

# Appendix for *Hidden in Plain Text: LLM-Assisted Detection of Discriminatory Municipal Laws*

## A Estimating Population

Municipal code sources use naming conventions that do not always align with Census jurisdiction names. To align each jurisdiction with a Census place or county, we apply a normalization and matching script (e.g., “hts” for “heights,” “twp” for “township”), which matches 98.5% of our corpus to Census records. For spatial boundaries, we map every matched city and county onto Census geographies using the Census’ TIGER 2023 shapefiles [2]. This mapping identifies which counties each city intersects and, for multi-county places, how much of the city falls within each county.

Population figures come from the Census Population Estimates Program (2023) for places (incorporated cities, towns, etc.) and counties, each identified by a GEOID, name, and 2023 population estimate [1]. To avoid double-counting, when both a county and one of its constituent cities appear in the dataset, we count only the county’s population; cities contained within a covered county contribute zero population. We estimate our codes cover approximately 252.8 million people or about 75.48% of the U.S. population.

## B Prompts

### B.1 Identifying References to Protected Categories and Differential Treatment Prompt

You are a legal research expert. Your task is to classify each legal provision by identifying whether it references protected categories, proxies, or socioeconomic groups, and whether it applies group-specific treatment. Follow the steps in order.

#### Step 1: Detect References

Check if the provision falls into one or more of the following categories:

##### 1.1 Protected Categories

A provision references a protected category if the provision does any of the following:

Directly mentions a protected class (race; religion; sex/gender/sexual orientation; marital status; national origin/alienage/ancestry; genetic information; age (over 40); disability/medical condition;).

Defines a term that incorporates a protected class.

Cross-references another law that explicitly names a protected class.

Uses gender-specific language (e.g. only using male pronouns for job descriptions)

Note: Do not consider age a protected category for ages not over 40 (eg. minors or children).

##### 1.2 Proxies

A provision references a proxy if the law uses a non-explicit criterion that commonly functions as a stand-in for a protected trait or enables discriminatory treatment.

Examples of proxies/enablers:

Language, literacy, English-only requirements (proxy for national origin).

Chastity (historical exclusions).

Residency duration/domicile rules (proxy for alienage/national origin).

Marital/parental status when irrelevant to the regulated task.

Grooming/hairstyle restrictions (cornrows, locs, beards (proxy for race/religion)).

Examples that do not include group-specific treatment for the specified protected categories:

Laws that allow both religious and non-religious representatives to perform marital services do not have group-specific treatment based on religion

Laws outlining conditions for divorce do not have group-specific treatment based on marital status

Laws with degree requirements for employment do not have group-specific treatment based on socioeconomic status.

##### 1.3 Socioeconomic references

A provision references socioeconomic status if the law references a group based on their socioeconomic status. For example, a statute points out a person's wealth, education, or poverty status, even if those are not formally recognized as protected categories.

```

If no protected categories, proxies, or socioeconomic references are found, skip Step 2 and set all extracted fields to NULL.

Step 2 - Detect group-specific treatment

If Step 1 found any categories, check whether the provision assigns different rules, requirements, accommodations, or definitions to the identified group(s) or certain sub-groups within the identified group(s).

Examples of group-specific treatment:
Marriage eligibility limited to heterosexual couples
Accommodations or programs for individuals with disabilities
Employment criteria that excludes non-citizens
Defining classifications of children based on parental marital status

Output only a single JSON object using the following structure:

{
  "Reasoning": "Text explanation of why each classification was made.",
  "Categories": [
    {
      "Type": "Protected_category | Proxy | Socioeconomic_reference",
      "Value": "race | religion | sex/gender/sexual orientation | marital status | national origin/alienage/ancestry | genetic information | age (over 40) | disability/medical condition | socioeconomic reference",
      "Group_specific_treatment": true | false
    },
    {
      "Law_category": ["Specific legal topics covered in relation to the protected category, e.g. 'witness eligibility'"],
    }
  ],

  "Contemporary": true | false, // Whether the law covers a topic commonly discussed in the 21st century

```

```

"Repealed": true | false //Whether the law contains notes stating it has been repealed
}

```

Note: For the value field only respond with one of the categories listed above.

Examples:

```

2024 Code of Alabama
Title 12 - Courts.
Chapter 21 - Evidence and Witnesses.
Article 1 - General Provisions.
Division 2 - Witnesses.
Division 2 - Competency.
Section 12-21-165 - Incompetent Witnesses.

```

Section 12-21-165

Incompetent Witnesses.

(a) Persons who have not the use of reason, such as idiots, lunatics during lunacy and children who do not understand the nature of an oath, are incompetent witnesses.

(b) The court must, by examination, decide upon the capacity of one alleged to be incompetent from idiocy, lunacy, insanity, drunkenness or infancy.

(Code 1907, 4013, 4014; Code 1923, 7727, 7728; Code 1940, T. 7, 439, 440.)

```

{
  "Reasoning": "The provision excludes persons with mental disabilities ('idiots,' 'lunatics') who cannot understand the nature of an oath from serving as witnesses. This is a protected category (disability/medical condition). The exclusion is explicit group-specific treatment because it denies this group eligibility as witnesses. There are no proxies or socioeconomic references. The law relates to witness eligibility. The terminology suggests it is not contemporary. The text does not indicate repeal.",

```

```

"Categories": [
  {
    "Type": "Protected_category",
    "Value": "disability/medical condition",
    "Group_specific_treatment": true,
    "Law_category": ["witness eligibility"]
  }
],
,
"Contemporary": false,
"Repealed": false
}

```

2023 Code of Alabama  
 Title 16 - Education.  
 Chapter 1 - General Provisions.  
 Section 16-1-14 - Removal, Isolation, or  
 Separation of Pupils Creating Disciplinary  
 Problems; State Approval Necessary for  
 Rules Implementing Such Measures;  
 Deprivation of Right to Equal and Adequate  
 Education May Not Result.

Removal, isolation, or separation of pupils  
 creating disciplinary problems; state  
 approval necessary for rules implementing  
 such measures; deprivation of right to  
 equal and adequate education may not  
 result.

Any city, county, or other local public school  
 board shall, consistent with Section  
 16-28-12, prescribe rules and regulations  
 with respect to behavior and discipline of  
 pupils enrolled in the schools under its  
 jurisdiction and, in order to enforce such  
 rules and regulations, may remove,  
 isolate, or separate pupils who create  
 disciplinary problems in any classroom or  
 other school activity and whose presence  
 in the class may be detrimental to the  
 best interest and welfare of the pupils of  
 such class as a whole. Any rules and  
 regulations adopted pursuant to this  
 section shall be approved by the State  
 Board of Education. Any such removal,  
 isolation, or separation may not deprive  
 such pupils of their full right to an  
 equal and adequate education.

```

{
  "Reasoning": "The provision authorizes  

  school boards to establish discipline  

  rules and to remove or separate pupils  

  who create disciplinary problems. It  

  does not reference protected categories,  

  proxies, or socioeconomic groups.  

  Therefore, no group-specific treatment  

  under Step 2 applies. The law concerns  

  student discipline. It appears  

  contemporary and is not marked as  

  repealed.",
  "Categories": [],
  "Law_category": ["student discipline", "  

  school governance"],
  "Contemporary": true,
  "Repealed": false
}

```

## B.2 Instruction-Tuned Embedding Model Prompt

Represent the legal provision for clustering  
 based on the protected category and type  
 of differential treatment

## B.3 Cluster Labeling Prompt

You will be provided with a sample of legal  
 provision descriptions that belong to the  
 same cluster. Your task is to generate a  
 short, descriptive label for the cluster  
 that summarizes the protected categories  
 and types of group-specific treatment  
 present in the provisions. Focus only on  
 these aspects.

Instructions:

Identify all explicitly mentioned protected  
 categories (e.g., age, disability, gender,  
 marital status, race, religion, etc.).  
 Be as specific as possible (e.g., say  
 people with a communicable disease  
 instead of people with medical  
 conditions).

Identify the type and mechanism of group-  
 specific treatment (e.g., eligibility  
 criteria, exemptions, licensing  
 requirements).

Describe how and in what context these groups are treated differently (e.g., Establishes exemptions from regulations governing social gatherings for religious institutions, Differentiates employee insurance eligibility based on age)

Note if the treatment is an affirmative action measure or provides accommodations for disadvantaged groups

If possible, include a brief reference to the policy or regulatory domain (e.g., employment, education, transportation, housing, public facilities), only when clearly inferable from the text.

Ignore unrelated details such as historical context, repeal status, or legislative dates.

Write a concise, specific label (one short phrase or sentence) that integrates:

Who the provisions apply to

What kind of treatment they describe

The domain or context (if clearly stated in the text)

Example 1:

Provisions:

Denver Municipal Code Chapter 18 - EMPLOYEE AND OFFICER PAY AND BENEFITS [1] ARTICLE XI. - OLD-AGE AND SURVIVOR'S INSURANCE [8] Sec. 18-376. - Adoption of federal social security plan; exceptions to coverage.

The provision extends 'old-age and survivor's insurance,' which directly references the protected category of age (over 40). It provides benefits specifically to this group, which constitutes group-specific treatment. The provision also uses the gender-specific term 'firemen' when listing exceptions to coverage ('firemen's pension fund'). This use of gender-specific language references the protected category of sex/gender. The exclusion of this group from the main social security plan is a form of group-specific treatment. The law concerns employee benefits. The references to 1953 and the Code of 1950 indicate it is not contemporary. The text does not indicate it has been repealed.

Denver Municipal Code Chapter 18 - EMPLOYEE AND OFFICER PAY AND BENEFITS [1] ARTICLE XI. - OLD-AGE AND SURVIVOR'S INSURANCE [8] Sec. 18-377. - Agreements with state department of employment security.

The provision authorizes the city to create a plan for extending the benefits of the 'federal old-age and survivor's insurance system' to its employees. This directly references 'old-age', which is a protected category for individuals over 40. The system is designed to provide benefits specifically to this age group, which constitutes group-specific treatment. The term 'survivor's insurance' also implicates marital status, as benefits are often provided to surviving spouses. Providing benefits to survivors is a form of group-specific treatment based on marital status. The law's reference to 1951 and 1953 dates it as not contemporary. The text does not indicate it has been repealed.

Denver Municipal Code Chapter 18 - EMPLOYEE AND OFFICER PAY AND BENEFITS [1] ARTICLE XI. - OLD-AGE AND SURVIVOR'S INSURANCE [8] Sec. 18-379. - Payment by city to contribution fund.

The provision's title explicitly mentions 'OLD-AGE AND SURVIVOR'S INSURANCE,' which directly references the protected category of age. The law authorizes payments into a contribution fund for a federal program (Federal Insurance Contributions Act) that provides benefits specifically to individuals upon reaching a certain age (retirement). This system, by design, treats a specific age group differently, which constitutes group-specific treatment. The law covers employee benefits and municipal finance. The reference to 1950s law and code indicates it is not contemporary. There is no indication that the law has been repealed.

Label:

Differentiates eligibility, funding, or benefits for employee insurance based on the employees age, gender, or marital status

Example 2:

Provisions:

Denver Municipal Code Chapter 11 - CHILD CARE ESTABLISHMENTS [1] ARTICLE IV. - DENVER COLLEGE AFFORDABILITY FUND [2] Sec. 11-31. - Expenditures of Denver college affordability tax to be made by contract with non-profit corporation.

The provision outlines the administration of the Denver College Affordability Fund. Subsection (b)(6) contains a non-discrimination clause that explicitly references numerous protected categories: race, religion, national origin, gender, military status, sexual orientation, marital status, and disability. For these categories, the law prohibits different treatment, so there is no group-specific treatment. The same clause references age but includes an exception allowing for age-based eligibility criteria for students, which constitutes group-specific treatment. The entire purpose of the Denver College Affordability Fund is to provide financial assistance to students, which is a reference to a socioeconomic group (students needing financial aid). The program is specifically designed for this group, which is a form of group-specific treatment.

Denver Municipal Code Chapter 11 - CHILD CARE ESTABLISHMENTS [1] ARTICLE III. - DENVER PRESCHOOL PROGRAM Sec. 11-22. - Program requirements to be included in contract.

The provision contains an explicit non-discrimination clause in section (h) that directly lists numerous protected categories: race, color, religion, national origin, gender, age, military status, sexual orientation, gender variance, marital status, and physical or mental disability. This clause prohibits group-specific treatment for these groups in the context of the preschool program. Section (h) also references 'protective hairstyle,' which is a proxy for race, and similarly prohibits discrimination on this basis. Furthermore, section (d)(2) references socioeconomic status by requiring that tuition credits be administered on a sliding scale based on family income. This is group-specific treatment because the amount of the benefit is directly determined by a family's socioeconomic status. The ordinance dates and terminology confirm it is contemporary, and there is no indication it has been repealed.

Label:

Establishes non-discrimination clauses in educational systems for multiple protected categories with special protections or exemptions for age and socioeconomic status

#### B.4 Discriminatory Attribute Detection Prompt

Your task is to classify each legal provision using the criteria listed below.

Criteria:

Explicit Facial Discrimination (The provision explicitly names a protected or identifiable group and imposes a ban, exclusion, or denial of fundamental rights )

No Remedial or Sovereignty Justification

Not an Accommodation or Affirmative Action Measure  
Unjustified Classification (Provision excludes , burdens, or privileges a group, or otherwise applies the law unequally, in a way not necessary to the legitimate function or task being regulated)  
Demeaning or Historically Oppressive Framing ( Provision uses demeaning terminology or mirror historically oppressive or exclusionary legal regimes (e.g., Jim Crow , apartheid, antimiscegenation laws)

Output only a single JSON object using the following structure:

```
{
  "Criteria": {
    "Explicit_Facial_Discrimination": "Yes|No",
    "No_Remedial_or_Sovereignty_Justification": "Yes|No",
    "Not_Accommodation_or_Affirmative_Action": "Yes|No",
    "Unjustified_Classification": "Yes|No",
    "Demeaning_or_Historically_Oppressive": "Yes|No"
  },
  "Reasoning": "Concise explanation justifying each Yes/No determination."
}
```

Examples:

551.4 Incapacity; persons woman prohibited from marrying. Sec. 4.

A woman shall not marry her father, brother, grandfather, son, grandson, stepfather, grandmother's husband, daughter's husband, granddaughter's husband, husband's father, husband's grandfather, husband's son, husband's grandson, brother's son, sister's son, father's brother, mother's brother, or cousin of the first degree, or another woman.

```
{
  "Criteria": {
    "Explicit_Facial_Discrimination": "Yes",
    "No_Remedial_or_Sovereignty_Justification": "Yes",

```

```
    "Not_Accommodation_or_Affirmative_Action": "Yes",
    "Unjustified_Classification": "Yes",
    "Demeaning_or_Historically_Oppressive": "Yes"
  },
  "Reasoning": "The provision explicitly singles out women and imposes categorical prohibitions on whom they may marry, including an express ban on marrying another woman, which constitutes facial discrimination affecting a fundamental right. It is not framed as a remedial or sovereignty-based measure, nor as an accommodation or affirmative action policy. The categorical exclusion of same-sex marriage and the gender-specific framing go beyond what is necessary to regulate marriage and reflect historically oppressive legal regimes, including bans on same-sex marriage and other status-based marriage restrictions."
}
```

2023 Code of Alabama  
Title 16 - Education.  
Chapter 1 - General Provisions.  
Section 16-1-14 - Removal, Isolation, or Separation of Pupils Creating Disciplinary Problems; State Approval Necessary for Rules Implementing Such Measures; Deprivation of Right to Equal and Adequate Education May Not Result.

Removal, isolation, or separation of pupils creating disciplinary problems; state approval necessary for rules implementing such measures; deprivation of right to equal and adequate education may not result.

Any city, county, or other local public school board shall, consistent with Section 16-28-12, prescribe rules and regulations with respect to behavior and discipline of pupils enrolled in the schools under its jurisdiction and, in order to enforce such rules and regulations, may remove, isolate, or separate pupils who create disciplinary problems in any classroom or other school activity and whose presence in the class may be detrimental to the best interest and welfare of the pupils of such class as a whole. Any rules and regulations adopted pursuant to this section shall be approved by the State Board of Education. Any such removal, isolation, or separation may not deprive such pupils of their full right to an equal and adequate education.

```
{
  "Criteria": {
    "Explicit_Facial_Discrimination": "No",
    "No_Remedial_or_Sovereignty_Justification": "Yes",
    "Not_Accommodation_or_Affirmative_Action": "Yes",
    "Unjustified_Classification": "No",
    "Demeaning_or_Historically_Oppressive": "No"
  },
  "Reasoning": "The provision does not name a protected or identifiable group, but instead applies to pupils based on conduct (creating disciplinary problems), so it is not facially discriminatory. It is framed as a regulatory measure aimed at maintaining classroom order and welfare, with state oversight and explicit safeguards, which constitutes a legitimate justification rather than an unjustified exclusion. It is not framed as a remedial or sovereignty-based measure. It is not an accommodation or affirmative action measure. The exclusionary actions permitted are conditional, discretionary, and explicitly constrained by the requirement that pupils retain their right to an equal and adequate education, indicating tailoring to the task rather than unnecessary exclusion. The language is neutral and does not use demeaning terms or mirror historically oppressive legal regimes."
}
```

```
}
```

## B.5 Soft Legality Rating Prompt (1)

Analyze the following legal provision to determine whether it would be found unconstitutional under the Fourteenth Amendment to the U.S. Constitution and/or whether it would violate applicable federal anti-discrimination law. Base your reasoning strictly on the text of the provision and established constitutional or statutory doctrine. Do not assume unstated facts or conditions.

For constitutional analysis, only consider the Fourteenth Amendment (do not analyze the provision under any other constitutional amendments).

In your analysis, address three key dimensions:

Precedential Strength (Supreme Court / Federal Authority):

- 1a. How strong and authoritative is the controlling precedent governing identical or highly similar treatment under either: the Equal Protection or Due Process Clauses of the Fourteenth Amendment, or applicable federal anti-discrimination statutes? (If there is no precedent, respond None).
- 1b. Consider whether controlling Supreme Court or federal precedent directly addresses this type of law or treatment. If so, list specific cases, statutes etc. (If there is none identical or highly similar to this treatment or this treatment is not related to Equal Protection/federal anti-discrimination law, respond None.)

Note: If your responses were None to section 1 a. and 1b, respond None to steps 2-3, and likely satisfies for "Overall judgment."

Level of Scrutiny Likely Applied:

Identify what level of judicial scrutiny would most likely apply under Equal Protection or Due Process.

Use these general rules:

Strict Scrutiny: race, religion, national origin, some alienage

Intermediate Scrutiny: gender or legitimacy.

Rational Basis (with or without bite): disability, age, economic regulation, or neutral classifications.

Accepted Contexts:

3a. Identify whether courts have recognized any accepted contexts where such a distinction is considered constitutionally or statutorily permissible (e.g., alienage in political-function roles). Are there exceptions that would allow the treatment in this provision to be constitutional?

3b. Does the provision meet these exceptions? Is there even a slight chance that one could reasonably infer (even if not directly stated in the text) that the provision may meet these exceptions?

Note: For all steps if there are multiple treatments/categories, analyze all of them. If any treatment is unconstitutional, consider the entire provision as unconstitutional.

Output only a single JSON object using this structure:

```
{
  "Reasoning": "Text explanation for each dimension and overall conclusion. Write 2-3 sentences per section 1a, 1b, 2, 3a, 3b, and overall conclusion"
  Overall judgement: Violates |
                    Ambiguous | Likely satisfies
}
```

Examples:

The following provision references the following protected categories: sex/gender /sexual orientation -some or all which may experience differential treatment. Analyze the provision using the instructions above:

RS 9:272 - Covenant marriage; intent; conditions to create

A. A covenant marriage is a marriage entered into by one male and one female who understand and agree that the marriage between them is a lifelong relationship. Parties to a covenant marriage have received counseling emphasizing the nature and purposes of marriage and the responsibilities thereto. Only when there has been a complete and total breach of the marital covenant commitment may the non-breaching party seek a declaration that the marriage is no longer legally recognized.

```
{
  "Reasoning": {
    "1a": "There is strong Supreme Court precedent addressing laws that differentiate based on sex and gender. Cases like United States v. Virginia (1996) and Craig v. Boren (1976) establish that sex-based classifications are subject to intermediate scrutiny, which requires the government to show that the classification serves important governmental objectives and is substantially related to achieving those objectives.",
    "1b": "Relevant cases include United States v. Virginia (1996), Craig v. Boren (1976), and other sex discrimination cases under the Equal Protection Clause. Obergefell v. Hodges (2015) addresses marriage equality and affirms that same-sex couples cannot be denied the right to marry. Therefore, there is substantial precedent indicating that limiting marriage to one male and one female would likely be unconstitutional.",
  }
}
```

"2": "Sex/gender is a classification triggering intermediate scrutiny under the Equal Protection Clause. Laws limiting access to legal marriage based on sex must be substantially related to an important government interest. Differential treatment based on sex in marriage is tightly constrained by current Supreme Court jurisprudence.",

"3a": "There is no widely recognized exception allowing the restriction of marriage to opposite-sex couples while completely excluding same-sex couples from a comparable legal institution .",

"3b": "The covenant marriage provision does not meet any recognized exception . By explicitly restricting covenant marriage to one male and one female, it excludes same-sex couples entirely and imposes a sex-based limitation that cannot be justified by the neutral governmental objectives exceptions courts have upheld in other contexts.",

"Overall conclusion": "This provision imposes a sex-based restriction on access to a legal institution and would be subject to intermediate scrutiny. Given current constitutional doctrine, including Obergefell v. Hodges, this law would likely fail to meet the standard required for permissible sex-based distinctions in marriage and is therefore unconstitutional under the Fourteenth Amendment."

},

"Overall judgement": "Violates"

}

The following provision references the following protected categories: disability/ medical condition -some or all which may experience differential treatment. Analyze the provision using the instructions above:

Section 12-21-165

Incompetent Witnesses.

(a) Persons who have not the use of reason, such as idiots, lunatics during lunacy and children who do not understand the nature of an oath, are incompetent witnesses.

(b) The court must, by examination, decide upon the capacity of one alleged to be incompetent from idiocy, lunacy, insanity, drunkenness or infancy.

(Code 1907, 4013 , 4014; Code 1923, 7727 , 7728; Code 1940, T. 7, 439 , 440.)

{

"Reasoning": {

"1 a : Supreme Court precedent strongly establishes that laws categorically excluding people with intellectual or mental disabilities from fundamental rights based on outdated stereotypes violate equal protection. *Cleburne v. Cleburne Living Center* (1985) held that disability classifications based on prejudice rather than actual inability fail constitutional review. While *Cleburne* addressed zoning, its reasoning applies to categorical exclusions based on disability status. The archaic terms 'idiots' and 'lunatics' reflect the exact type of stigmatizing classifications the Court condemned. ,

1b : *Cleburne v. Cleburne Living Center*, 473 U.S. 432 (1985) is the primary controlling authority establishing that disability-based classifications cannot rest on prejudice or archaic assumptions. While this case does not directly address witness competency, it establishes that categorical disability-based exclusions using stigmatizing labels like 'idiot' fail constitutional scrutiny when they do not account for individual capability.

2 : Rational basis review applies to disability classifications under Cleburne. However, courts apply 'rational basis with bite' when disability classifications reflect prejudice rather than legitimate differences in actual capability. A categorical ban on testimony by 'idiots' and 'lunatics' without individualized assessment of actual competence appears to rest on archaic stereotypes rather than legitimate assessment of ability to understand and give truthful testimony. The provision does include subsection (b) requiring individual examination, which may save it from being a purely categorical exclusion. ,

3a : Courts have recognized that witness competency requirements serve legitimate state interests in ensuring reliable testimony and protecting the integrity of judicial proceedings. States may establish competency standards based on ability to perceive, remember, and communicate truthfully, and to understand the nature of an oath. However, these standards must be based on individual assessment of actual capability, not categorical exclusions based on diagnostic labels or disability status alone. ,

3b : The provision partially meets recognized exceptions because subsection (b) requires individualized examination by the court to 'decide upon the capacity' of persons alleged to be incompetent. This suggests the provision doesn't create an absolute categorical ban but rather establishes a presumption subject to rebuttal through judicial assessment. However, subsection (a) uses stigmatizing and archaic terminology ('idiots,' 'lunatics') that reflects outdated stereotypes rather than functional capacity. A reasonable inference might be that courts would interpret this provision to require individualized functional assessment of testimonial capacity, which could be constitutional, though the offensive terminology itself remains problematic. ,

Overall conclusion : This provision is ambiguous in its constitutionality. While the use of archaic, stigmatizing terms like 'idiots' and 'lunatics' reflects the prejudice-based classifications condemned in Cleburne, subsection (b) requires individualized judicial examination of actual capacity rather than categorical exclusion based on disability status alone . If interpreted and applied to require functional assessment of whether a person can perceive, remember, communicate, and understand oath obligations, rather than automatic exclusion based on diagnostic labels, the provision might survive rational basis review. However, the offensive terminology and apparent focus on disability status rather than functional testimonial capacity creates serious constitutional doubt under Cleburne's framework against stereotype-based classifications."

},

"Overall judgement": "Ambiguous"

}

## B.6 Soft Legality Rating Prompt (2)

Analyze the following legal provision to determine whether it would be found unconstitutional under the Fourteenth Amendment to the U.S. Constitution and/or whether it would violate applicable federal anti-discrimination law. Base your reasoning strictly on the text of the provision and established constitutional or statutory doctrine do not assume unstated facts or conditions.

For constitutional analysis, only consider the Fourteenth Amendment (do not analyze the provision under any other constitutional amendments).

In your analysis, address three key dimensions :

Precedential Strength (Supreme Court / Federal Authority):

1a. How strong and authoritative is the controlling precedent governing identical or highly similar treatment under either: the Equal Protection or Due Process Clauses of the Fourteenth Amendment, or applicable federal anti-discrimination statutes? (If there is no precedent, respond None).

1b. Consider whether controlling Supreme Court or federal precedent directly addresses this type of law or treatment. If so, list specific cases, statutes etc. (If there is none identical or highly similar to this treatment or this treatment is not related to Equal Protection/federal anti-discrimination law, respond None.)

Note: If your responses were None to section 1 a. And 1b, respond None to steps 2-3, a likely satisfies for "Overall judgement."

Level of Scrutiny Likely Applied:

Identify what level of judicial scrutiny would most likely apply under Equal Protection or Due Process.

Use these general rules:

Strict Scrutiny: race, religion, national origin, some alienage

Intermediate Scrutiny: gender or legitimacy.

Rational Basis (with or without bite): disability, age, economic regulation, or neutral classifications.

Accepted Contexts:

3a. Identify whether courts have recognized any accepted contexts where such a distinction is considered constitutionally or statutorily permissible (e.g., alienage in political-function roles). Are there exceptions that would allow the treatment in this provision to be constitutional?

3b. Does the provision meet these exceptions? Is there even a slight chance that one could reasonably infer (even if not directly stated in the text) that the provision may meet these exceptions?

Note: For all steps if there are multiple treatments/categories, analyze all of them. If any treatment is unconstitutional, consider the entire provision as unconstitutional.

Output only a single JSON object using this structure:

```
{
  "Reasoning": "Text explanation for each
  dimension and overall conclusion. Write
  2-3 sentences per section 1a, 1b, 2, 3a,
  3b, and overall conclusion"
  Overall judgement: Violates |
  Ambiguous | Likely satisfies
}
```

Examples:

The following provision references the following protected categories: sex/gender /sexual orientation -some or all which may experience differential treatment. Analyze the provision using the instructions above:

RS 9:272 - Covenant marriage; intent; conditions to create

A. A covenant marriage is a marriage entered into by one male and one female who understand and agree that the marriage between them is a lifelong relationship. Parties to a covenant marriage have received counseling emphasizing the nature and purposes of marriage and the responsibilities thereto. Only when there has been a complete and total breach of the marital covenant commitment may the non-breaching party seek a declaration that the marriage is no longer legally recognized.

```
{
  "Reasoning": {
```

"1a": "There is strong Supreme Court precedent addressing laws that differentiate based on sex and gender. Cases like United States v. Virginia (1996) and Craig v. Boren (1976) establish that sex-based classifications are subject to intermediate scrutiny, which requires the government to show that the classification serves important governmental objectives and is substantially related to achieving those objectives.",

"1b": "Relevant cases include United States v. Virginia (1996), Craig v. Boren (1976), and other sex discrimination cases under the Equal Protection Clause. Obergefell v. Hodges (2015) addresses marriage equality and affirms that same-sex couples cannot be denied the right to marry. Therefore, there is substantial precedent indicating that limiting marriage to one male and one female would likely be unconstitutional.",

"2": "Sex/gender is a classification triggering intermediate scrutiny under the Equal Protection Clause. Laws limiting access to legal marriage based on sex must be substantially related to an important government interest. Differential treatment based on sex in marriage is tightly constrained by current Supreme Court jurisprudence.",

"3a": "There is no widely recognized exception allowing the restriction of marriage to opposite-sex couples while completely excluding same-sex couples from a comparable legal institution .",

"3b": "The covenant marriage provision does not meet any recognized exception . By explicitly restricting covenant marriage to one male and one female, it excludes same-sex couples entirely and imposes a sex-based limitation that cannot be justified by the neutral governmental objectives exceptions courts have upheld in other contexts.",

"Overall conclusion": "This provision imposes a sex-based restriction on access to a legal institution and would be subject to intermediate scrutiny. Given current constitutional doctrine, including Obergefell v. Hodges, this law would likely fail to meet the standard required for permissible sex-based distinctions in marriage and is therefore unconstitutional under the Fourteenth Amendment."

},

"Overall judgement": "Violates"

}

The following provision references the following protected categories:disability/ medical condition -some or all which may experience differential treatment. Analyze the provision using the instructions above:

Section 12-21-165

Incompetent Witnesses.

(a) Persons who have not the use of reason, such as idiots, lunatics during lunacy and children who do not understand the nature of an oath, are incompetent witnesses.

(b) The court must, by examination, decide upon the capacity of one alleged to be incompetent from idiocy, lunacy, insanity, drunkenness or infancy.

(Code 1907, 4013 , 4014; Code 1923, 7727 , 7728; Code 1940, T. 7, 439 , 440.)

{

"Reasoning": {

"1a": "There is some Supreme Court precedent regarding the treatment of individuals with disabilities in legal proceedings, but no directly controlling case addresses the categorical exclusion of certain persons as witnesses based on mental capacity. Cases on competence to stand trial or provide testimony touch on related concepts, but the precise rule in this provision is not directly addressed at the Supreme Court level. Therefore, precedential strength is limited.",

"1b": "Relevant cases include *Drope v. Missouri* (1975), which addresses competency in legal proceedings, and *Faretta v. California* (1975) in the context of understanding legal processes, but neither directly endorses or rejects a blanket rule excluding witnesses based on mental condition. There is no identical Supreme Court precedent directly addressing this type of categorical witness competency exclusion. Consequently, there is no directly applicable federal authority.",

"2": "Disability or mental incapacity classifications are typically reviewed under rational basis review in the context of legal procedures. The classification here (excluding 'idiots, lunatics, or children who do not understand the nature of an oath') is arguably rationally related to the important governmental interest of ensuring reliable testimony in court, which may satisfy rational basis scrutiny.",

"3a": "Courts have recognized that individuals must have minimal mental capacity to understand and testify under oath; competency rules are generally permissible to ensure the integrity of judicial proceedings. This is an accepted context in which differential treatment based on mental capacity is constitutionally permissible.",

"3b": "The provision appears to fit within this exception, as it requires examination by the court to determine capacity, rather than imposing an automatic exclusion in all cases. While the language uses archaic terms, the text allows for judicial evaluation of competence, meaning there is a reasonable basis for the law to be constitutional under the rational basis standard.",

"Overall conclusion": "Although the provision references disability/mental condition and uses outdated and offensive terminology, it regulates witness competency in a neutral way aimed at ensuring reliable testimony. Under current constitutional doctrine, the law likely satisfies the Fourteenth Amendments requirements because it is rationally related to a legitimate governmental interest and provides judicial evaluation rather than absolute exclusion."

},

"Overall judgement": "Likely satisfies"

}

## B.7 Offensive Language Prompt (Post-RegEx)

You are labeling phrases extracted from U.S. municipal codes.

For each phrase, determine if it is or could plausibly signal:

- Derogatory, stigmatizing, or exclusionary language toward any group
- Historically discriminatory terminology (even if no longer commonly recognized)
- Language targeting protected classes (race, religion, disability, gender, etc.)

Return ONLY a JSON array containing the EXACT phrases that ARE potentially discriminatory.

Example:

Input: {"phrases": ["city hall", "negro", "parking lot", "crippled"]}

Output: ["negro", "crippled"]

IMPORTANT: Return the exact phrase strings, not indices. Copy them exactly as they appear in the input.

Be inclusive: when in doubt, include the phrase (recall over precision).

If NO phrases are discriminatory, return an empty array: []

## B.8 Offensive Language Prompt (Provisions with protected category references)

You are a legal research expert. Your task is to classify each legal provision by identifying whether it contains derogatory language. Carefully consider the definitions and instructions below.

Definitions:

### 1. Derogatory Language

A provision contains derogatory language if it uses outdated or offensive terms for a group of individuals.

This includes words and phrases historically used to oppress, stigmatize, discriminate, or marginalize individuals.

It also includes words and phrases that describe offensive or oppressive practices (e.g., 'minstrel shows', 'freaks exhibition')

Do not include provisions that are discriminatory if they do not use offensive terms.

Do not include 'alien', 'homosexuality', or 'handicapped' as a derogatory term for this task.

Example terms: 'insane', 'colored' (if referring to a person), 'retarded', 'idiot', 'pauper', 'oriental', 'womens chastity', 'minstrel shows', 'Indian blood', 'Gypsies', 'deformed', 'cripple', 'vagrant', 'Negro'

Examples:

Positive (contains derogatory language): No insane person shall inherit property.

Negative (does not contain derogatory language): Persons with mental disabilities may not inherit property.

Instructions:

Analyze each legal provision carefully.

Determine whether it contains derogatory language based on the definitions above.

Provide reasoning for each classification.

Output Format:

Output only a single JSON object using this structure:

```
{
  "Reasoning": "Text explanation of why each
  classification was made.",
  Derogatory_terms: True|False
}
```

Example:

2024 Code of Alabama  
Title 6 - Civil Practice.  
Chapter 5 - Actions.  
Article 19 - Seduction.  
Section 6-5-351 - Action by Father or Mother  
for Seduction of Daughter.

Section 6-5-351  
Action by Father or Mother for Seduction of  
Daughter.

```
The father or, in case of his death or
desertion of his family, or of his
imprisonment for a term of two years or
more under a conviction for crime, or of
his confinement in an insane hospital, or
of his having been declared of unsound
mind, the mother, may commence an action
for the seduction of a daughter under the
age of 19 years though she be not living
with or in the service of the plaintiff at
the time of the seduction or afterwards
and there is no loss of service; provided,
that an action by the daughter is a bar
to an action by the father or mother.
```

```
{
  "Reasoning": "The provision contains
    derogatory language because it uses
    outdated terms like 'insane hospital'
    and 'of unsound mind', which
    historically stigmatize individuals with
    mental illness.",
  "Derogatory_terms": True
}
```

```
2024 Minnesota Statutes
Chapters 540 - 552      Civil Procedure
Chapter 540           Parties To Actions
Section 540.07        Parent Or Guardian May Sue
                      For Seduction.
540.07 PARENT OR GUARDIAN MAY SUE FOR
SEDUCTION.
```

```
A parent may maintain an action for the
seduction of the son or daughter, and the
guardian may maintain an action for the
seduction of the ward, though the son or
daughter or ward is not living with, or in
the service of, the plaintiff at the time
of the seduction or afterwards, and there
is no loss of service.
```

```
{
  "Reasoning": "The provision does not contain
    derogatory language as it does not use
    any outdated or offensive terms to
    describe a group of individuals.",
  "Derogatory_terms": False
}
```

### C TF-IDF logistic regression classifier

To filter out laws unlikely to reference protected categories, we use a TF-IDF logistic regression classifier. We split a sample of

Gemini-labeled provisions into 90% training (1.7M rows) and 10% held-out validation (190K rows), preserving the class balance of the full dataset. We set the classification threshold to achieve 99.5% recall on the held-out set. At this threshold, the classifier removes 58% of negatives (provisions not referencing protected categories), reducing the corpus requiring LLM labeling by approximately 17% overall. We manually review a sample of the small number of false negatives and find them to be insubstantial losses, with many provisions referencing protected categories only in the context of repeal notes or historical annotations rather than operative legal text.

### D Validation Set from External Sources

We define our validation for this first step as the following:

**True Positives:** The provision references a protected category and has differential treatment based on this category. The model correctly identifies the protected category reference and positively identifies the differential treatment.

**False Positives:** The provision does not reference a protected category or provision references a protected category but there is no differential treatment. The model incorrectly identifies a protected category or falsely identifies differential treatment.

**True Negatives:** The provision does not reference a protected category or provision references a protected category but there is no differential treatment. The model does not identify the protected category reference or identifies the difference but does not identify differential treatment.

**False Negatives:** The provision references a protected category and has differential treatment based on this category. The model does not identify the protected category reference or identifies the protected category reference but does not identify differential treatment.

Note, even if there are multiple references of the same protected category throughout a given provision, we only count this as one reference for that protected category. However, there may be multiple references of different protected categories within the same provision. We also include proxies within the protected category references.

**Table 1: Proportion of laws referencing each protected category in validation set of 259 externally-sourced provisions.**

Protected Category	Proportion (%)
Race	34.4
Religion	3.1
Sex/Gender/Sexual Orientation	31.3
Marital Status	10.0
National Origin/Alienage/Ancestry	13.5
Genetic Information	0.8
Age (Over 40)	4.6
Disability/Medical Condition	35.1

**Table 2: Provision Ranking Based on LLM Legal Assessments**

Priority Level	Model Agreement Criteria	Human Annotation Criteria
Priority 1 (High)	Both models conclude that the provision violates the Fourteenth Amendment and/or federal anti-discrimination law.	Laws for which unconstitutionality or violation of federal anti-discrimination law can be determined almost definitively based on clearly established statutory or constitutional doctrine.
Priority 2 (Medium)	Both models classify the provision as ambiguous or each model outputs a different classification than the other.	Provisions for which constitutional validity or compliance with anti-federal law cannot be determined with confidence. This category includes provisions that are vague, internally inconsistent, or for which relevant judicial or statutory precedent is limited, conflicting, or nonexistent.
Priority 3 (Low)	Both models conclude that the provision likely satisfies the Fourteenth Amendment and federal anti-discrimination law.	Provisions that appear facially constitutional and are supported by existing judicial precedent or well-established legal doctrine.

Row	Text / Rationale	Verdict
	<b>Provision text.</b> That it shall be the duty of said board to establish a graded school for the white children and one for the colored children of said town, and to appropriate the funds derived from said special tax and all other sources for said graded schools for white and colored children so as to equalize school facilities between the two races.	
Gemini (run 1)	<p><b>1a:</b> The constitutional precedent governing state-mandated racial segregation in public education is exceptionally strong and authoritative. The Supreme Court has unequivocally held that such laws violate the Equal Protection Clause of the Fourteenth Amendment. This is one of the most settled areas of American constitutional law.</p> <p><b>1b:</b> The controlling Supreme Court precedent is <i>Brown v. Board of Education of Topeka</i>, 347 U.S. 483 (1954). This landmark case directly overturned the “separate but equal” doctrine in public education, holding that separate educational facilities are inherently unequal. The provision, which mandates separate schools for “white” and “colored” children, is a direct example of the type of law <i>Brown</i> declared unconstitutional.</p> <p><b>2:</b> Any law that establishes a classification based on race is subject to Strict Scrutiny. Under this standard, the government must prove that the law is narrowly tailored to achieve a compelling governmental interest. A state interest in maintaining racial segregation is not a compelling interest; in fact, it is an illegitimate one.</p> <p><b>3a:</b> There are no accepted contexts where state-mandated racial segregation in public education is constitutionally permissible. The doctrine that once permitted this, known as “separate but equal” from <i>Plessy v. Ferguson</i> (1896), was explicitly rejected in the field of public education by <i>Brown v. Board of Education</i>. Subsequent jurisprudence has consistently affirmed that racial segregation by the state is unconstitutional.</p> <p><b>3b:</b> This provision does not meet any recognized exception because no such exception exists. The provision’s own language mandating the establishment of separate schools for white and colored children is the very essence of the constitutional violation. The clause aiming “to equalize school facilities” is irrelevant, as the Court in <i>Brown</i> held that the segregation itself, not the material equality of the facilities, is what causes the constitutional harm.</p> <p><b>Overall conclusion:</b> This provision explicitly mandates racial segregation in public schools, a practice the Supreme Court has declared unconstitutional. Based on the controlling precedent of <i>Brown v. Board of Education</i>, this law is subject to strict scrutiny and fails because there is no compelling government interest in segregation. The provision is a clear and direct violation of the Equal Protection Clause of the Fourteenth Amendment.</p>	VIOLATES
Gemini (run 2)	<p><b>1a:</b> The constitutional precedent governing state-mandated racial segregation is exceptionally strong and authoritative. The Supreme Court has unequivocally held that laws establishing separate public facilities based on race are unconstitutional. The doctrine of “separate but equal,” which this provision attempts to codify, was explicitly rejected as inherently unequal.</p> <p><b>1b:</b> The controlling Supreme Court precedent is <i>Brown v. Board of Education of Topeka</i>, 347 U.S. 483 (1954). This landmark case directly held that state-mandated segregation of public schools based on race violates the Equal Protection Clause of the Fourteenth Amendment. Subsequent cases, such as <i>Loving v. Virginia</i> (1967), have reinforced the principle that racial classifications created by the state are presumptively invalid.</p> <p><b>2:</b> Strict scrutiny would apply to this provision. Any law that creates a classification based on race is subject to the most exacting level of judicial review. Under strict scrutiny, the government must demonstrate that the law is narrowly tailored to achieve a compelling government interest, a standard that state-mandated racial segregation in education cannot meet.</p> <p><b>3a:</b> There are no recognized contexts where state-mandated racial segregation in public education is considered constitutionally permissible. The Supreme Court has made clear that creating separate public institutions based on race is a per se violation of the Equal Protection Clause. While the court has addressed race-conscious policies in other limited contexts (e.g., affirmative action), de jure segregation is never permissible.</p> <p><b>3b:</b> This provision fails to meet any exception because no such exception exists for this type of racial classification. The provision’s text explicitly mandates the establishment of separate schools for “white children” and “colored children.” Its attempt to “equalize school facilities” directly invokes the “separate but equal” doctrine that was found to be inherently unequal and unconstitutional in <i>Brown v. Board of Education</i>.</p> <p><b>Overall conclusion:</b> This provision is a textbook example of de jure racial segregation explicitly declared unconstitutional by the Supreme Court. By mandating separate schools based on race, it creates a racial classification that is subject to strict scrutiny. Based on the controlling precedent of <i>Brown v. Board of Education</i>, the law fails to serve any compelling government interest and is a clear violation of the Fourteenth Amendment’s Equal Protection Clause.</p>	VIOLATES
Human labels	ANNOTATION_Reference_Exists: Yes; ANNOTATION_Same_holding: Yes; ANNOTATION_Verdict_Agreement: Agree	

**Table 3: Example annotation: racial segregation in public education.**

## E Clustering Motivation and Criteria

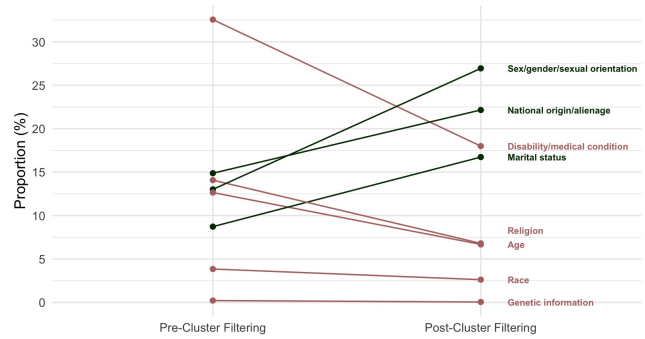
### Motivation

We embed the LLM generated reasoning rather than the statutory text of the provisions because discriminatory clauses in municipal law are often short segments embedded within longer legal texts covering multiple topics. As a result, clustering on the statutory text (and more conventional text embeddings) results in grouping laws by their shared topics or language rather than the specific treatment they impose. Additionally, because municipalities adopt legal language from other jurisdictions, we observe cases where laws across cities are nearly identical except for specific offensive or discriminatory language that one or more cities removed. Conventional text embedding approaches would likely group these laws

together due to their high lexical similarity. In preliminary analyses, we also clustered the LLM-generated reasoning using open source models such as Llama models and LegalBERT. We found that the instruction-tuned embedding model used in our approach yielded clusters that were better aligned with the legal treatment in the statutes, motivating our use of it.

### Filtering Criteria

We remove laws highly unlikely to be discriminatory, such as disability accommodations laws. We exclude laws that concern accommodations, including provisions related to individuals with disabilities, older adults, or those with medical conditions. Affirmative action provisions are also considered outside the scope of this project. In addition, we exclude clusters composed solely of provisions in which the only distinguishing feature is the use of gendered pronouns or other gendered language that does not establish a legal classification based on gender and may reasonably be interpreted as gender-neutral. Finally, we exclude anti-discrimination statutes, fair housing laws, provisions governing religious accommodations, laws pertaining to accessibility building standards, and religious-exceptions for smoking regulations.



**Figure 1: Proportion of Laws by Protected Category Reference Before and After Cluster Filtering.** Green indicates categories whose proportion in the data increased after clustering, while red indicates categories whose proportion decreased.

**Table 4: Cluster Labels and Inclusion Status. Shown below are a sample of cluster labels generated through our method, along with whether each cluster was included for further analysis.**

Cluster Label	Included
Establishes differentiated eligibility, documentation, and application requirements for various licenses, permits, and identification based on national origin (including citizenship, alienage, and residency), disability, medical condition, marital status, or socioeconomic status.	Yes
Imposes group-specific definitions, prohibitions, and regulations in public order, criminal, and nuisance law related to sexual conduct and public morality, targeting individuals based on sex/gender, sexual orientation, marital status and socioeconomic status.	Yes
Establishes processes for individuals with disabilities to receive reasonable accommodations, modifications, and exceptions in housing, land use, and zoning regulations to ensure equal opportunity and accessibility.	No
Prohibits discrimination in real estate and housing transactions, financing, and services based on race, religion, national origin, sex (including gender identity and sexual orientation), age, marital status, familial status, disability, socioeconomic status, and related characteristics like height, weight, or geographic location.	No

## F Discriminatory Attribute Classification

		LLM Prediction		
		LOW	MED	HIGH
Human	LOW	27 (8.8%)	34 (11.1%)	66 (21.6%)
	MED	11 (3.6%)	27 (8.8%)	38 (12.4%)
	HIGH	4 (1.3%)	8 (2.6%)	91 (29.7%)

**Figure 2: Priority validation results comparing human annotations to LLM predictions (n=306). Cell intensity reflects count.**

**Table 5: Annotation Agreement: Discriminatory Attribute Classification**

Attributes	Proportion of Classifications Agreed With
Facial Discrimination	90.2%
Unjustified Classification	92.2%
Affirmative Action	96.1%
Remedial or Sovereignty Justification	92.2%
Demeaning or Historically Oppressive	84.3%

## G RegEx and LLMs for Offensive Term Identification

Using a LLM enables the identification of stigmatizing language that may not appear offensive in isolation but becomes so when interpreted in context. For example, in the provision below, the model flagged the text not due to the presence of an explicit slur, but because migrants are framed alongside natural disasters and public health crises. Such contextual associations are difficult to capture using a purely RegEx-based approach.

*“Emergency’ includes any man-made or natural event or circumstance causing or threatening loss of life, injury to person or property, human suffering or financial loss, and includes, but is not limited to, fire, explosion, flood, severe weather, drought, earthquake, volcanic activity, spills or releases of oil or hazardous material as defined in ORS 466.605, contamination, utility or transportation emergencies, disease, blight, infestation, crisis influx of migrants unmanageable by the county, civil disturbance, riot, sabotage and war.”*

Examining the provisions initially captured by RegEx but subsequently filtered out by Gemini reveals that 79.63% of discarded matches originate from only 20 words. These terms are primarily common words that are typically offensive only in particular contexts, such as “black” or “colored.” These terms were only flagged by the LLM in 1% of appearances. We also observe cases in which the LLM’s reasoning focuses on a different phrase than the one originally matched by the RegEx. For example, in provisions containing multiple potentially offensive terms, Gemini may center its reasoning on only one term or identify an alternative stigmatizing phrase not captured by the initial pattern match. These patterns suggest that the RegEx and LLM components play complementary roles: The RegEx surfaces a broad set of candidate provisions, while the LLM refines this set by incorporating contextual interpretation. In practice, a larger lexicon that includes words not inherently offensive can still be effective for narrowing the search space when paired with an LLM. The RegEx also serves as a safeguard in cases where the LLM produces incorrect or incomplete judgments.

**Table 6: Additional examples of suspect provisions appearing in municipal codes and charters available online. Examples are edited for length.**

City	Law
Yazoo City, MS	“The city clerk shall procure registration books... indicating the date of registration, names of elector, <b>color... whether elector reads or understands when read to him, or can give a reasonable interpretation of the constitution of the state...</b> ” (City of Yazoo, Charter Sec. 9-8.)
Ostego, MI	“ <b>All women property holders</b> , and taxpayers <b>possessing qualifications of male electors</b> shall [be] <b>entitled to vote</b> ” (City of Ostego, Charter Sec. 2.)
Liberty, MO	“The City of Liberty... shall be divided into six wards... so as to include... the same number of <b>free white male inhabitants in each ward...</b> ” (City of Liberty, Charter Section 2.)
Sturgis, MI	“No licensee... shall... permit the licensed premises to be frequented by... known <b>prostitutes, homosexuals, vagrants...</b> ” (City of Sturgis, Code Sec. 6-13.)
Hazleton, PA	“Mandatory retirement. All employees, except elected officers... <b>shall retire at the age of 70 years.</b> ” (City of Hazleton, Code § 90-6)
Township of Rochelle Park, NJ	“No licensee shall employ... <b>any female persons</b> to perform... <b>any dances, ballet, acrobatics or public performances</b> of any kind in which body movement constitutes the principal or integral part of such performance;... nothing herein shall prevent a licensee from employing... <b>female persons playing musical instruments</b> ” (City of Township of Rochelle Park, Code § 58-17)
San Francisco, CA	“Every contract for any public work or improvement to be performed at the expense of the City and County... must provide... that <b>any person performing labor in the execution of the contract shall be a citizen of the United States.</b> ” (City of San Francisco, Charter A7.204)
Gibraltar, MI	“Family means to constitute a <b>man and wife or a father or a mother and their children...</b> and the parents of either or both...” (City of Gibraltar, MI, Code Sec. 44-24.)
Crawfordville, GA	“It shall be unlawful for any person to publicly appear... in any <b>dress not appropriate to such person’s sex...</b> ” (City of Crawfordville, GA, Code Sec. 28-20.)
El Paso, TX	“No person shall... knowingly furnish to any <b>male and female not husband and wife</b> any room, bed or other facilities <b>for sexual intercourse...</b> ” (City of El Paso, TX, Code 10.16.080)

**Table 7: Differential Treatment Detection Across Model Families. Performance was comparable for models from the Gemini, OpenAI, and Anthropic families along with several open models. However, smaller open-weight models produced lower quality reasoning traces, which motivated our preference for larger models in the clustering step. We chose to rely on Gemini models for our pipeline due to cost considerations and access to institutional API credits. We estimate our full pipeline cost approximately \$10,000.**

Model	Proportion of provisions correctly flagged for differential treatment
Gemini 2.5 Flash	98.3%
GPT 4o-mini	92.1%
GPT 4o	96.7%
GPT 5.1	98.3%
Claude Haiku 4.5	93.0%
Claude Sonnet 4.6	96.3%
Llama 3.1 8B Instruct	98.3%
Gemma 3 4B	98.4%
Qwen 3 8B	93.0%

**Table 8: Recall and Precision Over High Priority Laws Across Model Families. We conduct this experiment using the 306 sampled laws in the evaluation set. While not a fair end-to-end comparison, these results illustrate how legal prioritization varies across model families. Future work could explore enhancements such as utilizing a more complex LLM-based evaluation framework to better align performance on this task.**

Model	Recall	Precision
Gemini 2.5 Pro	88.0%	46.6%
Claude Sonnet 4.5	74.5%	57.2%
GPT 5.1	55.2%	63.7%

## References

- [1] U.S. Census Bureau. 2023. Population Estimates Program. <https://www.census.gov/programs-surveys/popest.html>. Accessed: 2026-01-31.
- [2] U.S. Census Bureau. 2023. TIGER/Line Shapefiles: States (and equivalent). <https://www.census.gov/cgi-bin/geo/shapefiles/index.php>. Accessed: 2026-01-31.