

# Extracting and mapping wildfire policies using NLP with Plan Integration for Resilience Scorecard (PIRS)

Bernette Chan, Math; Jun Noh, Statistics; Nidhi Shinde, Statistics; Zarina Yunis, Statistics; Alex Dekhtyar, Computer Science; Kelly Bodwin, Statistics  
California Polytechnic State University, San Luis Obispo

## Background

Wildfires are growing in frequency and intensity across California, demanding stronger resilience policies. However, county plans often have conflicting and wordy policies that are time-consuming to extract.

**Goal:** Partially automate the process of extracting and mapping policies using NLP techniques and topic modeling.

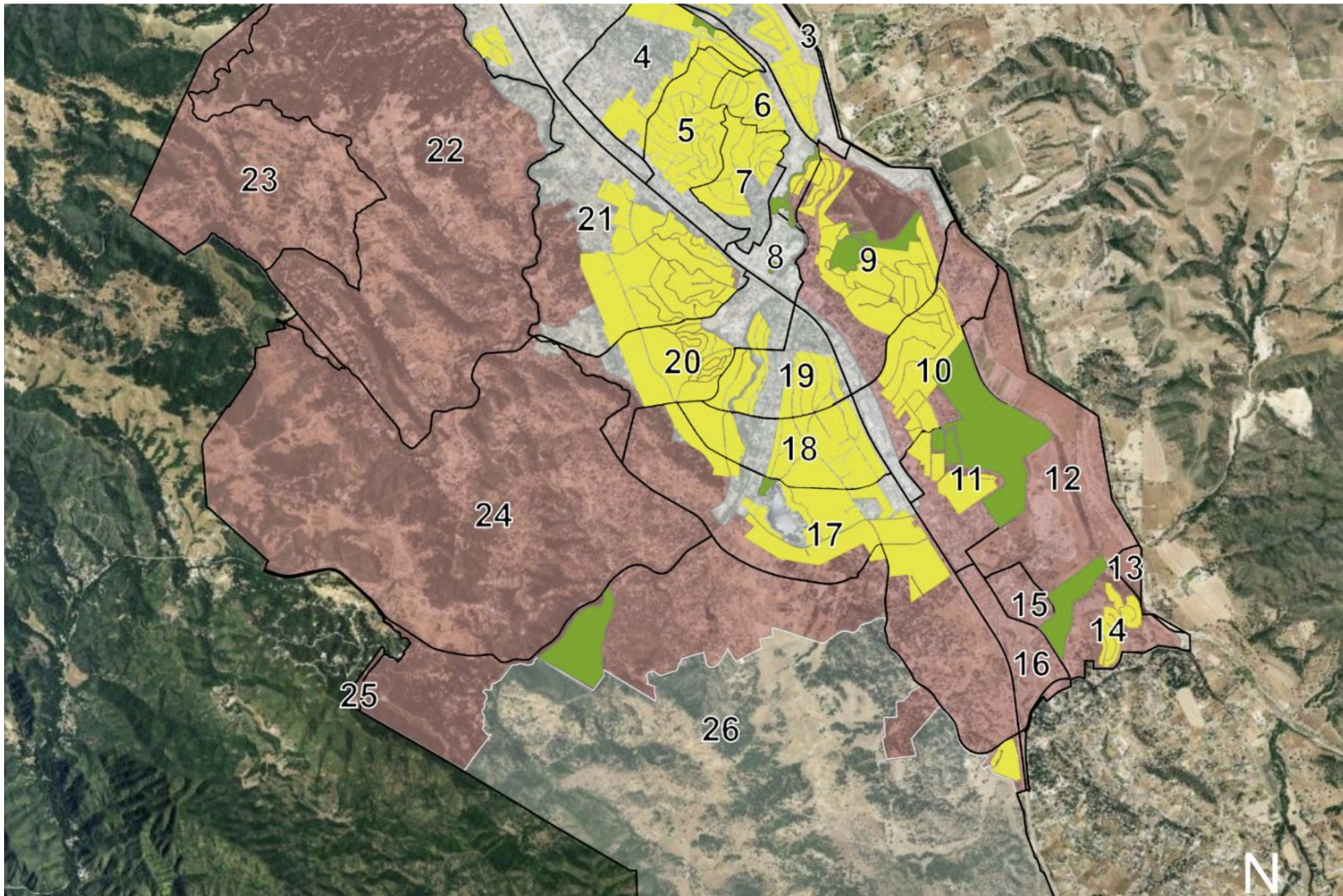


Figure 1: Atascadero Planning Districts and Wildfire Hazard Zones. Red shows Wildfire Urban Interface (WUI) zones used in the PIRS™ analysis. Overlays: green = resilience-supportive (e.g., open space); yellow = mixed (e.g., residential zones); highest hazard alignment occurs in western and southern WUI areas.

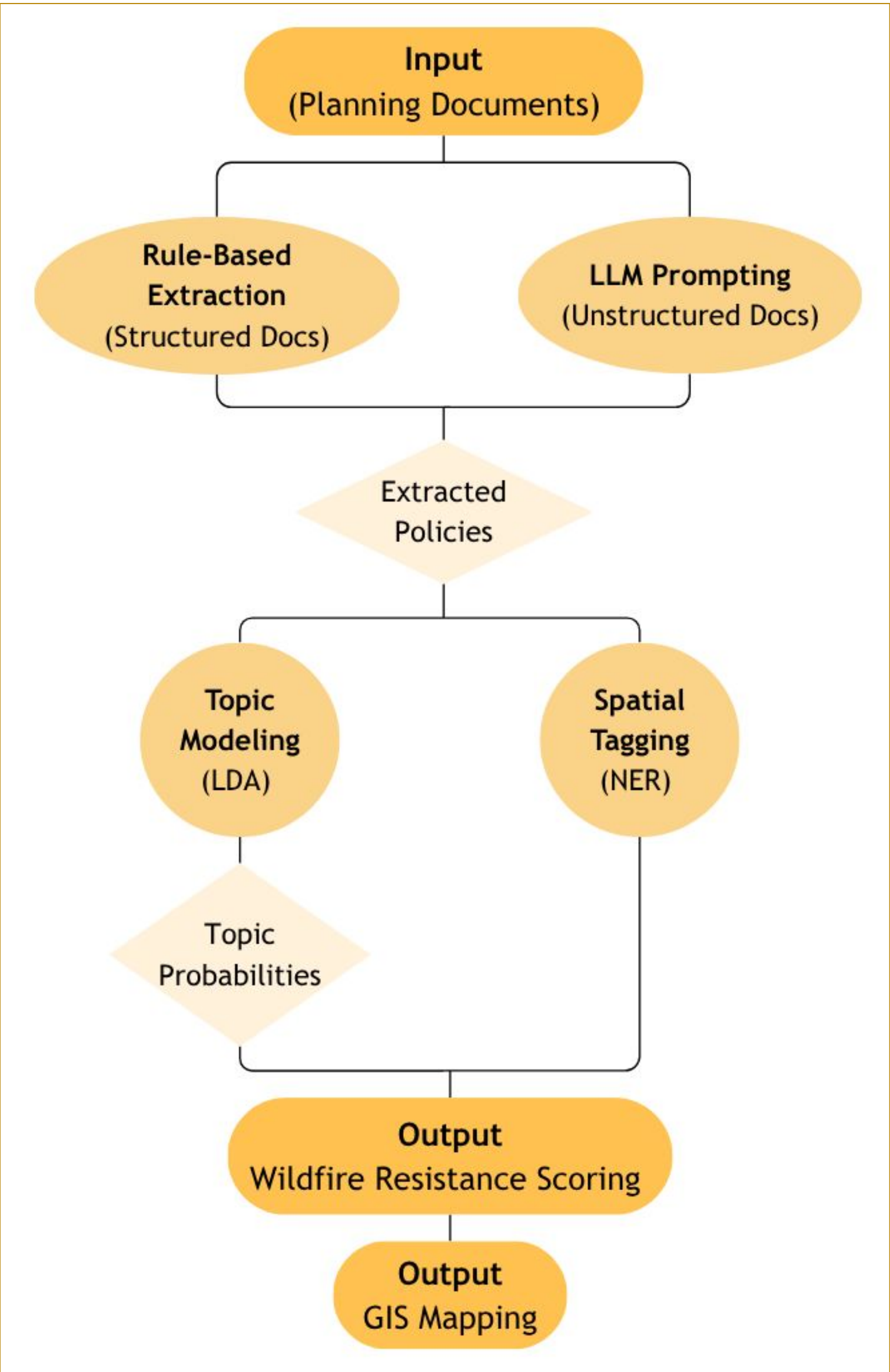


Figure 2: NLP Pipeline for Wildfire Policy Analysis  
Diverse document formats are processed to extract relevant wildfire resilience policies, infer topics via LDA modeling, and prepare outputs for integration into spatial policy maps.

## Structured & Unstructured Policy Extraction

**Structured:** We chunked each document line-by-line and then used regular expressions to identify policies. The two structured Atascadero documents emphasized words that indicate policies and action items. To accommodate different document formats, we altered the code accordingly.

**Unstructured:** For documents with long paragraphs and inconsistent headings, we used Gemini’s Large Language Model (LLM) to interpret unstructured documents like a human reader. By splitting the document into lines and querying Gemini to find actionable items, we extracted policies and manually checked them for accuracy.

Document	Topic	Probability
Policy 1.1. Preserve the rural atmosphere of the community and assure “elbow room” in areas designated for lower density development by guiding new development into the Urban Core to conform to the historic Colony land use patterns of the City and to respect the natural environment, hillside areas, and existing neighborhoods.	Noise Regulations & Land Use	0.83
[6.4.]4. Utilize the Secretary of the Interior’s Standards and Guidelines for Rehabilitating Historic Properties to assess proposed improvements to historic properties.	Housing Policy & Neighborhood Compatibility	0.78
Policy 5.1: Reduce multi-family densities and increase single-family lot sizes as site slope increases.	Environmental Conservation & Emergency Response	0.78

Table 1: Estimated topic probabilities for policies based on LDA output. Each entry indicates the likelihood that a given topic is present in a policy.

Topic	Specificity
Environmental Conservation & Emergency Response	0.93
Residential & Mixed Use Development	0.92
Open Space & Recreation Planning	0.92
Historic Preservation & Community Identity	0.92
Noise Regulations & Land Use	0.91
Zoning & Aesthetic Standards	0.90

Table 2  
Top six topics with the highest specificity scores, indicating how uniquely each topic is associated with the Atascadero General Plan. See definition of specificity in the LDA section on the left.

## Latent Dirichlet Allocation (LDA) Model

LDA generates meaningful topics within a document based on a distribution of words. We selected 16 topics for the Atascadero documents to ensure each policy has meaningful topic associations.

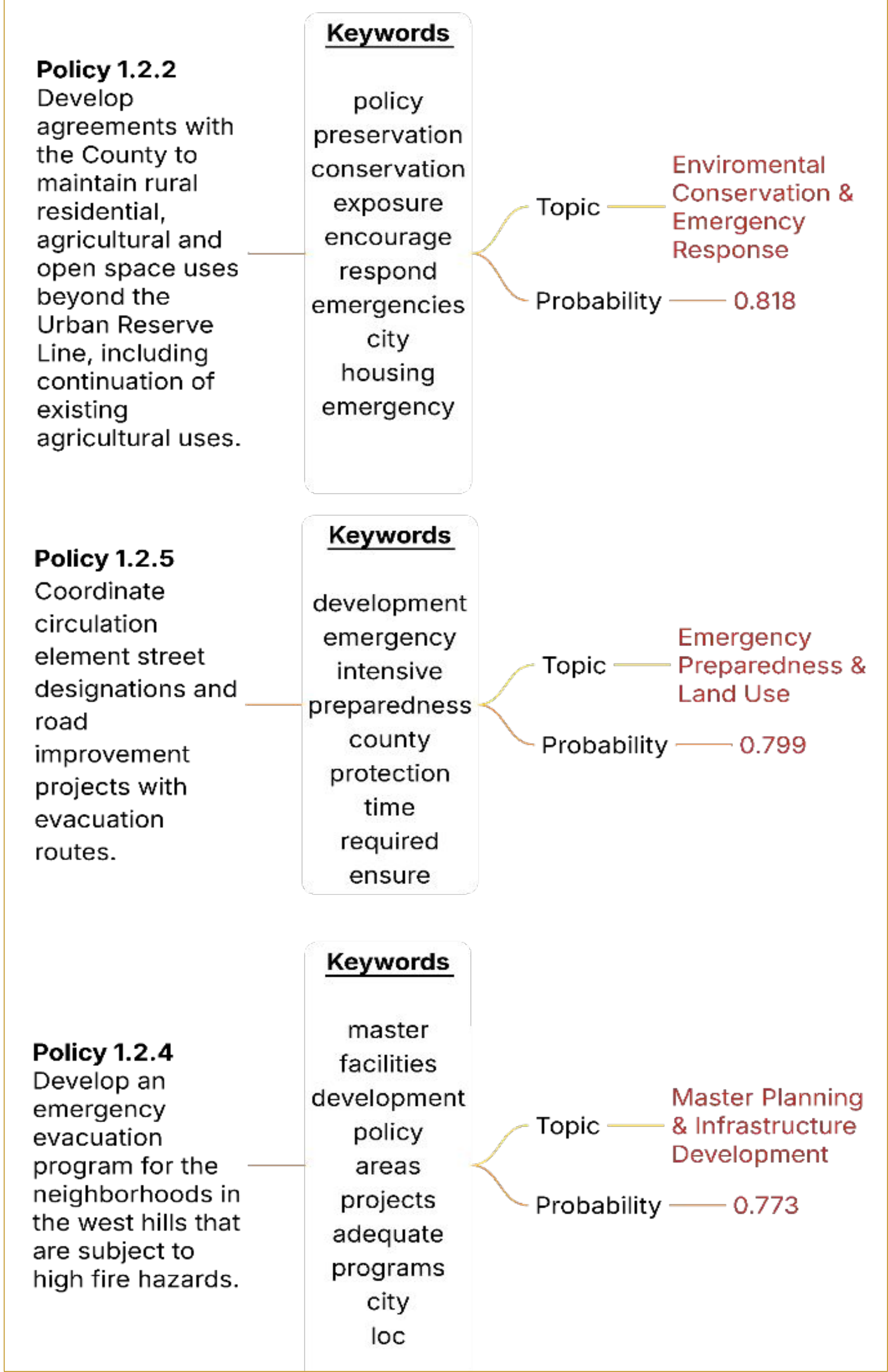
Process:

- Topic probabilities are assigned based on phrase prevalence
- Each document is assigned to its most representative topic
- Topics are manually labeled for their relevance to wildfire mitigation
- Topics are compared to client-identified PIRS policies (ground truth) for model evaluation

Evaluation Metric:

- Specificity measures how well a topic avoids including irrelevant (non-PIRS policies)
- Higher scores → fewer policies are incorrectly grouped into topics

## LDA Results



## Future Work

- Use GIS to map scored policies spatially, helping counties prioritize high fire-risk areas.
- Expand policy extraction across more documents.
- Refine policy categorization for improved scoring.
- Evaluate topic model performance across counties to identify improvements.

## References

[1] Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan), 993-1022.  
[2] City of Atascadero. Atascadero General Plan 2025 [PDF]. Atascadero Planning Department, 2025. <https://www.atascadero.org/sites/default/files/2023-06/Atascadero%20GP%202025.pdf>  
[3] San Luis Obispo County. San Luis Obispo County Multi-Jurisdictional Hazard Mitigation Plan. 2019. <https://www.slocounty.ca.gov/Departments/Planning-Building/Forms-Documents/Plans-and-Elements/Elements/Local-Hazard-Mitigation-Plan/San-Luis-Obispo-County-Multi-Jurisdictional-Hazard.pdf>  
[4] City of Atascadero. Community Wildfire Protection Plan. 2019. Unpublished manuscript, available via personal archive.  
[5] City of Atascadero. Final Climate Action Plan. 2014. Unpublished manuscript, available via personal archive.

## Acknowledgements

We sincerely thank Bill Siembieda, Margot McDonald, and Andrew Fricker for their guidance and support throughout this project.