prompt design:



## 3. prompt refinement (using human labeling sets)



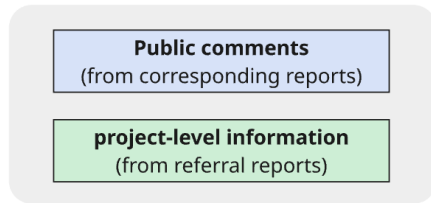
LLM performance in **topic classification** using a temperature setting of 0.4.



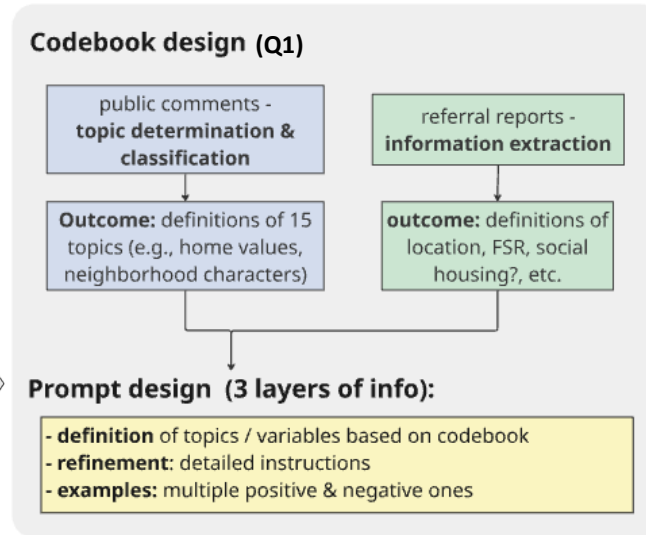
The performance of **variable extraction** from referral/policy reports

# Analytical Workflow

## 1. Data collection

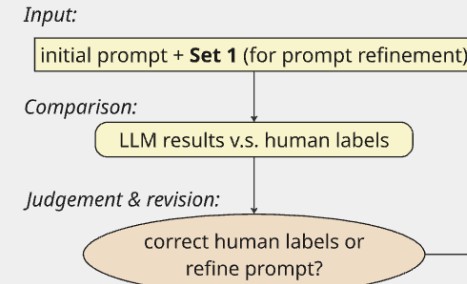


## 2. codebook + prompt design:

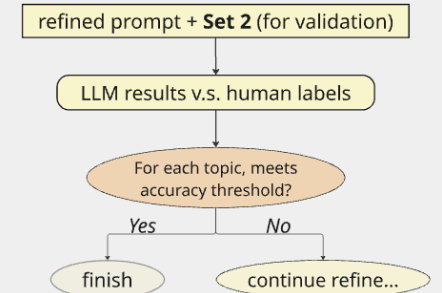


## 3. prompt refinement (using human labeling sets)

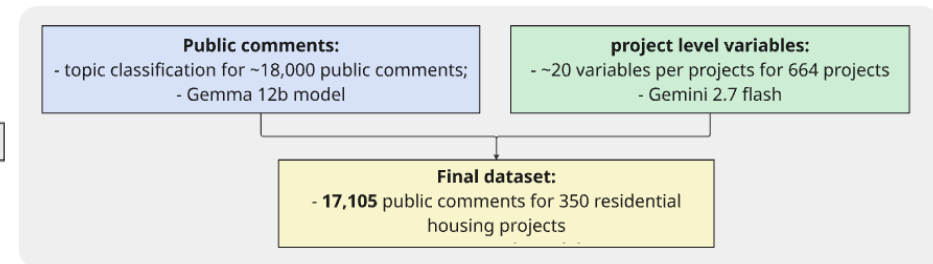
### Step1- a consensus making between human and LLM



### Step2 - testing the external validity



## 4. Final datasets preparation and cleaning



## 5. Formal analysis

### Descriptive statistics (Q2)

- How often are different topics mentioned?

### Multinomial logit model (Q3)

- Modeling whether topics are mentioned in supportive or oppositional comments

# Q1. What topics do Vancouverites raise when commenting on new housing proposals?

- **Theory driven:** Grounded in existing literature and prior conceptual frameworks (Einestín et al., 2024).
- **Data informed:** derived from reviewing and synthesizing public comments
- **Iteratively refined** by combining theory with empirical patterns

## Codebook based on public comments

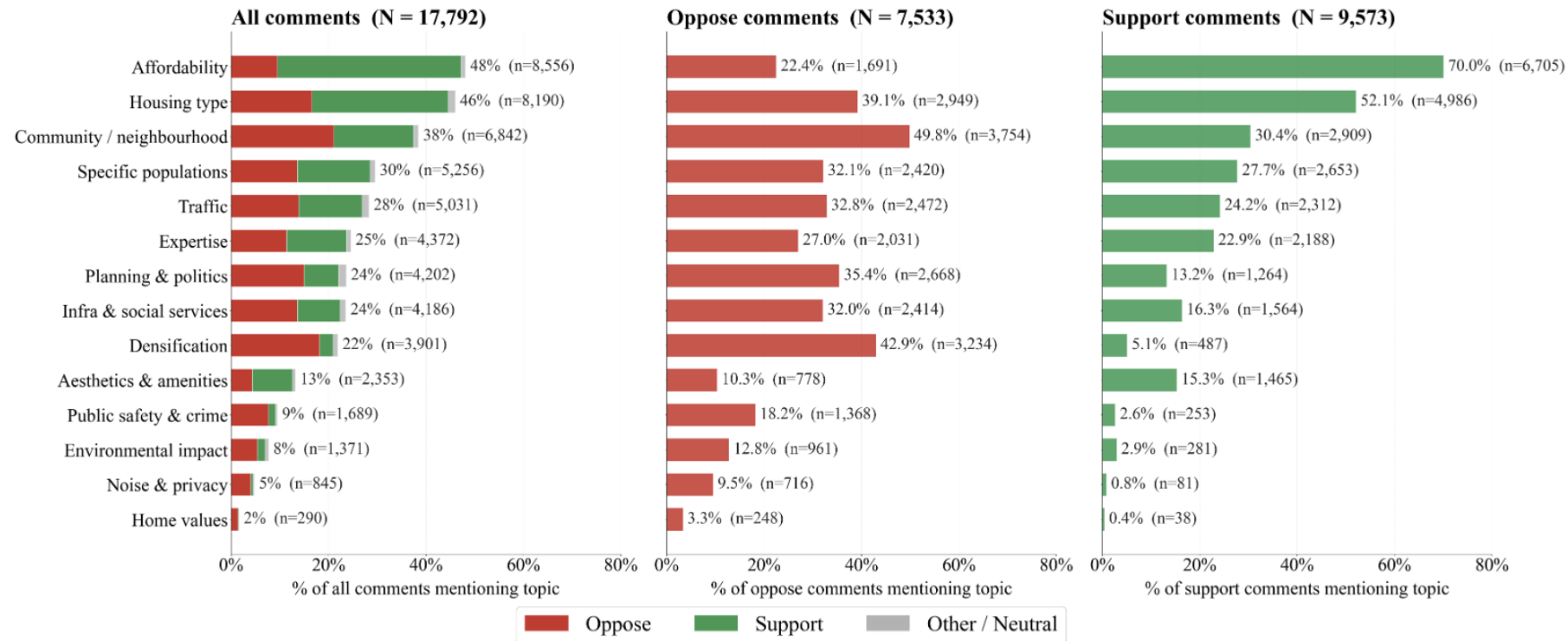
| Topics                                               | Definition                                                                                                                                                                                                                                                                                                                                                                                                  |
|------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Community &amp; neighborhood</b>                  | Arguments examine how the proposed development will affect both the physical character and the social fabric of the neighborhood. Physical character emphasizes the neighborhood's feel, identity, built environment, and quality of life. Social fabric focusses on people, including mentions of changes to the neighborhood's socio-demographic composition, diversity, displacement, and gentrification |
| <b>densification</b>                                 | Arguments address the level of added density the project will bring, including evaluations of whether the proposed density is too high, too low, or appropriateness for the site, with attentions to height and the overall intensity of development in the area.                                                                                                                                           |
| <b>Housing stock &amp; development type</b>          | Arguments examine the specific type of housing being proposed, such as rental units, the size of units, whether the development is residential or commercial in nature, and the number of units dedicated to secured rental or social housing.                                                                                                                                                              |
| <b>affordability</b>                                 | Arguments focus on the cost of housing as being too high—especially exceeding what average people can reasonably afford—, insufficient housing supply, unmet housing needs and demand, and the broader housing crisis.                                                                                                                                                                                      |
| <b>Development aesthetics &amp; amenities</b>        | Arguments focus on the visual appeal and design of the proposed building, debating whether it is aesthetically pleasing or not, as well as the private amenities. also includes concerns about building management and maintenance.                                                                                                                                                                         |
| <b>social infrastructure &amp; cultural services</b> | Arguments examine the adequacy of community services and public amenities, including schools, childcare facilities, access to public services like community centers, libraries, pools, and arts facilities, as well as access to neighborhood greenspace, streetside landscaping and trees, community gardens, mountains, and the ocean.                                                                   |
| <b>Traffic</b>                                       | Arguments address transportation-related impacts, including concerns about cars, traffic congestion, transit accessibility, biking and walking infrastructure/ walkability, and parking availability.                                                                                                                                                                                                       |
| <b>Specific populations</b>                          | Arguments identify how the proposed development will affect particular demographic groups, including populations identified by race, ethnicity, immigration status, homelessness, those with substance use disorders, those with mental illness, First Nations, LGBTQ+ residents, families, children, parents, newcomers, commuters, current residents, young professionals, renters, and foreign investors |
| <b>Home values</b>                                   | Arguments consider how a development proposal might affect property values, land values, or homeowners' financial interests in the surrounding area.                                                                                                                                                                                                                                                        |
| <b>Environmental impact</b>                          | Arguments focus on sustainability considerations, climate change adaptation, disaster risk exacerbation, mitigation strategies, adaptation measures, greenhouse gas emissions, removal/retention of trees, the lack of environmental impact assessment, environmental degradation from development, sustainable construction and demolition management                                                      |
| <b>Public safety &amp; crime</b>                     | Arguments address policing concerns, safety issues, crime rates, and drug use problems that might be associated with or affected by the proposed development.                                                                                                                                                                                                                                               |
| <b>Noise &amp; privacy</b>                           | Concerns focus on disruptions either during the construction phase, as an ongoing result of the proposed development, or ongoing concerns from increasing local population, addressing how the project might affect residents' peace and privacy.                                                                                                                                                           |
| <b>Planning &amp; politics</b>                       | Arguments address procedural and policy issues, including compliance or non-compliance with existing plans and policies, concerns about government and corporate corruption, and comments about the appropriateness of the development approval process. These also include arguments about efficiency or the timeliness of planning processes.                                                             |
| <b>Expertise</b>                                     | This includes strong local, experiential knowledge, i.e. relevant lived experience as well as technical expertise that commenters mention to argue why they have an informed opinion about the proposed development.                                                                                                                                                                                        |

## Q2. Which topics are most frequently mentioned in support and opposition?

### **Top 2 oppositional comments: Community/neighborhood, Density**

- **Community/neighborhood**

- “This does not fit the neighbourhood.” / “It will destroy the character of the neighbourhood.” / the zoning change you’re requesting does not fit the neighbourhood
- opposition is often framed as place protection, with residents defending neighborhood character, social ties, and everyday livability.



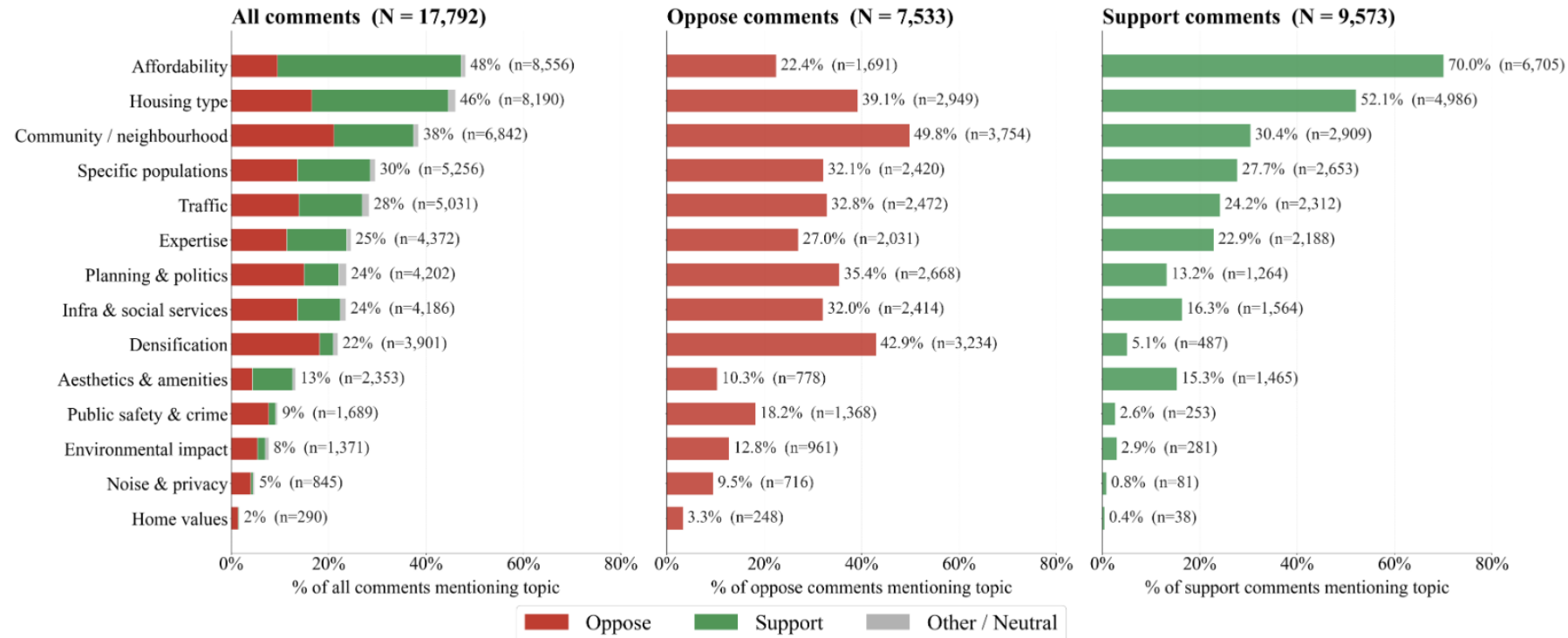
**Panel 1-2** = % of comments within that position group mentioning the topic; **Panel 3** = % of ALL comments mentioning each topic, stacked by position. Topics sorted by overall mention rate (highest at top)

## Q2. Which topics are most frequently mentioned in support and opposition?

### Top 2 oppositional comments: Community/neighborhood, Density

#### • Density

- “I agree we need more rental homes. However, I do not agree with increasing the FSR, height of building or footprint size.” / “The building is too tall and out of scale.” (highly relates to neighborhood concerns)
- Opposition is less about growth itself but more about excessive scale—commenters argue that the proposed height, density, or number of units exceeds local capacity.



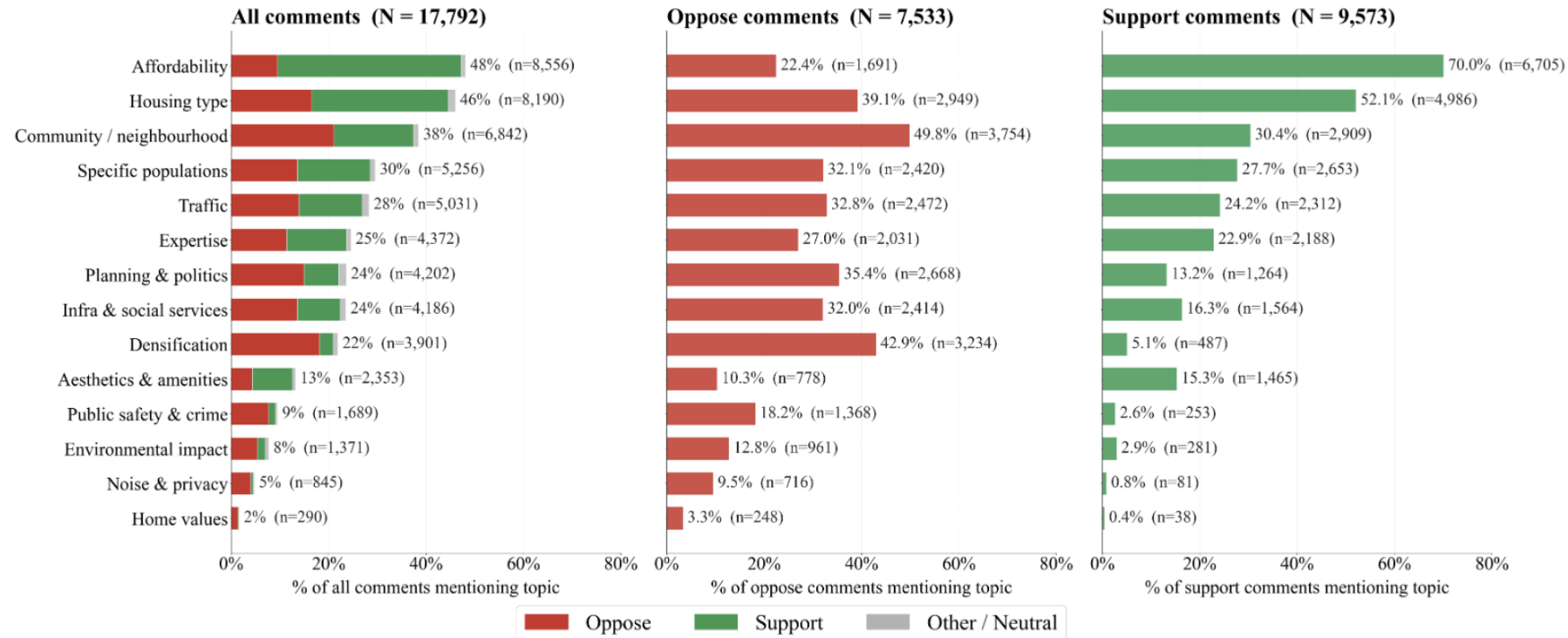
**Panel 1-2** = % of comments within that position group mentioning the topic; **Panel 3** = % of ALL comments mentioning each topic, stacked by position. Topics sorted by overall mention rate (highest at top)

## Q2. Which topics are most frequently mentioned in support and opposition?

### Top 2 supportive comments: Affordability, Housing type

- **affordability**

- “We don’t have enough housing in the city.” / “New supply should be built to ease pressure on rents.”
- support is framed as the response to the housing crisis, with commenters arguing that more housing supply is necessary to reduce rent/price pressure in Vancouver.



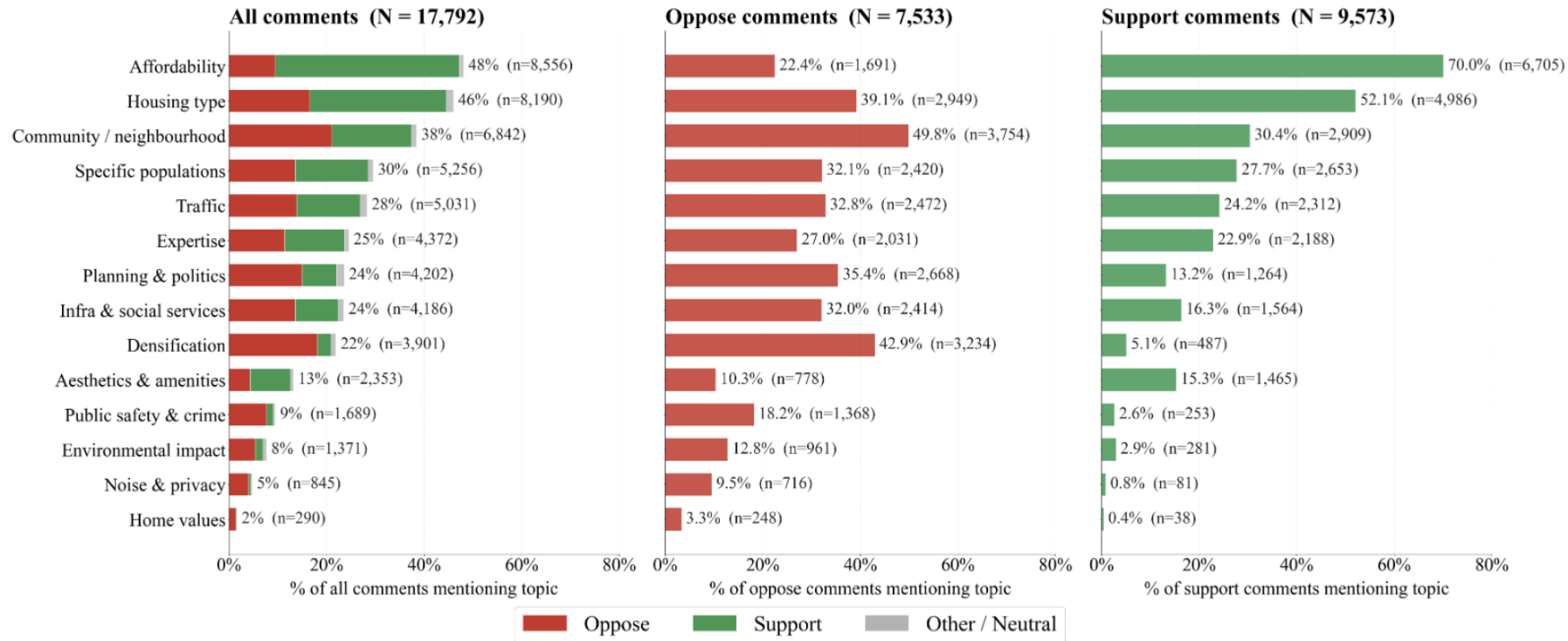
**Panel 1-2** = % of comments within that position group mentioning the topic; **Panel 3** = % of ALL comments mentioning each topic, stacked by position. Topics sorted by overall mention rate (highest at top)

## Q2. Which topics are most frequently mentioned in support and opposition?

### Top 2 supportive comments: Affordability, Housing type

- Housing type**

- ““We need more homes of all types and sizes.” / “Family-oriented secured rental units are exactly the right supply.” / “This neighbourhood needs more secured market rental housing.”
- support emphasizes the need for the “right supply,” especially rental, affordable, mixed-income, and family-oriented housing.



**Panel 1-2** = % of comments within that position group mentioning the topic; **Panel 3** = % of ALL comments mentioning each topic, stacked by position. Topics sorted by overall mention rate (highest at top)

# Q3. How do topic-attitude patterns vary by project characteristics?

## Dependent variable

- For each comment  $i$  on housing project  $j$  and topic  $k$ , define:
  - $Y_{ijk} = \begin{cases} 0, & \text{topic } k \text{ is not mentioned,} \\ 1, & \text{topic } k \text{ mentioned in supportive comments} \\ 2, & \text{topic } k \text{ mentioned in oppositional comments} \end{cases}$

## Multinomial logit model:

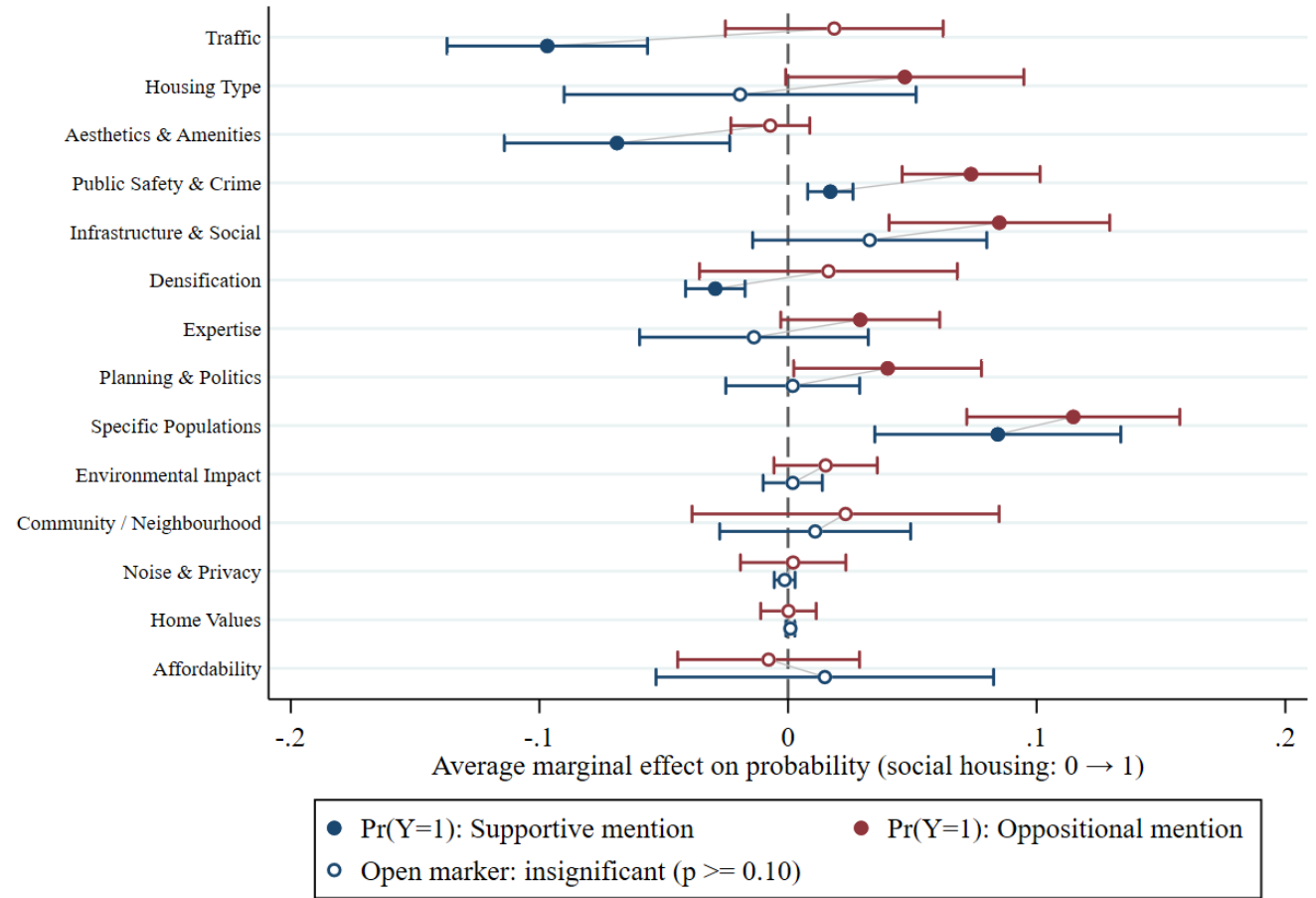
- Use “not mentioned” as the reference category:
  - $\log\left(\frac{\Pr(Y_{ijk}=1)}{\Pr(Y_{ijk}=0)}\right) = \alpha_{1k} + X_j' \beta_{1k} + \gamma_t + \delta_n$
  - $\log\left(\frac{\Pr(Y_{ijk}=2)}{\Pr(Y_{ijk}=0)}\right) = \alpha_{2k} + X_j' \beta_{2k} + \gamma_t + \delta_n$
- where:
  - $X_j$ : a set of **project characteristics – including social housing, Building heights, urban-rural divide, Land use change**;
  - $\gamma_t$ : year fixed effects;
  - $\delta_n$ : neighborhood / area fixed effects.

## What can this model answer?

- Q3. Under what project conditions ( $X_j$ ) are different arguments more likely to be mentioned, and are they mentioned as support or opposition?

# Q3. How social housing (versus non-social) shapes Topic Mentions

- Average marginal effects (AME) on probability (social vs. non-social)



Note: Topics estimated without neighborhood fixed effects: Environmental Impact and Noise & Privacy.

# Q3. How social housing (versus non-social) shapes Topic Mentions

- Average marginal effects (AME) on probability (social vs. non-social)

## Three Key Patterns:

### 1. More opposition in social housing

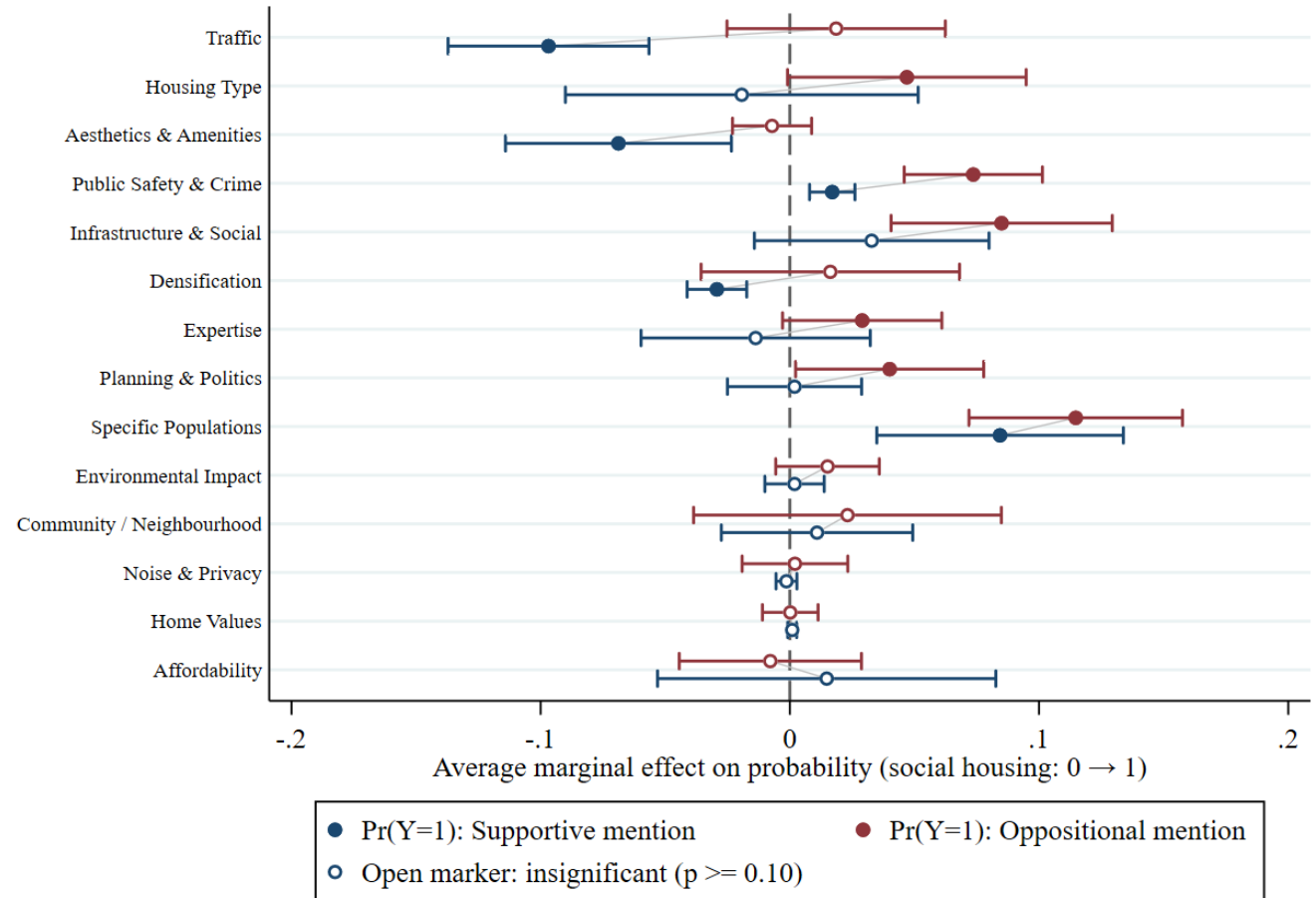
- Social housing projects generate significantly more oppositional mentions on housing type, public safety, social infrastructure, expertise, and planning & politics.

### 2. marginalized topics

- Environmental impact, noise & privacy, and home values do not meaningfully distinguish social from non-social housing projects and are rarely mentioned.

### 3. highly contested topics: both sides speak up

- *Specific populations* and *public safety* appear more often in both supportive and oppositional comments.



Note: Topics estimated without neighborhood fixed effects: Environmental Impact and Noise & Privacy.

# Q3. How social housing (versus non-social) shapes Topic Mentions

- Average marginal effects (AME) on probability (social vs. non-social)

## Public safety

### • **supportive:**

- “Giving people a safe place to sleep will save lives.” / “This project will help battle the opioid crisis.” / “Concerns that social housing will increase crime are overblown and unfounded.”

### • **Oppositional:**

- “This project will attract more crime and drugs.” /
- “The location is too close to schools and seniors’ facilities.”

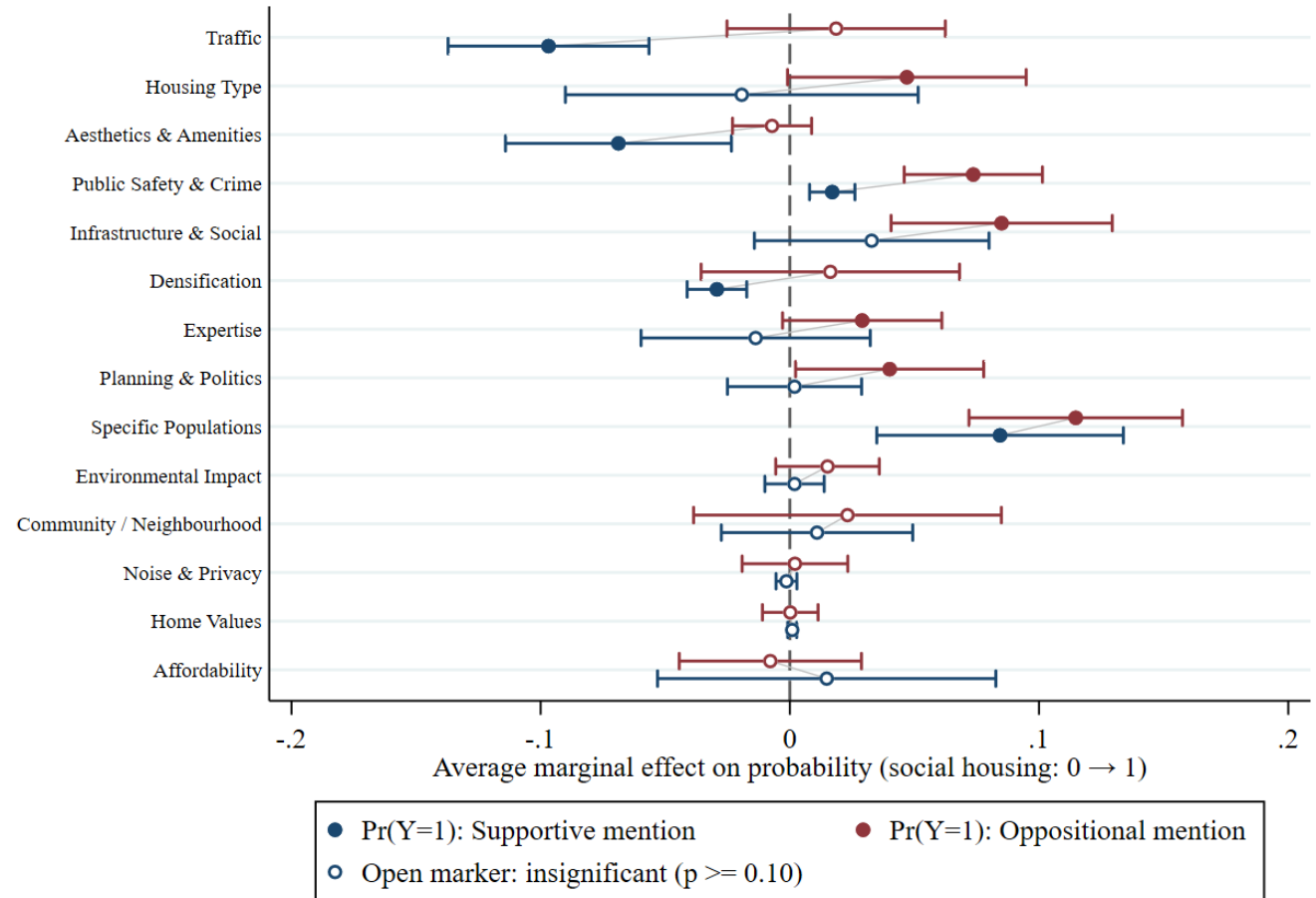
## Specific Population

### • **supportive:**

- “These groups are part of the community and need stable housing, care, and support.” /
- “We have a duty to take care of marginalized and vulnerable community members.”

### • **Oppositional:**

- “The project is not affordable enough for low-income residents.” /
- “The neighbourhood has many families and children, so this is not the right location.”



Note: Topics estimated without neighborhood fixed effects: Environmental Impact and Noise & Privacy.

# Findings and next steps.

## ***Key Findings***

- Opposition centers on neighborhood character and scale, while support emphasizes housing need and affordability.
- Topics like public safety and specific populations are highly contested—both supporters and opponents mobilize these arguments, but with distinct framing.

## ***Next Steps:***

- Associate people's arguments with Vancouver's unique spatial characteristics (e.g., projects within transit-oriented areas, and/or within the zones of city-wide large development plans such as the Broadway Plan and Cambie Corridor Plan).
- Explore how local versus nonlocal commenters differ in their arguments on the same project, to further examine how localism shapes public discourse on housing.

# Questions and Comments

Thank you!

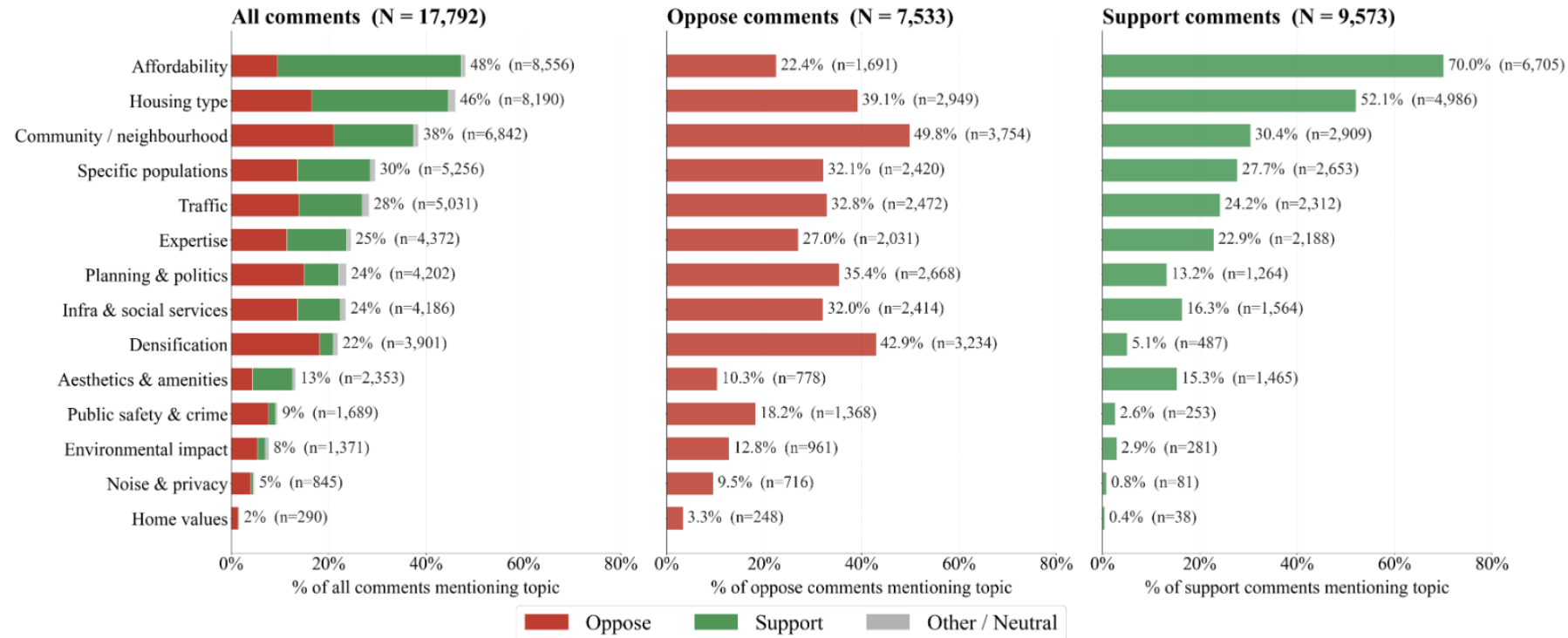
Contact info:

Wenzheng Li ([Wenzheng.li@ubc.ca](mailto:Wenzheng.li@ubc.ca))

## Q2. Which topics are most frequently mentioned in support and opposition?

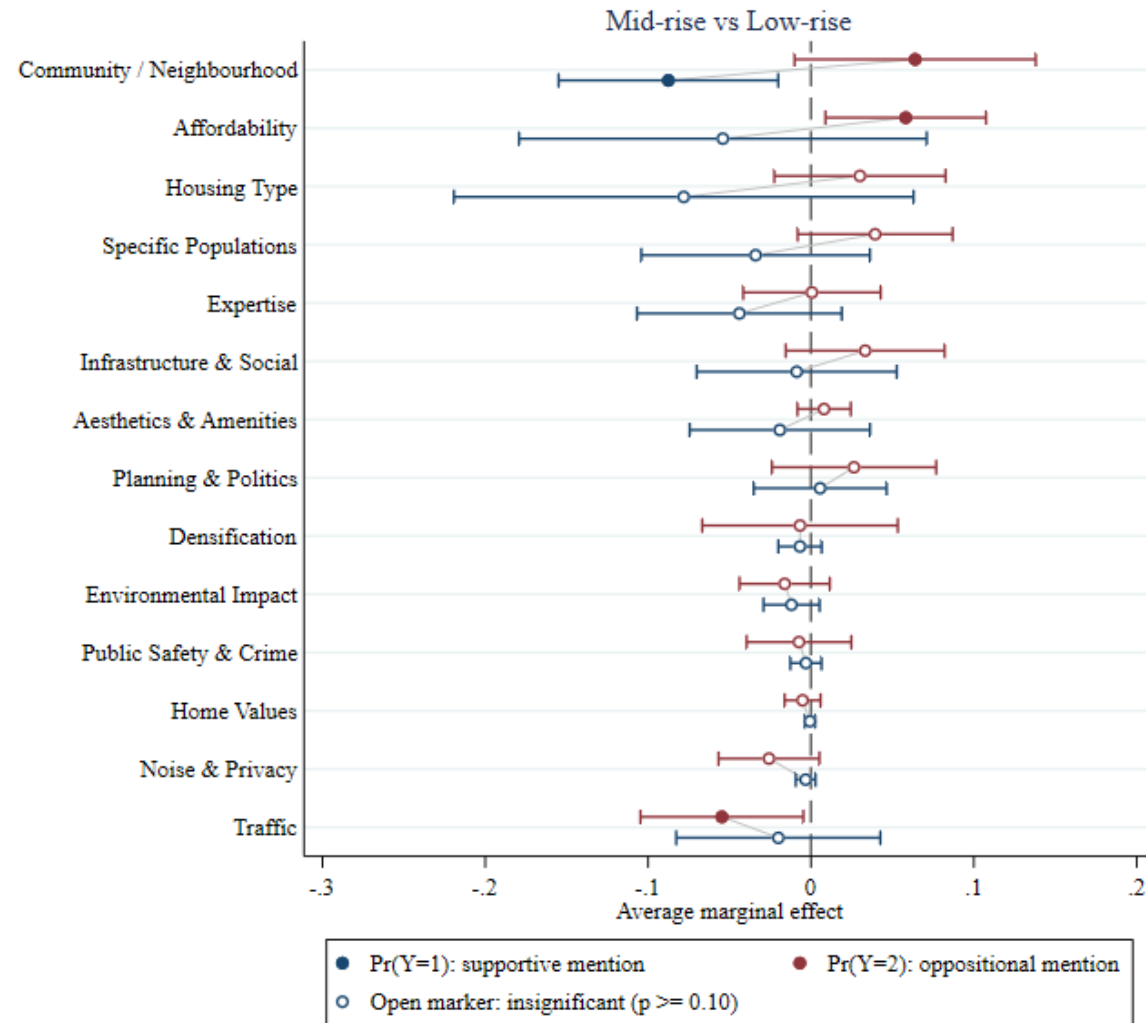
### **Top 2 supportive comments: Affordability, Housing type**

- Community / neighborhood
  - “This project will make the neighbourhood more vibrant and exciting.” / “Central, walkable neighbourhoods like this are the best places to build more homes.” / “The project will provide new residents to support small businesses in the area.”
  - support is framed as community improvement, with commenters arguing that new housing can make neighborhoods more vibrant, inclusive, walkable, and socially diverse.

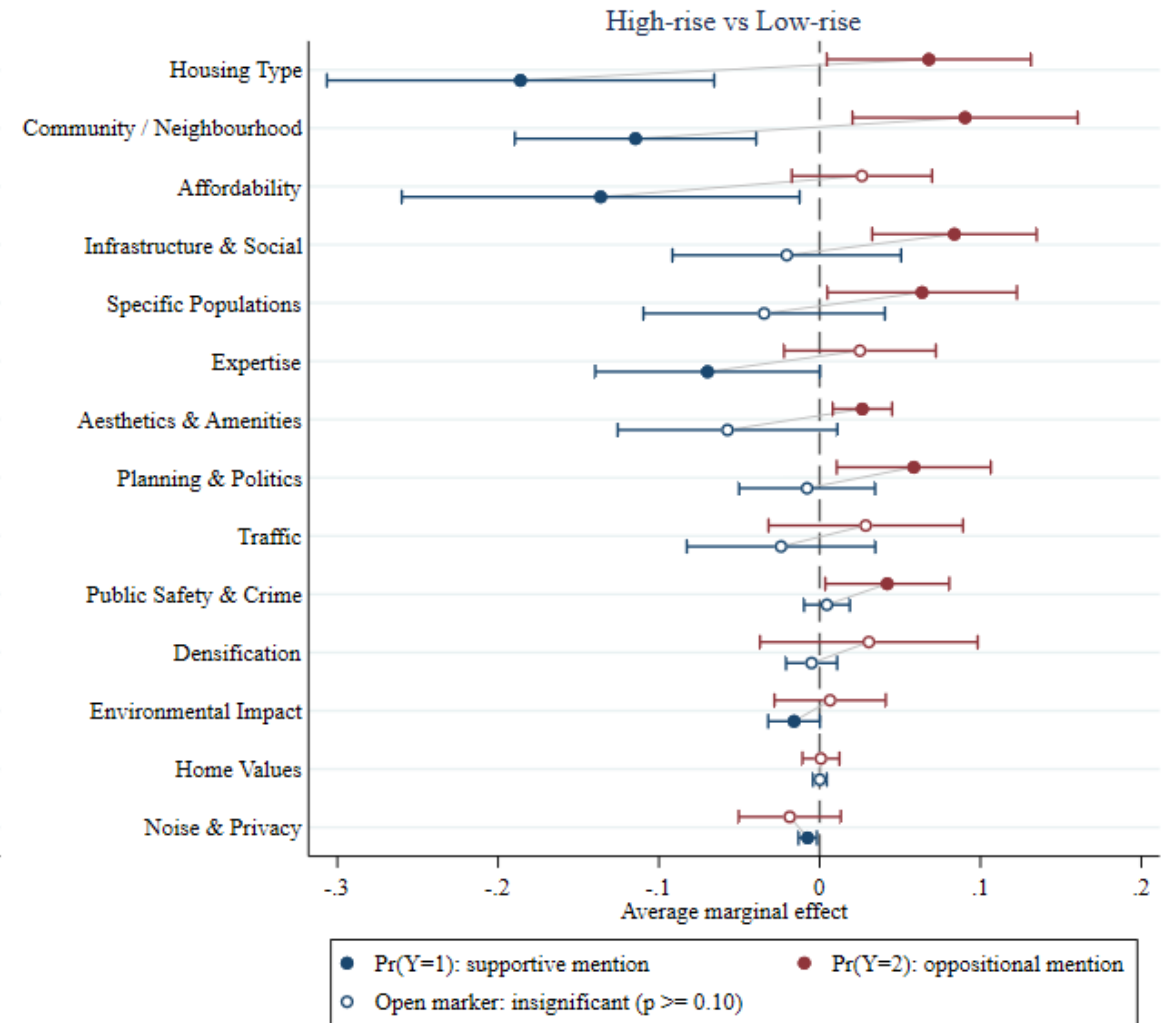


**Panel 1-2** = % of comments within that position group mentioning the topic; **Panel 3** = % of ALL comments mentioning each topic, stacked by position. Topics sorted by overall mention rate (highest at top)

### Q3. AME on probability (Building Heights)

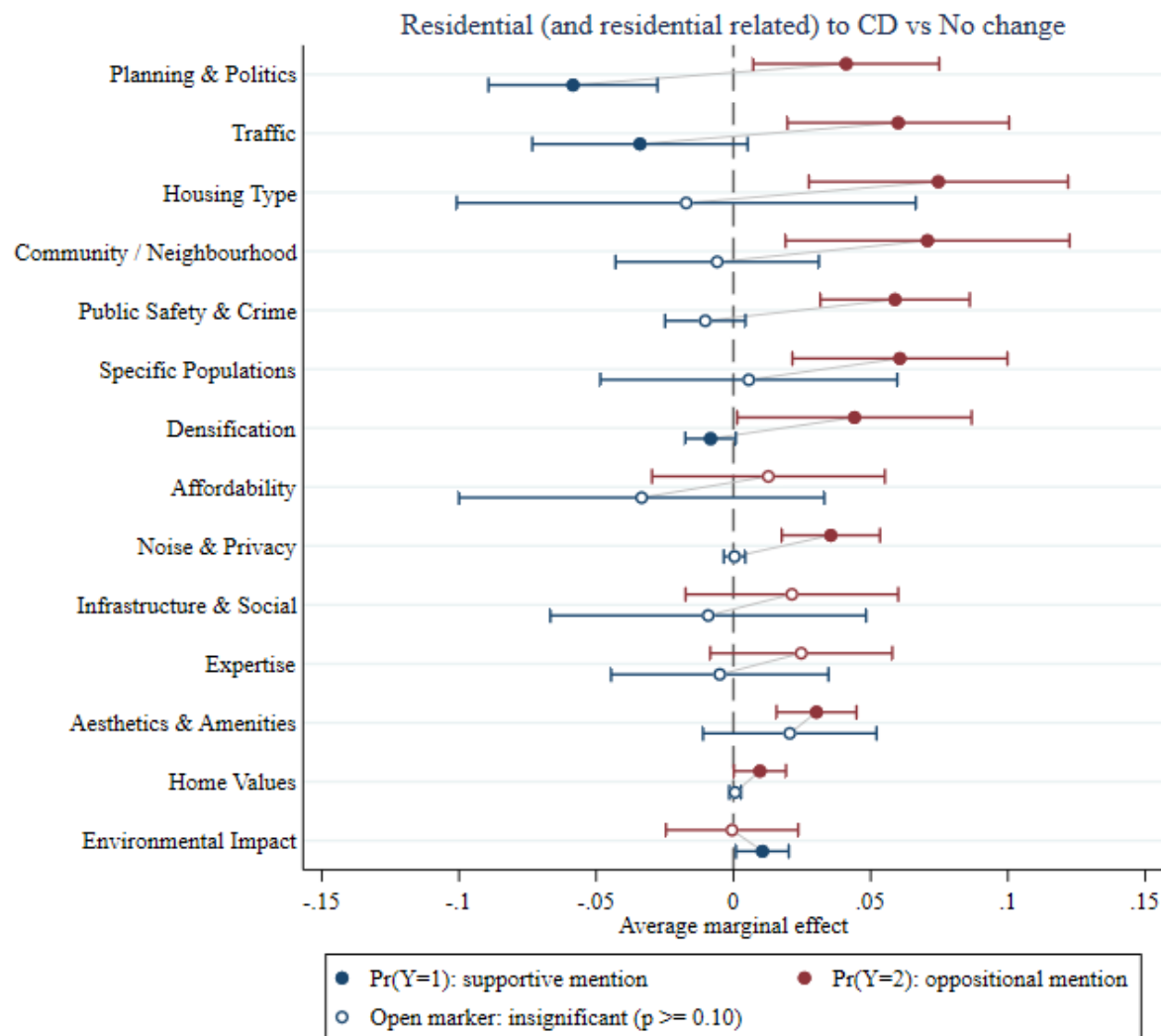


Topics estimated without neighborhood fixed effects: Environmental Impact and Noise & Privacy.

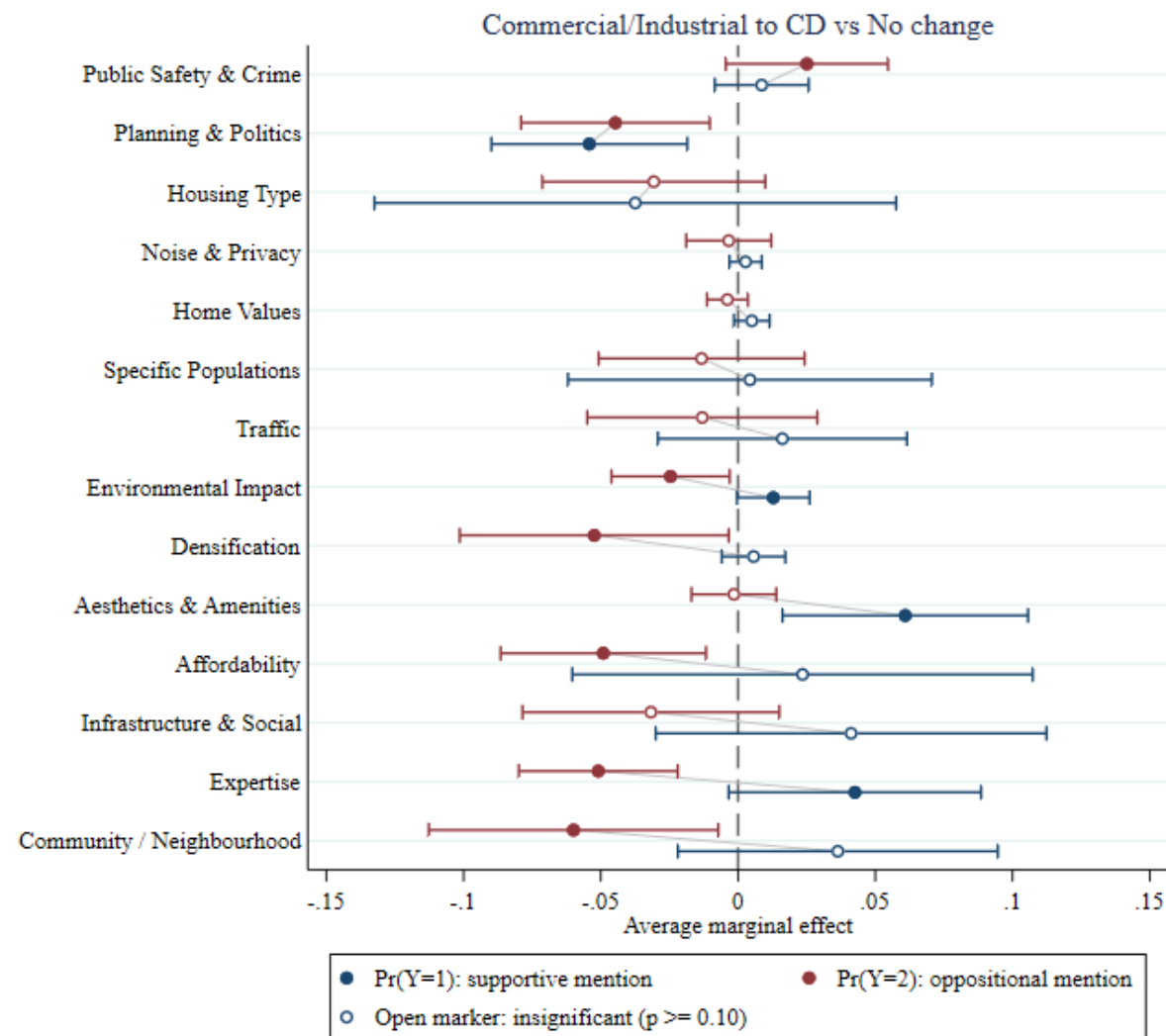


Topics estimated without neighborhood fixed effects: Environmental Impact and Noise & Privacy.

### Q3. AME on probability (Land use Change)

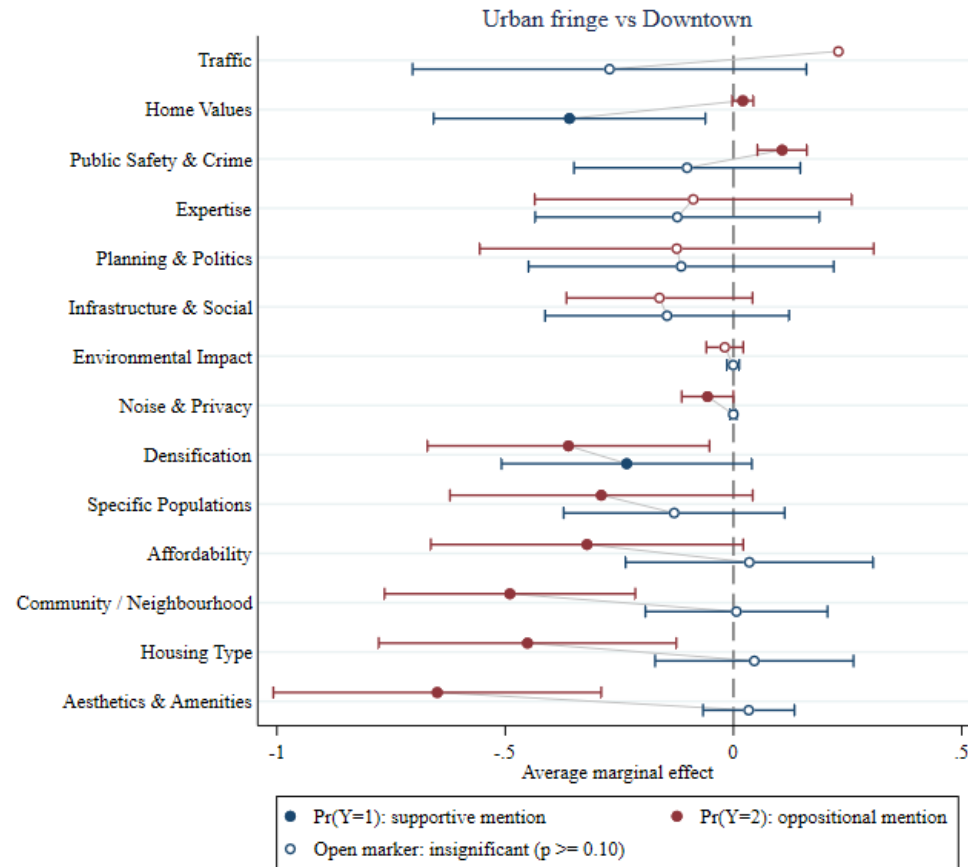


Topics estimated without neighborhood fixed effects: Environmental Impact and Noise & Privacy.

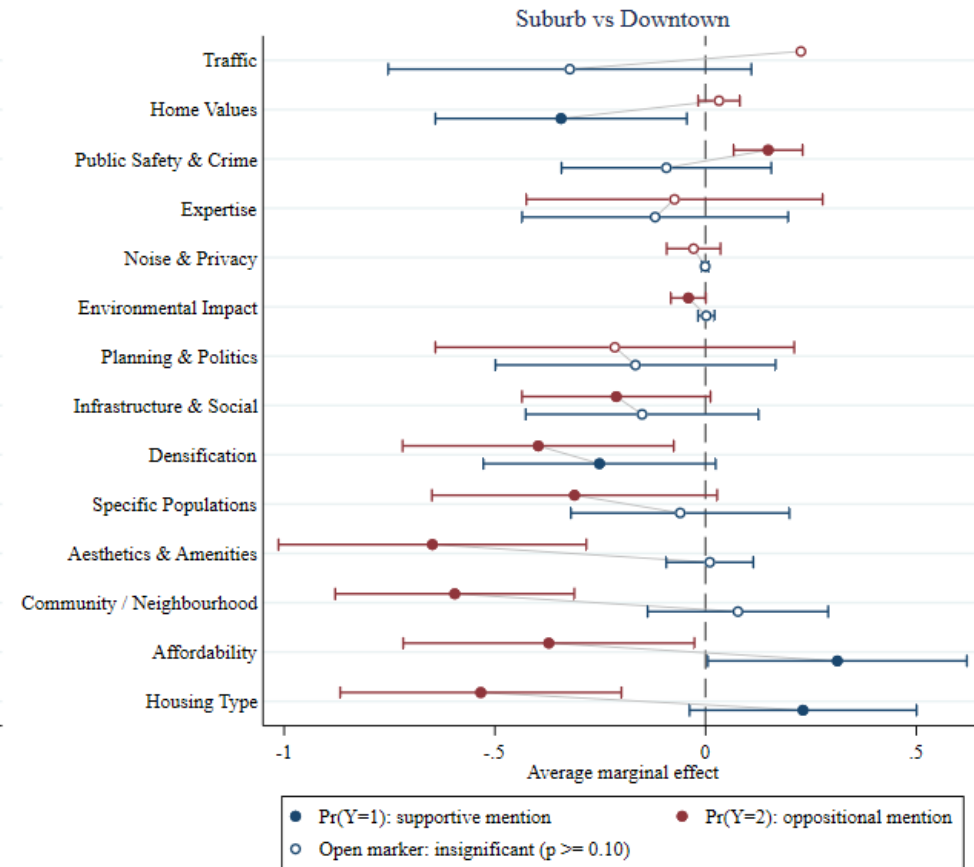


Topics estimated without neighborhood fixed effects: Environmental Impact and Noise & Privacy.

### Q3. AME on probability (location)



Topics estimated without neighborhood fixed effects: Environmental Impact and Noise & Privacy.

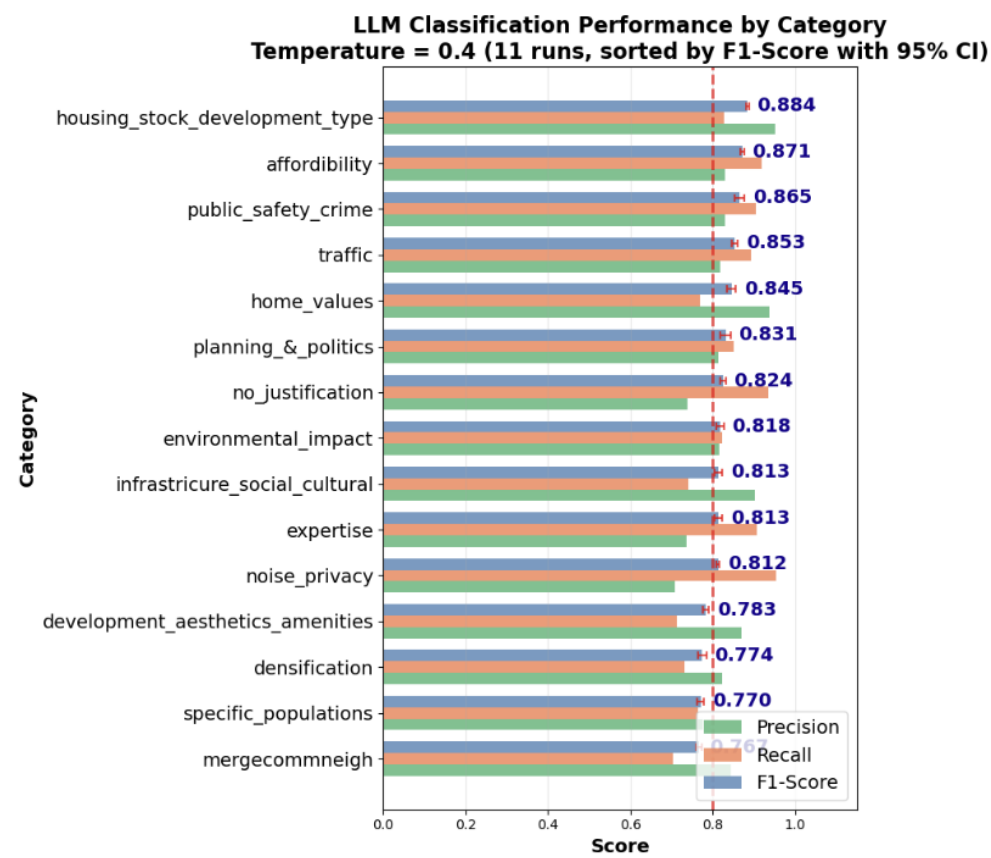


Topics estimated without neighborhood fixed effects: Environmental Impact and Noise & Privacy.

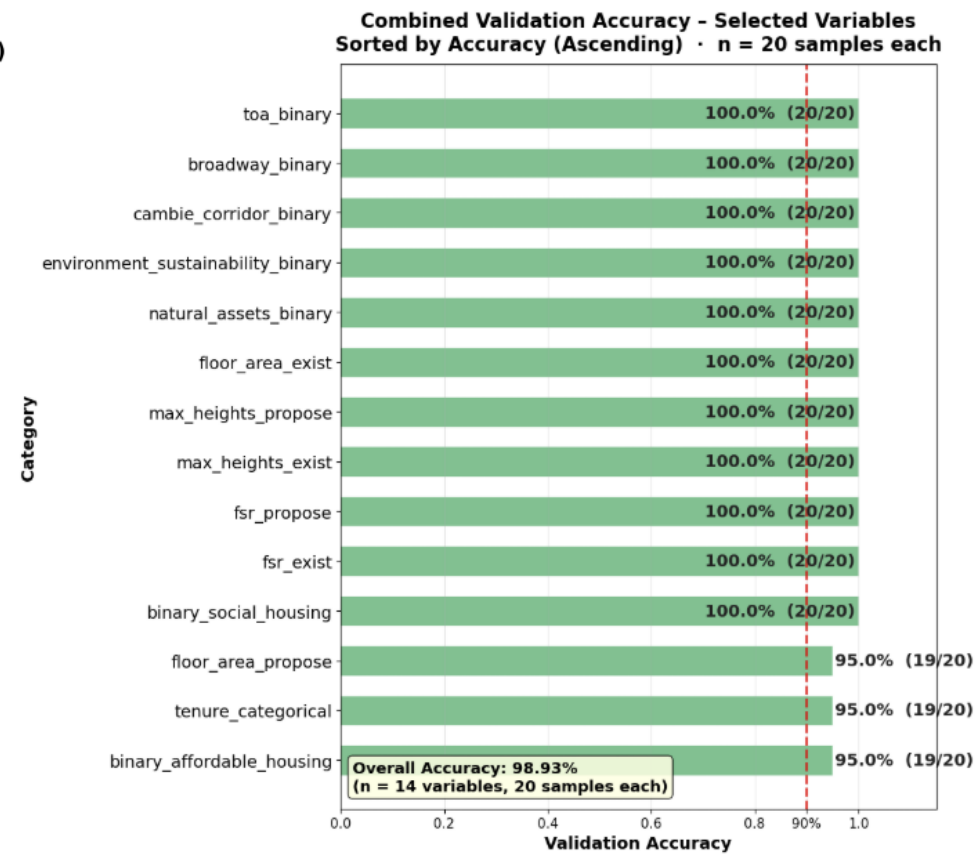
# ANOVA and post-hoc analysis

| Variable / Group                                                    | N      | Opp. Rate | Post-hoc / Test                        |
|---------------------------------------------------------------------|--------|-----------|----------------------------------------|
| binary_social_housing [a=Market rate, b=Social]                     |        |           |                                        |
| (a) Market rate                                                     | 12,353 | 42.1%     | a vs b: p=0.0000 ***                   |
| (b) Social                                                          | 4,753  | 49.0%     |                                        |
| Welch t-test                                                        |        |           | t(17,104) = -8.150, p = 0.0000 ***     |
| housing_typology [a=Low-rise/Multiplex, b=Mid-rise, c=High-rise]    |        |           |                                        |
| (a) Low-rise / Multiplex                                            | 2,767  | 36.4%     | (a) vs (b): p = 0.0000 ***             |
| (b) Mid-rise                                                        | 3,567  | 49.0%     | (b) vs (c): p = 0.0000 ***             |
| (c) High-rise                                                       | 10,714 | 44.3%     | (c) vs (a): p = 0.0000 ***             |
| ANOVA test                                                          |        |           | F(2, 17,045) = 51.030, p = 0.0000 ***  |
| urb_rural [a=Downtown, b=Urban fringe, c=Suburban]                  |        |           |                                        |
| (a) Downtown                                                        | 2,134  | 36.5%     | (a) vs (b): p = 0.0000 ***             |
| (b) Urban fringe                                                    | 13,088 | 47.6%     | (b) vs (c): p = 0.0000 ***             |
| (c) Suburban                                                        | 1,884  | 27.5%     | (c) vs (a): p = 0.0000 ***             |
| ANOVA test                                                          |        |           | F(2, 17,103) = 166.932, p = 0.0000 *** |
| is_local [a=Non-local (0), b=Local (1)] (missing excluded: n=6,227) |        |           |                                        |
| (a) Non-local (0)                                                   | 6,840  | 29.6%     | a vs b: p=0.0000 ***                   |
| (b) Local (1)                                                       | 4,039  | 66.6%     |                                        |
| Welch t-test                                                        |        |           | t(10,877) = -39.879, p = 0.0000 ***    |

# Data and Methods Workflow - LLM performance



LLM performance in **topic classification** using a temperature setting of 0.4. As robustness check, we tested temperatures ranging from 0.0 to 1.0 in increments of 0.1 and found that 0.4 yielded the best overall performance.



The performance of **variable extraction** from referral/policy reports

# Global regression

|                                        | (1)<br>OLS no FEs    | (2)<br>OLS FEs       |
|----------------------------------------|----------------------|----------------------|
| lnmax_heights                          | .0506*<br>(.0303)    | .0615*<br>(.0367)    |
| has_station                            | -.0736<br>(.0496)    | -.1123*<br>(.0589)   |
| natural_assets                         | -.1856***<br>(.0696) | -.2191***<br>(.0779) |
| Ownership                              | .1274***<br>(.0466)  | .1326**<br>(.0595)   |
| binary_social                          | .09*<br>(.051)       | .0983*<br>(.059)     |
| Land use change (benchmark: unchanged) |                      |                      |
| 1.residential related to CD            | .0796*<br>(.0438)    | .1238**<br>(.0519)   |
| 2.commercial/industrial to CD          | -.0939*<br>(.0516)   | -.0344<br>(.0552)    |
| Local comments?                        |                      |                      |
| lnwater_dist                           | -.0441*<br>(.0261)   | -.0443<br>(.0405)    |
| lnmed_home_price                       | .0871*<br>(.0491)    | .1274<br>(.0923)     |
| lnvis_minority                         | .0455<br>(.0601)     | -.0328<br>(.1175)    |
| Year FEs                               | Y                    | Y                    |
| Neighborhood FEs                       |                      | Y                    |
| cons                                   | .14<br>(1.6187)      | 1.1872<br>(2.2816)   |
| Observations                           | 335                  | 335                  |
| R-squared                              | .1919                | .2523                |

Robust standard errors are in parentheses

\*\*\* p<.01, \*\* p<.05, \* p<.1