Hidden in Plain Text: LLM-Assisted Detection of Discriminatory Local Laws

Dan Bateyko*
Cornell University
Ithaca, NY, USA
drb348@cornell.edu

Yasmine Mabene*
Stanford University
Stanford, CA, USA
ymabene@law.stanford.edu

Derek Ouyang
Stanford University
Stanford, CA, USA
douyang1@law.stanford.edu

Daniel Ho
Stanford University
Stanford, USA
deho@law.stanford.edu

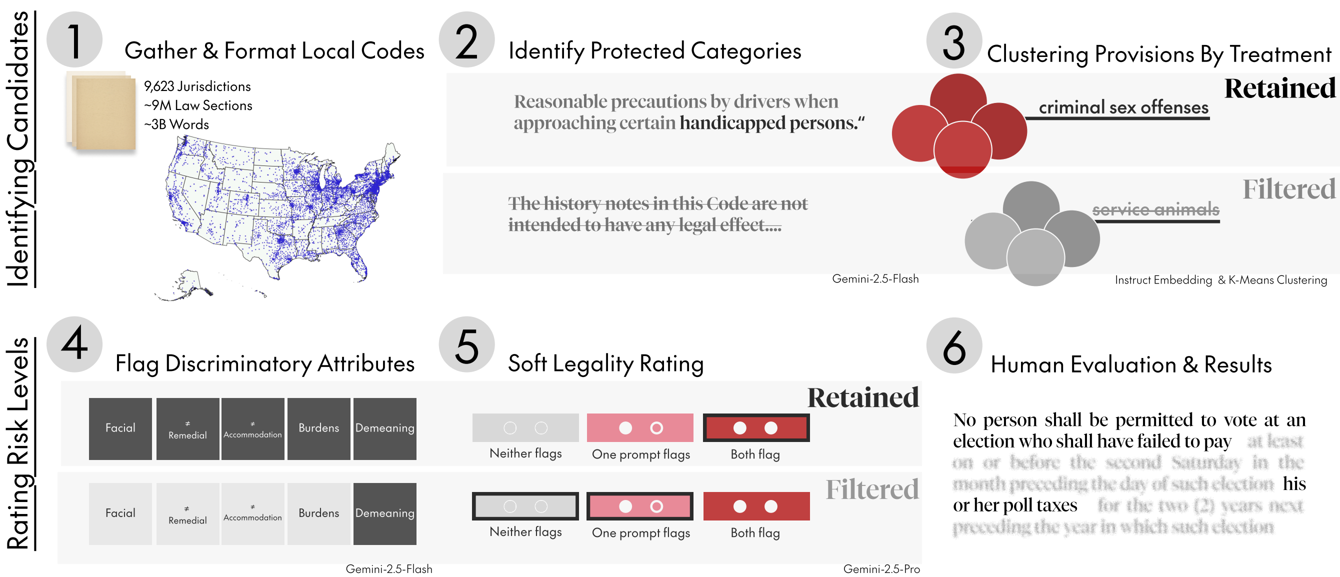

Figure 1: Overview of our LLM-assisted pipeline for detecting potentially discriminatory laws in ~9M law sections across 9,623 local jurisdictions. We identify and cluster laws mentioning differential treatment of protected categories, flag key attributes—whether laws are facially discriminatory, lack a remedial purpose or accommodation, impose unjustified burdens, or use demeaning language—and rate nearly 10,000 as high priority, *e.g.*, poll taxes.

Abstract

Legal reformers, for repeal campaigns and litigation, identify discriminatory laws through painstaking manual searches. This approach has proven effective at the federal and state levels, but becomes infeasible for local law owing to its scale. We make three contributions. First, we assemble the largest corpus of U.S. local codes and charters, spanning 9,623 jurisdictions and covering approximately 75% of the U.S. population. Second, we introduce a large language model (LLM)-assisted pipeline that combines *(i)* high-recall candidate identification with *(ii)* soft legality ratings, prioritizing provisions for human review. We detect laws that explicitly reference protected categories and impose group-specific treatment, and apply few-shot, chain-of-thought prompting to assign priority ratings for expert review. We find that this framework *(1)* achieves high recall (0.91) and precision (0.91) on a held-out, curated validation set of historical and expert-identified laws, but *(2)* is less reliable in drawing substantive assessments of antidiscrimination law compared to human reviewers. Third, we show

*Both authors contributed equally to this research. Access to our results can be found at: https://hidden-in-plain-text.reglabapp.com.

ICAIL '26, Singapore, Singapore

ACM ISBN 978-1-4503-XXXX-X/2018/06
https://doi.org/XXXXXXX.XXXXXXX

that this system uncovers a wide range of discriminatory provisions, including poll taxes, race and gender-based voting rules, gendered occupational restrictions, and citizenship-based licensing exclusions. Together, these results demonstrate the potential of LLM-assisted workflows to scale review and reform of local law.

CCS Concepts

• **Applied computing** → **Law**; • **Computing methodologies** → *Natural language processing*.

Keywords

local laws, Large Language Models, Discrimination, Human-LLM-Agreement, Chain-of-Thought, LLM-Assisted Legal Review

ACM Reference Format:

Dan Bateyko, Yasmine Mabene, Derek Ouyang, and Daniel Ho. 2026. Hidden in Plain Text: LLM-Assisted Detection of Discriminatory Local Laws. In *Proceedings of The 21st International Conference on Artificial Intelligence and Law (ICAIL '26).* ACM, New York, NY, USA, 10 pages. https://doi.org/XXXXXXX.XXXXXXX

Warning: This paper includes offensive, derogatory, and outdated language from local laws.

1 Introduction

Why do facially discriminatory laws persist in local codes for decades? In 2010, a Florida resident encountered a racial slur embedded in Palm Beach County's online code, where it had remained since 1959 [66]. In Nevada, a 1917 "sundown" ordinance barring Indigenous people from town after 6 p.m. stayed in place for more than fifty years [60]. Across local jurisdictions, laws that from their plain meaning appear discriminatory can sit untouched. Sometimes the explanation why is mundane: no one looked.

When researchers *do* search for discriminatory laws, typical keyword searches and manual review are time-consuming and produce uneven results. In the 1950s, Pauli Murray compiled over 700 pages of segregation laws—a volume Justice Thurgood Marshall called the "bible" of *Brown v. Board*—but only through demanding review constrained by the limits of printed legal indexes [54, 62]. In the 1970s, Ruth Bader Ginsburg co-led the U.S. Commission on Civil Rights' effort to excise sex bias from federal law, but relied on a federal database of only 59 gender-related keywords [74]. In the 2000s, Professor Gabriel Chin cataloged vestigial Jim Crow laws, prompting removals in several states, while acknowledging that his team missed unusually phrased laws [16]. As recently as 2019, a Virginia commission tasked with identifying sixty years' worth of facially discriminatory laws succeeded largely by brute force, mobilizing law students from three universities, staff from the governor's office, practicing attorneys, and a sitting judge [2].

The scale of local law makes surveys rare; one scholar describes comprehensive review across thousands of cities as "practically impossible" [23]. Yet local laws govern consequential matters: who can vote, who can work particular jobs, who can attend which schools, and who can be detained or punished. Systematic review of these codes would offer a road-map for litigation and reform. Litigators could identify jurisdictions where suspect classifications exist and where residents may not realize that a restriction is legally suspect. Even when a law has no attached legal consequence, residents may wish to see laws expressing outdated or discriminatory sentiments removed.

We assess how large language models (LLMs) can scale this review process. We propose a pipeline designed to focus expert attention on provisions where discriminatory intent is evident from the text itself. We focus on overtly discriminatory language that reviewers can quickly assess, excluding facially neutral laws with discriminatory effects. We apply this pipeline to local codes in two stages: identifying candidates, which removes irrelevant legal sections, then rating risk levels, an interpretive screen which prioritizes candidates for review. In the first stage, we arrive at a set of sections that reference protected categories and impose differential treatment using conventional natural language processing (NLP) methods, few-shot prompting, and embedding-based clustering. Dictionary methods and topic modeling further help flag explicit slurs and group sections by subject matter. In the second stage, we use LLM-generated reasoning to rate sections for priority review.

We make three contributions:

- *Corpus.* We assemble a novel corpus of local codes and charters spanning 9,623 jurisdictions and covering an estimated 75% of the U.S. population, roughly three times the next largest reported collection.
- *Method.* We introduce an LLM-assisted pipeline to prioritize local laws likely to be discriminatory or offensive and demonstrate that the pipeline can reduce approximately nine million sections to 9,980 high-priority candidates, a 99.9% reduction making legal review tractable. For the highest-priority stage, when evaluated against human raters, the pipeline achieves 0.89 recall and 0.47 precision, with fair inter-rater agreement (Cohen's $\kappa = 0.31$). These results are consistent with our design choice to prioritize recall over precision at the final stage. We characterize model–human disagreement patterns (legal ambiguity, historical terminology, deference to governmental interests) to inform future refinements.
- *Substantive findings.* We document discriminatory provisions across areas of law, including Jim Crow and pre-suffrage holdovers, such as segregated schools, literacy tests, poll taxes, and male-only voter eligibility. We further identify potentially suspect and more contemporary citizenship-based occupational licensing restrictions across over 1,600 jurisdictions which govern 49 million residents. We inventory offensive language, including ethnic slurs, anachronistic terminology, and references to bygone practices such as minstrel and "freak" shows.

These findings provide a starting place for further legal research and reform across U.S. local jurisdictions.

2 Background

This study builds on two strands of scholarship. First, computational legal scholarship increasingly supports statutory analysis but faces challenges on interpretive legal questions. Second, legal scholarship on local law identifies structural reasons why discriminatory provisions arise and stay on the books, suggesting the problem may be widespread. Together, these point to an opportunity for large-scale, LLM-assisted review.

2.1 Computational Methods for Statutory Interpretation

Computational methods offer tools for narrowing the set of laws requiring expert review. Traditional NLP approaches like lexical dictionaries and embedding-based clustering can flag terminology or group provisions by subject matter [25, 30]. More recent work develops and evaluates LLMs for semantic processing and reasoning across qualitative and legal settings [63, 64, 82]. Applied to statutes specifically, LLM-assisted approaches have shown promise in legal retrieval and reasoning tasks, including finding state and local laws [71], identifying racially restrictive covenants in property records [4], analyzing state unemployment schemes [35], classifying municipal zoning regulations [10], and general statutory reasoning tasks [11, 22, 32, 39].

Yet fewer studies focus on LLMs' capacity for interpretive analyses like assessing a provision's legality. Here, generative interpretation poses distinct challenges. LLM prompting begins by translating a concept into concrete criteria the model can apply, but legal concepts are often "ambiguous or contested by experts" [45]. This labeling task exhibits what Guerdan *et al.* call *rating indeterminacy*: even with identical instructions, reasonable annotators may disagree with one another, or even with themselves over time, about whether a provision counts towards a contested definition [31]. How well classifiers match human judgments, therefore, depends on how precisely the task is defined [34, 61, 72]. Developing that specification is itself a costly research problem [33], and in legal settings, even when annotators review many examples together, they may continue to disagree [72]. Conversely, under-specified tasks, such as zero-shot prompting, defer discretion to the model, risking errors that are hard to detect downstream [34]. Technical choices, including model selection and prompt design, further pose engineering problems that can meaningfully affect generative statutory interpretation outcomes [18, 29, 37, 77].

For these reasons, scholars have proposed generative interpretation as an assistive tool rather than a replacement for legal judgment [29, 59]. Coan and Surden suggest that constitutional analysis in particular benefits from "focused queries" where attorneys reason alongside model outputs on narrower questions [18]. When researchers engage with model outputs in the style of grounded theory or qualitative content analysis, LLM-assisted interpretation offers additional benefits. Iterative review of topics and labels can help refine the research task [55, 64], and computational approaches can surface legally relevant concepts that reviewers might otherwise miss but recognize as relevant once flagged [80]. LLMs' broad parametric knowledge and multi-step reasoning can support legality assessment, reaching beyond the four corners of a provision to consider relevant real-world context. Our approach builds on these insights. We develop an LLM-assisted pipeline through iterative prompt design, with authors reviewing outputs over the course of project conceptualization and providing few-shot examples in prompts.

2.2 Why Do Discriminatory Local Laws Persist?

Local lawmakers operate with limited budgets, little or no dedicated legal staff, and minimal outside oversight [23, 24, 81]. Most Americans are governed by part-time local lawmakers[24, 81]. Local legislatures can be unrepresentative of the populations they govern, increasing the risk of "noninclusive legislation" that excludes political or demographic minorities [24]. Yet local lawmakers are often on the front line of hot-button social conflicts and may translate political disputes into politically-charged rules. These conditions create multiple pathways for discrimination to enter and remain in the code: some jurisdictions discriminate deliberately, while others retain discriminatory language inadvertently.

Over time, such enactments can sediment into a kind of policy sludge. Legal scholars describe these provisions as "dormant," "dead," or "zombie" laws, and even when unenforced, they pose risks [9, 12, 26, 28, 41, 44, 65, 70, 79]. They can misfire when precedent shifts, muddy the legal landscape by creating ambiguity about legal obligations, and communicate outdated moral judgments through law's expressive function.

Local jurisdictions do occasionally review their codes, with such efforts often taking one of two forms: wholesale recodification oriented toward modernizing the code, or targeted reforms aimed at discrete issues, like unnecessarily gendered language [27, 42, 50]. These efforts reveal how rarely codes get sustained attention; as Contra Costa County discovered in its review of gender-specific pronouns, the code had not been changed in parts since 1970 [46].

Absent code revision, cities improvise. Some pass blanket ordinances that sweep changes across the code without revising individual sections. In Norwalk, Connecticut, for example, an official explained that the city pursued a "universal change" by revising a definition because it was more cost-effective than review [14]. Other cities turn to residents: Portland runs a hot-line for reporting discriminatory language in local ordinances [5]. Such inventive responses offer only partial fixes and place the onus on ordinary citizens to research legal provisions.

That jurisdictions infrequently review their codes reflects, among other things, the basic difficulty of searching thousands of legal sections. This difficulty helps explain the growing role of third-party code publishers, who offer services to flag obsolete provisions [3, 6, 17, 75]. As more jurisdictions rely on these vendors to host and maintain their codes, these providers become key access points for local law.

3 Data

No central clearinghouse exists for local codes. Though some jurisdictions still require an in-person visit to city hall to consult the official code [1], fortunately, many others publish their codes online, either on their own websites or through third-party codification providers. Scholars who have reviewed these codes to describe local regulations [51, 56], theorize about their origins and effects [36, 47, 52, 53], and analyze them as legal text [8, 21] have, with good reason, generally focused on a single state, policy area, or small set of cities. To study legal complexity, Ash, Goessmann, and Naidu gathered LexisNexis' complete collection, some 3,259 legal codes [8]. Because discriminatory provisions make up only a small

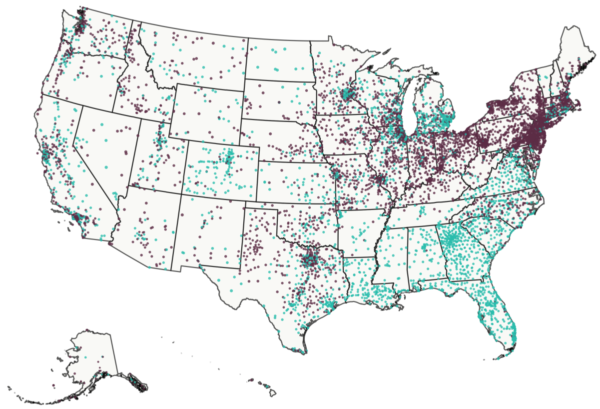

Figure 2: Geographic distribution of jurisdictions by code providers in our corpus. Provider coverage varies regionally, with more CivicPlus (teal) use in the South and more ICC (purple) in the Northeast. The number of mapped jurisdictions is fewer than in our corpus (see Appendix A).

fraction of local codes, we assemble a dataset nearly three times larger to ensure sufficient coverage.

We collect local laws from two major providers, the International Code Council (ICC) and CivicPlus, which publish codes for thousands of jurisdictions through various online portals. We scrape and parse codes, yielding a dataset of local codes and charters for 9,623 jurisdictions. These jurisdictions account for about 25% of all "local governments," a category that includes any "city, county, town, township, village, or other general-purpose political subdivision," which typically possess their own legislative authority [73, 81]. Although 25% is a minority of jurisdictions by count, coverage is concentrated in larger and more populous places. We estimate that the jurisdictions in our dataset govern approximately 75% of the U.S. population. We report detailed coverage statistics in Appendix A.

For each jurisdiction, we parse the code into their sections, often identified by a code's section headers (*e.g.*, § 50-31). Though we collected these codes in late 2025, not all codes are up-to-date; for example, some jurisdictions host PDFs which lag behind their corresponding web versions.

4 Methods

We develop a multi-step, LLM chain-of-reasoning prompting approach to identify, cluster, and rate provisions that are likely to be discriminatory or offensive. We use Google's family of models, Gemini, based on relative cost and other considerations (see Appendix Table 7).

4.1 Identifying References to Protected Categories and Differential Treatment

We prompt a smaller LLM, Gemini-2.5-Flash, to label all provisions that reference one of eight federally-protected categories, or proxies for such: race, religion, sex/gender/sexual orientation, marital status, national origin/alienage/ancestry, genetic information, age (40 and over), and disability or medical condition [19]. We also prompt to determine whether each provision involves differential treatment, which we define as the assignment of distinct rules, requirements, accommodations, or definitions to specified groups. We further instruct the model to produce explicit reasoning to justify each classification (see Appendix B.1). Because labeling our full dataset is costly, we first label a 40% sample of sections and train a TF–IDF logistic-regression classifier to screen the remaining sections for protected-category references. By setting the classification threshold to achieve 99.5% recall on a held-out set, we filter 17% of the corpus before labeling the remainder (see Appendix C).

4.2 Clustering Provisions by Treatment

We then use Instructor [68], an instruction-tuned embedding model, on the text of the LLM-generated reasoning to cluster provisions identified as involving differential treatment. We prompt the model to create embeddings for provisions according to two dimensions: the protected categories referenced and the type of treatment. We embed the reasoning rather than the statutory text because individual laws can address multiple topics; reasoning text better isolates the treatment of protected categories (see Appendix E). We then apply k-means clustering to the resulting embeddings. We test different values of k, balancing the need for clusters that capture distinct kinds of law and cohere together (as measured by high mean cosine similarity) against the need for clusters large enough that removing them shrinks the set of laws needing review. From each cluster, we sample 20 laws and use Gemini-2.5-Flash to generate a descriptive label summarizing the form of treatment represented within the cluster.

Two authors then manually review the resulting clusters, evaluating cluster labels and the statutory text of selected provisions across all protected categories. We exclude clusters whose underlying legal content does not align with our analytical criteria (see Appendix E), including clusters that reflect legislative accommodations or remedial efforts (see Appendix Table 4). We retain clusters recommended by at least one author. We document our prompt, review criteria, and examples of removed clusters in Appendices B and E.

4.3 Rating Discriminatory Risk Levels

Antidiscrimination law is contested territory. Different theories—anticlassification versus antisubordination—and different legal approaches can support distinct arguments about whether a provision passes constitutional muster. We therefore generate a diverse set of signals and aggregate them into a priority rating fitting our goal for review.

4.3.1 Discriminatory Attributes. In this step, we classify whether provisions exhibit features indicative of discriminatory treatment in the absence of a recognized justification.

We instruct the LLM to assess whether each provision exhibits any of the following features: *(1)* explicit facial discrimination, defined as the express identification of a protected or otherwise identifiable group coupled with a differential treatment such as a ban, exclusion, or denial of fundamental rights; *(2)* absence of a remedial or sovereignty-based justification; *(3)* lack of accommodation or affirmative action intent; *(4)* unjustified classification, whereby a provision burdens, excludes, or privileges a group in a manner not necessary to the legitimate regulatory function at issue; and *(5)* demeaning or historically oppressive framing, including language or structures that mirror historically exclusionary legal regimes. We instruct the LLM to indicate whether each criterion is present and to provide supporting reasoning for its determinations (see Appendix B.4).

4.3.2 Soft Legality Rating. Next, we use a more capable LLM, Gemini-2.5-Pro, to rate provisions identified as having unjustified classifications from the previous step. We prompt to assess whether a provision *satisfies*, *violates*, or presents an *ambiguous* case under the Equal Protection Clause or federal anti-discrimination law, and to provide supporting sources, including relevant Supreme Court cases.

We set up two prompts to capture uncertainty in legal interpretation. One prompt steers the model to prefer the *ambiguous* label when legal precedent is weak or indeterminate. The second prompt asks the model to more readily distinguish between "violates" and "satisfies" determinations. We provide different few-shot examples to each model in order to steer their judgments towards our own assessments (see Appendices B.5 and B.6). We combine assessments between the two prompting strategies to produce an overall priority rating of provisions (see Appendix Table 2). We caution that LLM-assisted statutory interpretation is an area of active research; accordingly, we use the rating with due skepticism.

4.4 Offensive Language

To identify provisions containing offensive language, we use two complementary strategies: searching for laws with potentially offensive words and labeling laws directly. For offensive terms to search, we construct a candidate list by gathering all unigrams and a sample of bigrams from the corpus, and prompt an LLM to label the words which are potentially offensive. We further supplement from public lists of offensive terms. We search all potentially offensive terms, then prompt an LLM to label the matching laws as offensive or not. Second, to cover offensive language not captured by the one or two-word phrases in the first approach, we gather a sample of provisions which an LLM labels directly. Further details are available in Appendix B.7, B.8, and G.

4.5 Evaluation

We evaluate two distinct tasks: *(1)* whether the model correctly detects provisions that explicitly reference a protected category and encode group-specific treatment and *(2)* whether our pipeline's priority ratings align with human judgments about what warrants legal review.

4.5.1 Validating Candidate Identification Using External Sources. We build out our validation set using three external lists and our own purposive sampling of state and local law. The first list consists of laws that discriminate against persons with disabilities. Over 100 legal historians and scholars assembled this list in their *amici curiae* brief for *Board of Trustees of the University of Alabama v. Garrett* (2001). The scholars catalog 398 examples of state statutes, session laws, and constitutional provisions from the late nineteenth century through 2000 which they argue are discriminatory [40]. The second list contains state-level same-sex marriage restrictions from two sources: *(i)* a Wikipedia compilation and *(ii)* a Congressional Research Service aid [20]. The third list is an expert-labeled training set of Jim Crow laws in North Carolina session laws between 1866 and 1967, assembled by University of North Carolina (UNC) researchers [38]. The UNC team, as part of a machine learning project to identify Jim Crow laws, constructed the dataset from foundational legal bibliographies, including Pauli Murray's, and a random sample of laws reviewed by hand.

To supplement these lists, we purposively sampled additional laws using dictionary searches, embedding-based searches, and web searches, then drew our final validation set from the combined pool, yielding 259 laws. These laws consist of positive and negative examples, span 45 states (all states excluding Illinois, Iowa, Kentucky, Rhode Island, and Virginia) and Puerto Rico, and include laws with differential treatment based on socioeconomic status. Because sustained legal attention at the local level is limited, the set consists mostly of state statutes and does not overlap with our local code corpus. We use the same prompting, instructions, and output format on this set as in the main pipeline. We discuss the tradeoffs of validating a local-code pipeline against state statutes and provide the distribution of protected categories within the validation set in Appendix D.

Using this set, we report precision and recall for correctly identifying laws which differentially treat protected categories. Because a single legal section can reference multiple protected categories, a section may contribute both true positives and false positives. The model flags sections correctly when the model identifies at least one correct category-treatment pair in a section (see Appendix D).

4.5.2 Validating Risk Ratings Using Pipeline Output. For the risk rating task, we perform two separate validations for robustness. First, we take our previously assembled validation set of external sources, draw a 20% stratified sample ensuring representation across all protected categories, and produce human and LLM labels following Section 4.3. For the discriminatory attribute detection, we report the proportion of agreement for each attribute. For the soft legality rating, we compute the proportion of laws for which human annotators agree with the LLM output along three dimensions: *(1)* whether the cited legal reference exists and is relevant to the provision; *(2)* whether the model accurately interprets the referenced case law or legal holding; and *(3)* whether annotators agree with the model's assigned priority rating. Second, we construct an additional validation set by sampling 306 provisions rated by our pipeline. We stratify our sample across risk ratings (HIGH-/MED/LOW) and by legal topic. Our pipeline functions as triage for human review, so we evaluate its alignment with annotator judgment. We measure inter-rater agreement on priority labels (Cohen's κ) and characterize Type I and Type II errors.

5 Findings

The pipeline reduces a corpus of over nine million provisions to a tractable set for expert review. Of those provisions, it classifies 933,047 (9.8%) as referencing protected categories and 455,673 (4.8%) as imposing differential treatment based on those categories. Within this subset, the pipeline flags nearly 10,000 provisions as high priority. Below, we describe the discriminatory laws and offensive language these high-priority flags surface, then turn to validation.

5.1 High Priority Discriminatory Laws

Our approach rates as high priority nearly 10,000 laws, spanning myriad domains and all federally-protected categories, as illustrated in Table 1 and Appendix Table 6. We highlight striking examples below.

First, 70 years after *Brown v. Board of Education* (1954), we still find provisions mandating racial segregation. Memphis, Tennessee's Charter provides for *"separate schools for the use and accommodation of the white and colored youths of the City..." (Charter § 1000)* While shocking, this finding is consistent with findings of state laws by Gabriel Chin [16].

Second, we find numerous provisions imposing race-based voting restrictions. For example, poll taxes have long been held unconstitutional, with *Harper v. Virginia State Board of Elections* (1966) finding that wealth-based discrimination for voting violates the Equal Protection Clause. However, Valparaiso, Florida, provides: *"No person shall be permitted to vote [without] pay[ing] … poll taxes." (Charter § 4)*. A similar race-based restriction remains in Carrollton, Georgia, which requires that *"[The clerk] shall keep...separate lists of white and colored voters, who are entitled to vote in said city… " (Charter § X)*.

Third, we uncover an extensive set of provisions expressly barring non-citizens in occupational licensing. As Calvo-Friedman discusses [15], courts have largely treated alienage-based licensing requirements as unconstitutional. Alienage classifications are subject to strict scrutiny, *Graham v. Richardson*, 403 U.S. 365 (1971), and states may not condition professional licenses on citizenship, see, e.g., *Bernal v. Fainter*, 467 U.S. 216 (1984) (notary public); *In re Griffiths*, 413 U.S. 717 (1973) (attorney), absent a narrow exception for positions involving democratic self-governance, see, e.g., *Foley v. Connelie*, 435 U.S. 291 (1978); *Ambach v. Norwick*, 441 U.S. 68 (1979) (both upholding citizenship requirements for public employment). Yet we find a myriad of restrictions in occupations far outside the bounds of recognized exceptions. Non-citizens may be prohibited from operating businesses such as go-cart tracks; bowling alleys; taxi or limousine services; bus or tow truck companies; casinos; pawnshops; used car dealerships; hotels; bars; arcades; dance halls; cabarets; massage parlors; jewelry shops; tattoo studios; boarding houses; and private investigator agencies, to name a few. These laws exist across over a thousand jurisdictions (see Table 1).

Fourth, we identify a wide range of laws discriminating on the basis of sex or gender. These laws include provisions restricting voting eligibility to men (City of Opp, Alabama Charter § 5), as well as gender-segregated employment, such as a local code in Massachusetts specifying the composition of a police department as "not less than 85 and not more than 113 patrolmen, a matron of police, [and] two police women" (City of Holyoke, MA, Code § 50-31). Some of these provisions may have begun as reformist efforts to expand women's role in society [13], but are now leftover artifacts, illustrating the interpretive complexity we return to in our analysis of model–human disagreement.

Finally, we identify provisions that likely conflict with federal age discrimination law. For example, one Pennsylvania city requires that: *"All employees, except elected officers, eligible for full pension hereunder, shall retire at the age of 70 years." (Hazleton, PA § 90-6)* Because of how broadly this provision sweeps, it likely violates the Age Discrimination in Employment Act, which prohibits mandatory retirement.

Table 1: Examples of discriminatory provisions appearing in local codes and charters available online. J indicates a lower bound on the number of jurisdictions featuring each provision type; R indicates number of residents living in those jurisdictions. More examples are provided in Appendix Table 6.

Type	J	R	Example
Racial segregation	15	1.8M	"The town council shall have the power to enact ordinances....… **providing separate cemeteries for white and black**" *(Smithfield, NC, Charter § 37)*
Race or gender-based voter eligibility	29	101k	"Voters… shall consist of **all male persons** of the age of twenty-one and upwards" *(Opp, AL, Charter § 5)*
Race-based voting restrictions	18	189K	"No person shall be permitted to vote [without] pay[ing]… **poll taxes**." *(Valparaiso, FL, Charter § 4)*
Gender-based indecent language restrictions	19	400k	"It shall be unlawful… to use indecent or obscene language in the presence… of **any woman or child**." *(Walled Lake, MI, Code § 50-157)*
Women-only police roles	50	2.6M	"The police department shall consist of… patrolmen… **a matron of police, two police women**" *(Holyoke, MA, Code § 17-16)*
Citizenship-based occupational licensing	1.6k	49M	"**No [junk dealer] license shall be issued**… if the applicant is… **an alien**" *(Alameda County, CA, Code § 3.48.060)*

5.2 Offensive Language

Our pipeline flags 30,584 provisions as containing potentially offensive language. We note two caveats before presenting examples. First, "offensive" is a broad and contested category; some terms flagged by our pipeline have been reclaimed by the communities they once targeted or remain contested within them. Second, some flagged provisions contain language that is demeaning but does not operate against a protected category, while others may expressly target specific communities.

Some provisions (see Table 2) contain explicit slurs or derogatory labels which often also evidence histories of exclusion. For example, the corpus includes ordinances referencing "gypsies" and "gypsy cunning"—a pejorative exonym for Romani people—within provisions restricting fortune-telling, an occupation associated with Romani communities, and thereby targeting that community. We also identify repeated use of the word *"Dago,"* a term historically directed at Italian immigrants and Italian Americans, appearing in our corpus in reference to *"Dago bombs"*, a type of firework.

The persistence of this language reflects the uneven pace of code revision. In the 1980s, California cities acknowledged copying anti-Romani references from neighboring jurisdictions' codes; while some like Los Angeles moved to revise their ordinances, the language remained in others, such as Arcadia [78]. A parallel pattern appears with model codes: officials removed "dago" from the

National Fire Prevention Code in the 1980s [43], but despite its national-level removal, over 100 local provisions in our corpus still contain the term.

Table 2: Sample of local laws with offensive language.

City	Law
Parkersburg, WV	"No circus, show or street carnival shall be licensed which has any **gypsy fortunetellers** or fortunetellers of any kind, or **gypsies connected therewith in any manner**." *(Code § 713.01)*
Houston, TX	"The City of Houston shall have power… to **restrain and punish vagrants, mendicants, beggars and prostitutes**." *(Code § 16)*
Lewisburg, OH	"Each person, desiring to conduct… a circus, menagerie, carnival, sideshow, musical or **minstrel entertainment, or exhibition of monsters or freaks of nature**… shall first obtain a license" *(Code § A7.204)*
Euclid, OH	"Every person… housing… **invalids**, children, aged persons or individuals adjudged… as **feeble-minded, insane, lunatics, imbeciles or idiots**, shall… comply with all the laws of the State" *(Code § 753.13)*
Texarkana, TX	"The term 'disability'…does not apply to an individual because… that individual is a **transvestite**." *(Code § 16.3)*
Linn County, IA	"Skin type: III. Skin reactions…: Burns moderately and tans about average. Examples: Normal average **Caucasoid**." *(Code Appendix B)*

Additionally, some provisions rely on terminology that was once treated as technical or descriptive but is now widely recognized as outdated or inaccurate, including: *"feeble-minded,"* historically used to describe individuals presumed to lack intelligence; *"transvestite,"* a stigmatizing term; and *"Caucasoid,"* an obsolete, unscientific racial classification.

The corpus also contains references to historically discriminatory social practices embedded within statutory text. Examples include ordinances regulating *"minstrel shows,"* which are racially caricatured performances, as well as references to exhibitions of *"monsters"* and *"freaks of nature,"* phrases historically used to describe individuals with disabilities or unique physical traits in a dehumanizing manner. Additionally we surface laws with archaic and charged terminology such as *"buggery"* for sodomy. Other provisions employ language which marginalizes and polices perceived moral deviance, including statutes penalizing *"beggars,"* or *"vagrants."*

5.3 Validation using External Sources

Our prompting approach demonstrates strong performance in identifying provisions involving protected-category-based differential treatment. The model achieves an overall recall of 0.91 and an overall precision of 0.91 for identifying provisions involving differential treatment based on protected categories. In addition, 98.8% of provisions containing protected-category-based differential treatment are correctly flagged by the model. When restricting evaluation to provisions in our validation set that were drawn from existing datasets (excluding those obtained through manual searches), performance is comparable: recall is 0.90, precision is 0.91, and 98.3% of provisions containing protected-category-based differential treatment are correctly identified.

We find that Gemini-2.5-Flash's classifications of discriminatory attributes in this validation set exhibit strong alignment with human annotations (see Appendix Table 5). Additionally, we find that 91.6% of the provisions that the model classified as satisfying all five discriminatory criteria are either unconstitutional or in violation of federal anti-discrimination law. In addition, among the validation set provisions evaluated, 98.0% of model interpretations include cited legal references that annotators find relevant to the form of treatment described in the provision. For 90.2% of provisions, the model's interpretations of the cited authorities reflect, in general terms, the underlying legal holdings. Finally, we agree with the LLM-generated HIGH/MED/LOW ratings for 96.1% of provisions.

5.4 Validation Using Pipeline Output

Among flagged provisions, how well do LLMs fare in making substantive assessments under antidiscrimination law? Compared with human annotations, the pipeline is over-inclusive in flagging provisions as discriminatory (see Figure 3): it labels 64% of provisions as high priority, compared with 34% under our human annotations. Correspondingly, agreement on the binary high-priority classification is fair ($\kappa = 0.31$).

For provisions labeled as high priority by human annotators, the LLM achieves high recall (88%). However, its precision is substantially lower (46%). Annotators judged more than half of the provisions the model flagged as high priority to be ambiguous or defensible. This reflects our design choice; we tuned the pipeline to prioritize recall so that fewer discriminatory provisions escape human review.

Model performance varies considerably across substantive clusters. Certain clusters—most notably those involving explicit citizenship-based restrictions—exhibit near-perfect precision and recall, reflecting strong alignment between the model and human annotators. Agreement is substantially lower in clusters related to gender, where the model frequently assigns high-priority labels to gendered language that does not impose substantive differential treatment.

We also observe signs of distributional shift between the validation set and the provisions surfaced by the scaled pipeline. This observation suggests that the broader corpus contains a larger share of legally ambiguous provisions than we could source in our validation set; given the wide range of topics in local law, anticipating every edge case in advance is difficult.

The value of our pipeline, then, lies in surfacing provisions for review rather than in rendering substantive legal assessments. By reducing the corpus by roughly three orders of magnitude, the pipeline makes human review tractable, leaving the interpretive work of applying antidiscrimination doctrine to reviewers.

5.5 Model–Human Disagreement Patterns

Type I Error: Constitutionally suspect under other doctrines. In our evaluation sample, the model flags as HIGH priority a Woodbury, Georgia charter provision empowering the mayor to prohibit *"the pursuing of one's ordinary avocation or labor on the Sabbath" (Charter § 19)* and a City of Chapin, South Carolina, law prohibiting *"cults, masked faces or organizations practicing discrimination" (Code § 14.1003)* from assembling. Human raters instead assigned these provisions MEDIUM or LOW priority, reasoning that while constitutionally suspect, these were primarily First Amendment concerns rather than Equal Protection violations. This disagreement likely stems from under-specification in our prompt: we instruct the

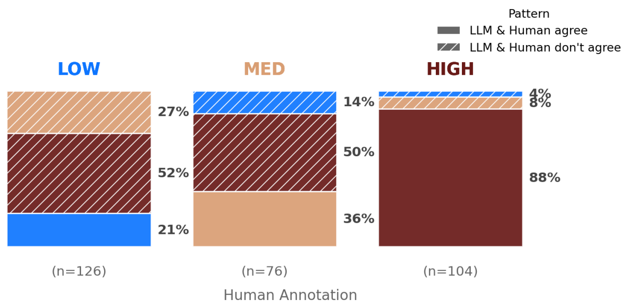

Figure 3: LLM and human priority agreement. Solid fills indicate the proportion of human classifications at that priority level which the LLM classified similarly; hatched segments show unaligned annotations.

model to analyze provisions "under the Fourteenth Amendment"; the model reasonably interprets Fourteenth Amendment analysis to include incorporated First Amendment rights, while human raters apply the narrower Equal Protection lens our study intends.

Type I Error: Historical terminology misinterpreted as discriminatory. The model sometimes misreads historically contingent language as discrimination. A 1931 Paulsboro, New Jersey ordinance regulating *"Shows and Exhibitions"* requires permits for merchandise sales by *"persons commonly known as fakirs." (Code § 64-3)* The model assigns HIGH priority, reasoning that "fakir" references religion and therefore implicates religious classification. Human raters instead assign LOW priority, interpreting "fakir" in its historical context as a period term for street-show performers, rather than a regulation targeting Muslim religious practice. These errors illustrate how archaic terminology may produce false positives.

Type II Error: Deference to governmental interests. These Type II errors occur when the model accepts an offered government justification too readily and assigns a lower priority to provisions that may still warrant scrutiny. For example, in evaluating a Puerto Rico law stating that *"the Authority may only issue pilot licenses to natural persons, citizens of the United States, of age, and of good repute," (Code § 2404)* the model reasons that citizenship requirements may fall within the political-function exception. It concludes that harbor pilots *"exercise significant discretionary authority and responsibility for public safety and commerce"* and therefore lowers the provision's priority. This reasoning under-emphasizes the breadth of the exclusion; it is not obvious that all licensed harbor pilots qualify under the model's reasoning. In such cases, the model's deference to plausible governmental justifications can de-prioritize provisions that merit closer human review.

6 Discussion

We present a structured pipeline for aiding in what to date has been done by intensive, manual work: crafting the litigation blueprint for civil rights. Our results reveal significant promise for LLM-assisted detection of potentially discriminatory laws to scale review and reform.

We focus on antidiscrimination law, but a similar approach applies to other surveys where initial filtering can rely on textual features and irrelevant clusters are easy to spot. Antidiscrimination law suits this approach well. First, the field has increasingly turned toward the principle of anticlassification—the idea that the law may not classify people by protected categories—which, with attention to doctrinal subtleties, makes potential violations tractable to flag through textual methods. Second, when a survey's criteria focuses on what laws do, clustering on that dimension makes the legal question at hand clear for further research. Our survey fits this pattern. For example, one cluster contains laws requiring litigants to self-fund English interpreters in legal settings; on further review, we found the Obama-era Department of Justice Civil Rights Division identified the practice as inconsistent with the agency's expectations under Title VI of the Civil Rights Act [57]. Surveys that hinge on specific wording, by contrast, may not cluster as cleanly.

We acknowledge several limitations. First, some might object that many of these provisions may not be in effect. Indeed, while we exclude provisions annotated as repealed, some laws may signal repeal in other ways, such as placement in a historical section of the code. Our pipeline may miss these rare indicators. Additionally, the laws may simply be not in practical effect at all. This may be true, but we surface large numbers of provisions (*e.g.*, citizenship restrictions in licensing) that some jurisdictions have encountered and recently moved to repeal. For example, Highland Park, Illinois, removed its citizenship requirement for liquor licenses in 2026 after a non-citizen applicant raised the issue [67]. Even if not practiced, laws have an expressive function [7, 69] stigmatizing groups in ways that conflict with the promises of equal protection.

Second, our analysis operates at the section level rather than the code level. We evaluate provisions without regard for cross-references to other sections that may supersede or otherwise qualify a provision's apparent meaning [39, 48]. While this design choice enables our large-scale analysis, it risks overstating the legal effect of provisions that cities have constrained elsewhere in the code.

Third, our approach does not capture discrimination expressed through facially neutral language. As lawmakers increasingly avoid overtly discriminatory terms, explicit exclusions have often given way to what Sofia Martos describes as "coded codes"—statutes that appear neutral on their face while achieving discriminatory ends in practice [49]. For example, cities in the 1990s and 2000s targeted undocumented migrants using trespassing and anti-solicitation ordinances, with courts later striking down those applications as unconstitutional [76]. Our analysis also does not reach to discriminatory enforcement.

Finally, we show that LLM capacity to make substantive legal assessments remains limited. Constitutional law evolves, sometimes dramatically. That said, the fluidity and uncertainty of the doctrine is also what poses challenges to human reviewers in prioritizing provisions. As Susannah Pollvogt observes, "the central challenge of equal protection jurisprudence is not to account for prejudices already apparent to us…but for those that today seem natural, familiar, and fair" [58].

Regardless, the search problem, especially for local law, has stymied code cleanup efforts for decades. We show that in one important domain—antidiscrimination law—modern AI systems offer considerable promise.

Generative AI assistance

We used ChatGPT and Claude for assistance with editing and for supporting scripting tasks (*e.g.*, automating PDF parsing quality checks; assisting with crosswalking place names to Census data; stylizing graphics and figures).

Acknowledgments

We thank Mohamed Afane, danah boyd, Allison Casasola, Gabriel J. Chin, Lindsey Gailmard, Isabel Gallegos, James Grimmelmann, Julia Greenberg, Emma Harvey, Ananya Karthik, Karen Levy, Emma Lurie, Erin Maneri, Emily Robitschek, Danny Sallis, Judy Shen, Alexander Spangher, Lucia Zheng, members of the Cornell Artificial Intelligence, Policy, and Practice initiative, and the Google Cloud research credits program.

Supplementary Information

Appendices, including LLM prompts, are available at https://hidden-in-plain-text.reglabapp.com/Appendix.pdf.

References

[1] [n. d.]. *Arlington Texas City Charter & Code of Ordinances.* https://www.arlingtontx.gov/Government/City-Government/City-Secretary/City-Charter-Code-of-Ordinances
[2] 2020. Identifying and Addressing the Vestiges of Inequity and Inequality in Virginia's Laws. https://www.fairfaxcounty.gov/boardofsupervisors/sites/boardofsupervisors/files/assets/meeting-materials/2021/feb12-legislative-racial-inequity-in-va-law-report.pdf
[3] General Code [n. d.]. *Invest in an Enforceable Code of Ordinances.* General Code. https://www.generalcode.com/codification/
[4] [n. d.]. *Racial Restrictive Covenants Project, Washington State.* https://depts.washington.edu/covenants/
[5] City of Portland (Auditor's Office / Archives) [n. d.]. *Statement on Harmful and Bias Language in Archival Description.* City of Portland (Auditor's Office / Archives). https://www.portland.gov/auditor/archives/harmful-and-bias-language-statement
[6] CivicPlus [n. d.]. *What to Expect During Each Code Process.* CivicPlus. https://www.civicplus.help/codification/docs/what-to-expect-during-each-code-process
[7] Elizabeth S. Anderson and Richard H. Pildes. 2000. Expressive Theories of Law: A General Restatement. 148, 5 (2000), 1503–1575. jstor:3312748 doi:10.2307/3312748
[8] Elliott Ash, Christoph Goessmann, and Suresh Naidu. 2024. Scaling Laws: Legal and Social Complexity in US Localities. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 382, 2270 (April 2024), 20230151. doi:10.1098/rsta.2023.0151
[9] Alicia Bannon. 2025. Stare Decisis and Zombie Laws. *St. John's Law Review* 98, 7 (2025), [article 8].
[10] Alexander Bartik, Arpit Gupta, and Daniel Milo. 2025. The Costs of Housing Regulation: Evidence from Generative Regulatory Measurement. *Available at SSRN* (2025). doi:10.2139/ssrn.4627587 SSRN 4627587.
[11] Andrew Blair-Stanek, Nils Holzenberger, and Benjamin Van Durme. 2023. Can GPT-3 Perform Statutory Reasoning?. In *Proceedings of the Nineteenth International Conference on Artificial Intelligence and Law* (Braga Portugal, 2023-06-19). ACM, 22–31. doi:10.1145/3594536.3595163
[12] Maureen E. Brady. 2021. Zombie State Constitutional Provisions. *Wisconsin Law Review* 2021 (2021), 1063. doi:10.2139/ssrn.3920420 Harvard Public Law Working Paper No. 22-01.
[13] Cynthia Brenwall. 2023. Policewomen.
[14] Abigail Brone. 2021. *Norwalk Expands Gender Definition to Include 'She' and 'They' in City Codes.* Norwalk Hour. https://www.thehour.com/news/article/Norwalk-expands-gender-definition-to-include-16124608.php
[15] Jennesa Calvo-Friedman. 2014. The uncertain terrain of state occupational licensing laws for noncitizens: A preemption analysis. *Georgetown Law Journal* 102 (2014), 1597.
[16] Gabriel "Jack" Chin, Roger E. Hartley, Kevin Bates, Rona Nichols Kreamer, Ira J. Shiflett, and Salmon A. Shomade. 2009. Still on the Books: Jim Crow and Segregation Laws Fifty Years after Brown v. Board of Education a Report on Laws Remaining in the Codes of Georgia, Louisiana, Mississippi, Missouri, South Carolina, Virginia, and West Virginia. *Michigan State Law Review* (2009). https://papers.ssrn.com/abstract=1428586
[17] CivicPlus. 2022. *How to Neutralize the Language in Your Code of Ordinances.* CivicPlus. https://www.civicplus.com/blog/cs/neutralize-the-language-in-your-code-of-ordinances/
[18] Andrew Coan and Harry Surden. 2025. Artificial Intelligence and Constitutional Interpretation. 96, 2 (2025), 413–498. https://heinonline.org/HOL/P?h=hein.journals/ucollr96&i=412
[19] Cornell Law School Legal Information Institute. 2020. Protected Characteristic. https://www.law.cornell.edu/wex/protected_characteristic.
[20] Laura Deal. 2022. *Survey of State Marriage Laws Related to Same-Sex Marriage.* Legal Sidebar LSB10866. Congressional Research Service, Library of Congress. https://crsreports.congress.gov/product/pdf/LSB/LSB10866 Published November 22, 2022.
[21] Jesse M. Engebretson, Kristen C. Nelson, Kelli L. Larson, Riley Andrade, Megan M. Wheeler, Susannah B. Lerman, Dexter H. Locke, Tara L. E. Trammell, and Peter M. Groffman. 2023. Ambiguity and Clarity in Residential Yard Ordinances across Metropolitan Areas in the United States. 45, 5 (2023), 1022–1039. doi:10.1080/07352166.2021.1901590
[22] Christoph Engel and Richard H. McAdams. 2024. Asking GPT for the Ordinary Meaning of Statutory Terms. ssrn:4718347 doi:10.2139/ssrn.4718347 Also circulated as University of Chicago, Public Law Working Paper No. 848.
[23] Brenner M. Fissell. 2020. Local Offenses. 89, 3 (2020), 837–888.
[24] Brenner M. Fissell. 2023. Rightsizing Local Legislatures. 2023, 2 (2023), 393–445. https://heinonline.org/HOL/P?h=hein.journals/utahlr2023&i=389
[25] Jens Frankenreiter and Michael A. Livermore. 2020. Computational Methods in Legal Analysis. 16, 1 (2020), 39–57. doi:10.1146/annurev-lawsocsci-052720-121843
[26] Melissa S. Fry. 2001. Dormant Statutes and the Legal Concept of Desuetude. 20, 4 (2001), 67–84. doi:10.1300/J113v20n04_03
[27] Maggie Fusek and Patch Staff. 2024. *CoCo Supervisors Remove Gender-Specific Pronouns From Ordinance Code.* Concord, CA Patch. https://patch.com/california/concord-ca/coco-supervisors-remove-gender-specific-pronouns-ordinance-code
[28] Hillary Greene. 1997. Undead Laws: The Use of Historically Unenforced Criminal Statutes in Non-Criminal Litigation. 72 (1997). https://digitalcommons.lib.uconn.edu/law_papers/72
[29] James Grimmelmann, Benjamin L W Sobel, and David Stein. 2025. Generative Misinterpretation. 63, 1 (2025).
[30] Justin Grimmer and Brandon M. Stewart. 2013. Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. 21, 3 (2013), 267–297. doi:10.1093/pan/mps028
[31] Luke Guerdan, Solon Barocas, Kenneth Holstein, Hanna Wallach, Zhiwei Steven Wu, and Alexandra Chouldechova. 2025. *Validating LLM-as-a-Judge Systems in the Absence of Gold Labels.* arXiv:2503.05965 [cs] doi:10.48550/arXiv.2503.05965
[32] Neel Guha, Julian Nyarko, Daniel E. Ho, Christopher Ré, Adam Chilton, Aditya Narayana, Alex Chohlas-Wood, Austin Peters, Brandon Waldon, Daniel N. Rockmore, Diego Zambrano, Dmitry Talisman, Enam Hoque, Faiz Surani, Frank Fagan, Galit Sarfaty, Gregory M. Dickinson, Haggai Porat, Jason Hegland, Jessica Wu, Joe Nudell, Joel Niklaus, John Nay, Jonathan H. Choi, Kevin Tobia, Margaret Hagan, Megan Ma, Michael Livermore, Nikon Rasumov-Rahe, Nils Holzenberger, Noam Kolt, Peter Henderson, Sean Rehaag, Sharad Goel, Shang Gao, Spencer Williams, Sunny Gandhi, Tom Zur, Varun Iyer, and Zehua Li. 2023. LEGALBENCH: a collaboratively built benchmark for measuring legal reasoning in large language models. In *Proceedings of the 37th International Conference on Neural Information Processing Systems* (New Orleans, LA, USA) *(NIPS '23)*. Curran Associates Inc., Red Hook, NY, USA, Article 1915, 157 pages.
[33] Neel Guha, Julian Nyarko, Daniel E. Ho, and Christopher Ré. 2025. Building GenAI Benchmarks: A Case Study in Legal Applications. In *The Oxford Handbook of the Foundations and Regulation of Generative AI* (1 ed.), Philipp Hacker, Andreas Engel, Sarah Hammer, and Brent Mittelstadt (Eds.). Oxford University Press, 0. doi:10.1093/oxfordhb/9780198940272.013.0007
[34] Andrew Halterman and Katherine A. Keith. 2025. *What Is a Protest Anyway? Codebook Conceptualization Is Still a First-Order Concern in LLM-Era Classification.* arXiv:2510.03541 [cs] doi:10.48550/arXiv.2510.03541
[35] Emaan Hariri and Daniel E. Ho. 2025. AI for Statutory Simplification: A Comprehensive State Legal Corpus and Labor Benchmark. arXiv. doi:10.48550/arXiv.2508.19365
[36] Megan E Hatch and Joseph W Mead. 2021. Learning from Laboratory Mistakes: How Policy Entrepreneurs Catalyze City Ordinance Repeals in the United States. 36, 3 (2021), 361–378. doi:10.1177/0952076719840070
[37] Luxi He, Nimra Nadeem, Michel Liao, Howard Chen, Danqi Chen, Mariano-Florentino Cuéllar, and Peter Henderson. 2025. *Statutory Construction and Interpretation for Artificial Intelligence.* arXiv:2509.01186 [cs] doi:10.48550/arXiv.2509.01186
[38] Amanda Henley, Lorin Bruckner, Hannah Jacobs, Matthew Jansen, Brianna Nunez, Rolando Rodriguez, and Morgan Wilson. 2024. On the Books: Jim Crow and

Algorithms of Resistance, a Collections as Data Case Study. *Journal of Computing and Cultural Heritage* 16, 4, Article 85 (jan 2024). doi:10.1145/3631128

[39] Nils Holzenberger and Benjamin Van Durme. 2021. Factoring Statutory Reasoning as Language Understanding Challenges. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)* (Online, 2021-08), Chengqing Zong, Fei Xia, Wenjie Li, and Roberto Navigli (Eds.). Association for Computational Linguistics, 2742–2758. doi:10.18653/v1/2021.acl-long.213

[40] Morton Horwitz, Martha Field, Martha Minow, et al. 2000. Brief of Morton Horwitz, Martha Field, Martha Minow, and Over 100 Other Historians and Scholars as Amici Curiae in Support of Respondents. Brief submitted to the U.S. Supreme Court. https://www.raggededgemagazine.com/garrett/statediscrim/statediscrim.htm In University of Alabama at Birmingham Board of Trustees v. Garrett, 2000 WL 1154025.

[41] Philip K. Howard. 2014. *The Rule of Nobody: Saving America from Dead Laws and Broken Government.* W. W. Norton & Company, New York.

[42] Hydee Feldstein Soto. 2024. Revising the Los Angeles Municipal Code and the Los Angeles Administrative Code to Remove Specific References to Gender and Make Other Technical Errors. https://web.archive.org/web/20240706080116/https://clkrep.lacity.org/onlinedocs/2023/23-0047_rpt_atty_4-09-24.pdf

[43] United Press International. 1986. *'Dago bomb' stricken from code.* UPI Archives. https://www.upi.com/Archives/1986/07/30/Dago-bomb-stricken-from-code/1568523080000/

[44] Joel Johnson. 2022. Dealing with Dead Crimes. *Georgetown Law Journal* 111 (2022), 95. doi:10.2139/ssrn.4040948 Pepperdine University Legal Studies Research Paper No. 2022/20.

[45] Allison Koenecke, Jed Stiglitz, David Mimno, and Matthew Wilkens. 2025. *Tasks and Roles in Legal AI: Data Curation, Annotation, and Verification.* arXiv:2504.01349 [cs] doi:10.48550/arXiv.2504.01349

[46] Kristi Jourdan. 2024. *Board of Supervisors Modernizes County Ordinance Code by Removing Gender-Specific Pronouns.* Board of Supervisors. https://www.contracosta.ca.gov/DocumentCenter/View/85050/Board-of-Supervisors-Modernizes-County-Ordinance-Code-by-Removing-Gender-Specific-Pronouns-News-Release-1232024

[47] Sarah Kroeger and Giulia La Mattina. 2020. Do Nuisance Ordinances Increase Eviction Risk? 110 (2020), 452–456. doi:10.1257/pandp.20201119

[48] Sarah B. Lawsky. 2018. A Logic for Statutes. 21, 1 (2018), 1–48. https://scholarship.law.ufl.edu/ftr/vol21/iss1/2

[49] Sofia D. Martos. 2010. Coded Codes: Discriminatory Intent, Modern Political Mobilization, and Local Immigration Ordinances. 85, 6 (2010), 2099–2137. https://heinonline.org/HOL/P?h=hein.journals/nylr85&i=2111

[50] Ashley McBride. 2019. *Banned Words: Berkeley Drops Gendered Language from City Codes.* San Francisco Chronicle. https://www.sfchronicle.com/bayarea/article/Banned-words-Berkeley-drops-he-she-14102930.php

[51] Nathan McClintock, Esperanza Pallana, and Heather Wooten. 2014. Urban Livestock Ownership, Management, and Regulation in the United States: An Exploratory Survey and Research Agenda. 38 (2014), 426–440. doi:10.1016/j.landusepol.2013.12.006

[52] Joseph Mead, Megan Hatch, J. Rosie Tighe, Marissa Pappas, Kristi Andrasik, and Elizabeth Bonham. 2017. Who Is a Nuisance? Criminal Activity Nuisance Ordinances in Ohio. (2017). doi:10.2139/ssrn.3067028

[53] Bryant J Moy. [n. d.]. Racial Threat and the Emergence of Discriminatory Ordinances. ([n. d.]).

[54] Pauli Murray (Ed.). 1950. *States' Laws on Race and Color, and Appendices: Containing International Documents, Federal Laws and Regulations, Local Ordinances and Charts.* Women's Division of Christian Service, Board of Missions and Church Extension, The Methodist Church, Cincinnati, OH. https://books.google.com/books?id=0452AAAAMAAJ

[55] Laura K. Nelson, Derek Burk, Marcel Knudsen, and Leslie McCall. 2021. The Future of Coding: A Comparison of Hand-Coding and Three Types of Computer-Assisted Text Analysis Methods. 50, 1 (2021), 202–237. doi:10.1177/0049124118769114

[56] Arnaldo Mailes Neto, Thiago Santiago Gomes, Mônica Pertel, Louise A. V. P. Vieira, and Elen B. A. V. Pacheco. 2021. An Overview of Plastic Straw Policies in the Americas. 172 (2021), 112813. doi:10.1016/j.marpolbul.2021.112813

[57] Thomas E. Perez. 2010. Letter from Thomas E. Perez, Assistant Attorney General, Civil Rights Division, to Chief Justices and State Court Administrators. U.S. Department of Justice, Civil Rights Division. https://perma.cc/M24N-SXHK Archived at https://perma.cc/M24N-SXHK.

[58] Susannah W. Pollvogt. 2014. Beyond Suspect Classifications. 16, 3 (2014), 739–796. https://scholarship.law.upenn.edu/jcl/vol16/iss3/4

[59] Dasha Pruss and Jessie Allen. 2025. Against AI Jurisprudence: Large Language Models and the False Promises of Empirical Judging. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society* (2025), Vol. 8. 2055–2066. https://ojs.aaai.org/index.php/AIES/article/view/36695

[60] Ben Railton. 2023. *Considering History: Sundown Towns, Racism, and Exclusion.* The Saturday Evening Post. https://www.saturdayeveningpost.com/2023/07/considering-history-sundown-towns-racism-and-exclusion/

[61] Hope Schroeder, Marianne Aubin Le Quéré, Casey Randazzo, David Mimno, and Sarita Schoenebeck. 2025. Large Language Models in Qualitative Research: Uses, Tensions, and Intentions. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25).* Association for Computing Machinery, New York, NY, USA, Article 481, 17 pages. doi:10.1145/3706598.3713120

[62] Kathryn Schulz. 2017. The Many Lives of Pauli Murray. (2017). https://www.newyorker.com/magazine/2017/04/17/the-many-lives-of-pauli-murray

[63] Shreya Shankar, Tristan Chambers, Tarak Shah, Aditya G. Parameswaran, and Eugene Wu. 2025. DocETL: Agentic Query Rewriting and Evaluation for Complex Document Processing. *Proc. VLDB Endow.* 18, 9 (May 2025), 3035–3048. doi:10.14778/3746405.3746426

[64] Shreya Shankar, Bhavya Chopra, Mawil Hasan, Stephen Lee, Bjoern Hartmann, Joseph Hellerstein, Aditya Parameswaran, and Eugene Wu. 2025. Steering Semantic Data Processing With DocWrangler. In *Proceedings of the 38th Annual ACM Symposium on User Interface Software and Technology (UIST '25).* Association for Computing Machinery, New York, NY, USA, Article 84, 18 pages. doi:10.1145/3746059.3747625

[65] Michael L Smith. 2024. Constitutional Interpretation and Zombie Provisions. 40 (2024).

[66] Jennifer Sorentrue. 2010. *Offensive Place Name Being Deleted from County Records.* The Palm Beach Post. https://www.palmbeachpost.com/story/news/2010/04/23/offensive-place-name-being-deleted/7305502007/

[67] Joseph States. 2026. Highland Park removes liquor-license citizenship requirement: "Certain laws ... we just need to get off our books". *Chicago Tribune.* https://www.chicagotribune.com/2026/01/16/highland-park-liquor-licenses/

[68] Hongjin Su, Weijia Shi, Jungo Kasai, Yizhong Wang, Yushi Hu, Mari Ostendorf, Wen-tau Yih, Noah A Smith, Luke Zettlemoyer, and Tao Yu. 2023. One embedder, any task: Instruction-finetuned text embeddings. In *Findings of the Association for Computational Linguistics: ACL 2023.* 1102–1121.

[69] Cass R. Sunstein. 1996. On the Expressive Function of Law. 144, 5 (1996), 2021. jstor:3312647 doi:10.2307/3312647

[70] Cass R. Sunstein. 2003. What Did Lawrence Hold? Of Autonomy, Desuetude, Sexuality, and Marriage. *University of Chicago Law & Economics Working Papers* 615 (2003). https://chicagounbound.uchicago.edu/law_and_economics/615

[71] Faiz Surani, Mirac Suzgun, Vyoma Raman, Christopher D. Manning, Peter Henderson, and Daniel E. Ho. 2025. *AI for Scaling Legal Reform: Mapping and Redacting Racial Covenants in Santa Clara County.* doi:10.48550/arXiv.2503.03888

[72] Rosamond Thalken, Edward Stiglitz, David Mimno, and Matthew Wilkens. 2023. Modeling Legal Reasoning: LM Annotation at the Edge of Human Agreement. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing* (Singapore, 2023-12), Houda Bouamor, Juan Pino, and Kalika Bali (Eds.). Association for Computational Linguistics, 9252–9265. doi:10.18653/v1/2023.emnlp-main.575

[73] U.S. Census Bureau. [n. d.]. *Census of Governments (CoG).* Census.gov. https://www.census.gov/programs-surveys/cog.html

[74] U.S. Commission on Civil Rights. 1977. Sex Bias in the U.S. Code: A Report of the U.S. Commission on Civil Rights. (1977). https://www.usccr.gov/files/historical/1977/77-009.pdf

[75] Abigail VanderMolen. 2025. *Midland City Council Declines to Pursue Gender-Neutral Charter Change, Considers Code Updates.* Midland Daily News. https://www.ourmidland.com/news/article/midland-charter-change-code-updates-21094680.php

[76] Monica W. Varsanyi. 2008. Immigration Policing Through the Backdoor: City Ordinances, the "Right to the City," and the Exclusion of Undocumented Day Laborers. 29, 1 (2008), 29–52. doi:10.2747/0272-3638.29.1.29

[77] Angelina Wang, Michelle Phan, Daniel E. Ho, and Sanmi Koyejo. 2025. Fairness through Difference Awareness: Measuring Desired Group Discrimination in LLMs. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (Vienna, Austria, 2025-07), Wanxiang Che, Joyce Nabende, Ekaterina Shutova, and Mohammad Taher Pilehvar (Eds.). Association for Computational Linguistics, 6867–6893. doi:10.18653/v1/2025.acl-long.341

[78] Mike Ward. 1985. Council to Remove 'Gypsy' Reference. *Los Angeles Times* (29 September 1985). https://www.latimes.com/archives/la-xpm-1985-09-29-ga-18997-story.html Accessed: 2026-04-19.

[79] Howard M Wasserman. 2022. Zombie Laws. *Lewis & Clark Law Review* 25 (2022), 1047. https://lawcommons.lclark.edu/lclr/vol25/iss4/3

[80] Andrea W Wen-Yi, Kathryn Adamson, Nathalie Greenfield, Rachel Goldberg, Sandra Babcock, David Mimno, and Allison Koenecke. 2024. Automate or Assist? The Role of Computational Models in Identifying Gendered Discourse in US Capital Trial Transcripts. 7 (2024), 1556–1566. doi:10.1609/aies.v7i1.31746

[81] Kellen Zale. 2019. Part-Time Government. 80, 5 (2019), 987–1054. https://heinonline.org/HOL/P?h=hein.journals/ohslj80&i=1017

[82] Lucia Zheng, Neel Guha, Javokhir Arifov, Sarah Zhang, Michal Skreta, Christopher D. Manning, Peter Henderson, and Daniel E. Ho. 2025. A Reasoning-Focused Legal Retrieval Benchmark. In *Proceedings of the 2025 Symposium on Computer Science and Law* (Munich, Germany) *(CSLAW '25).* Association for Computing Machinery, New York, NY, USA, 169–193. doi:10.1145/3709025.3712219