From Debate to Statute: Tracing Legislative Influence Using Generative AI

Mitchell Bosley[1] Kevin M. Esterling[2] Ju Yeon Park[3]

March 1, 2026

[1]Centro Nacional de Inteligencia Artificial (CENIA), Chile
[2]Department of Political Science, University of California, Riverside
[3]Department of Political Science, The Ohio State University

Abstract

Whether congressional committees engage in genuine deliberation or mere position-taking—a question long debated in the literature—has fundamental implications for democratic theory and institutional design. Yet, empirical tests of these competing theories have been limited by the absence of scalable methods to trace influence from committee deliberation to bill text. To address this gap, we develop a computational approach that combines mechanical text processing with theory-driven AI analysis to identify plausible pathways through which committee deliberation influences changes in the bill texts. Using bill-hearing-bill triplets, the method classifies changes by type, measures influence along three dimensions (temporal uniqueness, marginal contribution, and correspondence strength), and identifies specific mechanisms through which statements may shape legislation. Applied to a proof-of-concept analysis of nuclear waste policy legislation (H.R. 3053), the methodology produces interpretable measures showing that substantive-technical statements influenced changes in the bill but not organizational and symbolic parts of it did, with problem framing rather than direct language adoption as the dominant mechanism. We test whether committees function as information-gathering bodies (Krehbiel 1991), venues for position-taking (Mayhew 1974), or forums for genuine deliberation (Habermas 1996) and find support for the informational and deliberative theories over position-taking explanations.

1 Introduction

Scholars of the U.S. Congress have long debated whether testimony before congressional committees informs lawmaking or merely provides a venue for political theater. Information theorists argue that committees exist to gather specialized knowledge that reduces uncertainty for the floor, with hearings serving as crucial mechanisms for acquiring technical expertise (Krehbiel, 1991; Esterling, 2007). Alternatively, scholars emphasizing electoral incentives suggest that hearings primarily serve as venues for position-taking, where legislators express predetermined stances for constituency benefit rather than engage in genuine information exchange (Mayhew, 1974), or for "grandstanding" through strategic messaging to the public (Park, 2022). Despite decades of theoretical debate, empirical resolution has proven elusive due to the methodological challenge of tracing influence from testimony to legislative text at scale. Previous efforts have been limited by practical constraints: studies using manual coding typically examine only hundreds of cases due to labor intensity (Bussing et al., 2015), while quantitative approaches often rely on indirect proxies such as amendment success rates (Roberts and Smith, 2003) or witness counts (Lewison et al., 2020) that cannot capture the substantive content of deliberation. Recent computational work has begun to extract argument structures from hearings (Irani et al., 2024), but has not yet connected testimony content to subsequent legislative changes. This question carries implications beyond academic theory: if committees genuinely deliberate, then expanding participation and improving information quality could enhance democratic governance; if they merely provide theatrical venues, then reform efforts should focus elsewhere.

Recent advances in large language models offer a new methodological approach to this enduring question. While earlier computational methods using topic models or word embeddings could identify thematic similarity, LLMs can assess whether testimony and legislative text share conceptual approaches, problem framings, or solution strategies—approximating the context-dependent meaning and argumentative structure that qualitative researchers identify through close reading. Building on recent work exploring LLM applications to qualitative coding (Bosley, 2025), we develop a computational framework to measure potential influence between testimony and legislation, though systematic validation of these AI-generated measures remains essential before large-scale application. Our approach proceeds in three main stages. First, we identify and classify changes: we detect changes to a bill by comparing text from before and after a hearing, use semantic clustering to group related mechanical changes into coherent policy modifications, and classify these changes by type (substantive-technical, organizational-structural, or symbolic-expressive). Second, we match testimony to changes: we use AI to identify which speeches in the hearing are relevant to each semantic change, filtering out testimony unrelated to specific provisions. Third, we assess influence: for each semantic change, we jointly evaluate all relevant speeches along three ordinal dimensions—when the issue was raised (temporal uniqueness), what was added to the discussion (marginal contribution), and how closely the testimony aligns with the change (correspondence strength)—and identify the specific mechanism through which testimony may influence each change (direct language, technical details, problem framing, solution approach, or conceptual framework). This method enables us to identify influential speakers on a given piece of legislation and test whether influence patterns vary by change type.

The multi-level datasets generated by this approach enable direct empirical tests of competing theoretical predictions. If committees function as information-gathering bodies (Krehbiel, 1991), we would expect technical changes to show strong correspondence with expert testimony, particularly when experts provide novel data or specifications. If hearings serve primarily as position-taking venues (Mayhew, 1974), influence patterns should concentrate on symbolic-expressive changes with predetermined partisan alignments regardless of testimony content. If committees engage in genuine deliberation (Habermas, 1996), we would observe testimony shaping substantive provisions through problem framing and solution development, with influence distributed across multiple speakers rather than concentrated in predetermined authorities. By measuring these distinct patterns, researchers can move beyond binary assessments of whether testimony matters to understand when, how, and for whom deliberation shapes policy outcomes. Applying this methodology as a proof of concept to H.R. 3053 (Nuclear Waste Policy Amendments Act of 2017, 115th Congress), we analyze changes between the bill as introduced and as reported, along with testimony from a September 26, 2017 hearing before the Subcommittee on the Interior, Energy, and Environment. We find preliminary evidence of mixed theoretical support: while information theory correctly predicts that influence concentrates on substantive-technical changes where expertise matters, deliberation theory better explains the mechanisms—problem framing dominates over direct language adoption and influence distributes across multiple witnesses rather than concentrating in single authorities. These patterns from a single case establish the methodology's feasibility while highlighting the need for systematic validation and application across larger samples.

The remainder of this paper proceeds as follows. Section 2 develops the theoretical framework and measurement strategy for assessing testimony influence. Section 3 presents the computational methodology that operationalizes this framework. Section 4 demonstrates the application to H.R. 3053, walking through how the methodology processes the September 2017 hearing and bill changes, including a preliminary evaluation of theoretical predictions. Section 5 discusses implications for understanding committee function and outlines how this approach can be embedded in research designs to test competing theories of legislative behavior. Section 6 concludes.

2 Theoretical Framework

We develop a computational approach that measures patterns of correspondence between congressional testimony and legislative changes, enabling systematic tests of competing theories about committee function. Rather than treating influence as binary—either testimony matters or it doesn't—we measure multiple dimensions that distinguish different theoretical predictions: whether influence concentrates among experts or distributes across participants, operates through technical details or conceptual framing, and varies systematically across different types of legislative changes.

Our framework addresses a fundamental empirical challenge in legislative studies: tracing potential influence through complex deliberative processes at sufficient scale for systematic analysis. Committee hearings generate hundreds of pages of testimony touching diverse topics, while bills undergo numerous scattered edits between versions. Connecting specific testimony to particular changes requires identifying which mechanical edits constitute meaningful policy modifications, determining which speeches address those modifications, and assessing multiple dimensions of potential influence. Manual approaches that carefully trace these connections typically examine only small samples, while automated text-matching methods miss the semantic relationships and argumentative structures that constitute actual influence.

We solve this challenge through a measurement framework that captures both where influence might operate and how it manifests. The resulting measurements support multiple analytical strategies: comparing influence patterns across change types, identifying which speakers carry influence through which mechanisms, and testing whether technical expertise, political alignment, or deliberative engagement better predict testimony incorporation. Applied systematically across committees and policy domains, this framework can adjudicate between longstanding theoretical debates about whether committees gather information, provide venues for position-taking, or enable genuine deliberation.

2.1 Measurement Strategy

Our framework captures both where testimony influence might operate and how it manifests, providing the granular measurements necessary to test competing theories of committee function. Table 1 organizes these measurements into four complementary components that together enable researchers to trace potential pathways from testimony to legislative text.

We classify legislative changes to identify where testimony influence might plausibly operate, recognizing that not all modifications are equally susceptible to deliberative influence. Changes are classified along two intersecting dimensions: substantive content (technical details requiring expertise versus political choices involving values) and functional purpose (operational provisions that govern behavior versus expressive statements that signal positions). This yields the four categories shown in Table 1. Substantive-technical changes involve empirical parameters and specifications amenable to expert input, while substantive-political changes involve distributional decisions often predetermined by partisan alignments. Organizational-structural changes reorganize text without policy implications, and symbolic-expressive changes signal values without enforceable requirements.

We then measure influence through three ordinal dimensions that capture different aspects of how testimony shapes legislation. Temporal uniqueness (0–3) identifies when ideas enter the discourse, distinguishing agenda-setters who first raise issues from those who echo established points. Marginal contribution (0–3) recognizes that influence may come through providing decisive evidence or critical specifications rather than being first. Correspondence strength (0–3) assesses how closely testimony aligns with the final legislative text, from direct language adoption to no meaningful connection. These dimensions capture complementary aspects of influence—a late speech providing essential technical details might score low on temporal uniqueness but high on marginal contribution and correspondence.

When correspondence exists, we identify the specific mechanism through which influence operates, distinguishing five pathways rather than treating all influence as identical. Direct language adoption involves verbatim or near-verbatim incorporation of testimony phrasing. Technical details capture when specific numbers, thresholds, or procedures from testimony appear in provisions. Problem framing occurs when testimony's characterization of an issue shapes the legislative approach. Solution approach manifests when a strategic method advocated in testimony guides provision design. Conceptual framework applies when an underlying theory or model from testimony structures the legislation. These distinctions matter theoretically: information theory predicts influence through technical details and direct language, while deliberation theory emphasizes problem framing and conceptual frameworks.

Finally, we synthesize individual assessments into collective patterns that reveal how deliberation operates at the aggregate level. Influence may concentrate in one or two key speakers or distribute broadly across witnesses, measured through concentration indices. Temporal patterns distinguish between ideas that cascade through sequential refinement versus those emerging from independent consensus. These aggregate measures reveal not just whether testimony matters but how collective deliberation shapes policy outcomes—whether

Table 1: Measurement Framework for Testimony Influence

Dimension	Scale/Type	Description
Change Classification (Where Influence Operates)		
Substantive-Technical	Categorical	Changes involving empirical parameters, technical specifications, quantitative thresholds modifiable by evidence
Substantive-Political	Categorical	Resource allocation, jurisdictional assignments, distributional decisions reflecting value choices
Organizational-Structural	Categorical	Document reorganization, section renumbering without substantive policy change
Symbolic-Expressive	Categorical	Values signaling, findings, sense of Congress without enforceable requirements
Influence Dimensions (How Influence Manifests)		
Temporal Uniqueness	0-3 ordinal	When ideas enter discourse: 0=not discussed, 1=late/redundant, 2=early discussion, 3=first to raise
Marginal Contribution	0-3 ordinal	What speech adds: 0=nothing new, 1=minor elaboration, 2=significant evidence, 3=critical information
Correspondence Strength	0-3 ordinal	Alignment with change: 0=none, 1=tangential, 2=clear alignment, 3=direct match
Influence Mechanisms (Pathways of Influence)		
Direct Language	Categorical	Verbatim or near-verbatim text adoption from testimony
Technical Details	Categorical	Specific numbers, thresholds, procedures incorporated
Problem Framing	Categorical	Issue characterization shapes legislative approach
Solution Approach	Categorical	Strategic method advocated guides design
Conceptual Framework	Categorical	Underlying theory or model structures provision
Collective Patterns (Aggregate Influence)		
Distribution	Categorical	Concentrated (1-2 speakers) vs distributed (multiple speakers)
Temporal Pattern	Categorical	Single source, cascade, consensus, or contested
Overall Correspondence	0-3 ordinal	Weighted aggregate of individual correspondence scores

through expert dominance, distributed contribution, or iterative refinement.

2.2 Theoretical Implications

Our measurement framework enables researchers to adjudicate between competing theories by examining not just whether testimony influences legislation, but how influence operates across different contexts. The theories make distinct predictions because they posit fundamentally different logics of committee behavior: information theory views committees as specialized units reducing uncertainty for the floor, position-taking theory sees them as venues for electoral credit-claiming, and deliberation theory treats them as forums for collective reasoning. These different logics generate distinguishable empirical patterns in our measurements.

Information theory (Krehbiel, 1991) predicts that committees exist to develop and transmit expertise, implying that influence should concentrate where information asymmetries are greatest. We would observe substantive-technical changes showing strong correspondence with expert testimony, particularly through technical details and direct language mechanisms, as legislators defer to specialized knowledge they lack. Influence should concentrate among witnesses with recognized expertise—agency officials, technical specialists, industry engineers—rather than distribute broadly. Temporal patterns should show early technical testimony setting parameters that subsequent discussion accepts rather than challenges. Critically, influence should be minimal on substantive-political changes where legislators already have fixed preferences based on constituency interests.

Position-taking theory (Mayhew, 1974) predicts that hearings serve electoral rather than informational purposes, with legislators using testimony to justify predetermined positions. We would observe little correspondence between testimony and substantive changes of any type, as these decisions are made through partisan bargaining rather than deliberation. Influence should concentrate on symbolic-expressive provisions—findings, purpose statements, sense of Congress resolutions—that enable credit-claiming without binding policy commitments. The witnesses who appear influential should be those who provide quotable support for predetermined positions rather than those who offer novel information. Mechanisms should emphasize problem framing that supports electoral messaging rather than technical details that voters ignore.

Deliberation theory (Habermas, 1996) predicts that committees enable collective reasoning where participants can be persuaded through better arguments. We would observe testimony influencing substantive provisions through problem framing and solution approach mechanisms, as witnesses shape how legislators understand issues rather than just providing technical specifications. Influence should distribute across multiple speakers as ideas develop through discourse, with later speakers building on earlier contributions rather than simply repeating them. The marginal contribution dimension should matter more than temporal uniqueness, as deliberation values argument quality over primacy. Different types of changes might show different patterns, with technical issues more amenable to deliberation than partisan redistributions.

These competing theories generate distinct testable hypotheses:

- **H1 (Information Theory):** Substantive-technical changes will show significantly higher correspondence scores with testimony than substantive-political changes, as committees exist to gather specialized knowledge where information asymmetries are greatest.
- **H2 (Position-Taking Theory):** Influence will concentrate on symbolic-expressive provisions rather than substantive changes, with correspondence patterns determined by partisan alignment rather than technical expertise, as hearings serve electoral rather than informational purposes.
- **H3 (Deliberation Theory):** Changes showing high testimony correspondence will exhibit problem framing and solution approach mechanisms more frequently than direct language adoption, with influence distributed across multiple speakers based on argument quality (high marginal contribution scores) rather than formal authority.

Finding support for H1—technical changes responsive to expert testimony—would align with information theory. Finding support for H2—influence limited to symbolic provisions following partisan lines—would align with position-taking theory. Finding support for H3—distributed influence through conceptual mechanisms based on argument quality—would align with deliberation theory. Mixed patterns, such as H1 holding for technical changes while H2 holds for political changes, would suggest committees serve multiple functions depending on issue characteristics, with important implications for institutional design.

3 Methodology

The theoretical framework presented above requires a computational approach that can measure the multidimensional nature of influence at scale. Operationalizing these measurements confronts several practical challenges. Legislative changes manifest as scattered mechanical edits—a single policy decision might require adding definitions, modifying procedures, establishing penalties, and updating references across multiple bill sections. Analyzing these piecemeal edits rather than coherent policy changes would miss the forest for the trees. Additionally, committee hearings contain hours of testimony on diverse topics, most unrelated to specific legislative provisions, requiring methods to identify which speeches are relevant to which changes. Most fundamentally, assessing whether testimony and legislation share problem framings or solution approaches requires understanding meaning in context, not merely counting keyword overlaps.

The scale of legislative data compounds these challenges. Manual coding approaches that carefully trace testimony influence typically examine hundreds of cases at most (Bussing et al., 2015), insufficient for systematic comparison across committees, policy domains, or time periods. Automated approaches using keyword matching or topic models can process larger corpora but cannot capture the argumentative structure and contextual meaning necessary for assessing influence mechanisms. Recent advances in large language models offer a solution: the ability to assess semantic meaning, argumentative structure, and conceptual alignment at scale while maintaining consistency across cases.

This section describes how we operationalize the measurement framework through a three-stage pipeline that transforms hearing transcripts and bill versions into structured datasets capturing testimony-change correspondence. We combine mechanical text processing for reproducibility with AI-powered semantic analysis where human-like understanding is essential, enabling systematic measurement of the complex influence patterns that distinguish competing theories of committee function.

3.1 Data Requirements and Structure

Our methodology requires two primary data sources that are typically available for congressional proceedings. First, we need multiple versions of legislative bills that bracket committee deliberation—minimally, the version before committee consideration (as introduced or referred) and after committee action (as reported or marked up). These versions must be in structured text format that preserves the hierarchical organization of titles, sections, subsections, and paragraphs, enabling precise identification of where changes occur. Plain text or XML formats from congress.gov or the Government Publishing Office provide suitable structure, though researchers must ensure version timestamps align with committee activity.

Second, we need hearing transcripts with identified speakers and their complete statements. Official committee transcripts from GPO or committee websites typically provide speaker names, organizational affiliations, and testimony in chronological order. The transcripts must distinguish between witness testimony and member statements, as our framework focuses on external witness influence rather than member deliberation. While verbatim transcripts are ideal, substantially accurate transcripts that preserve argumentative content are sufficient, as our semantic analysis does not depend on exact wording.

Critical to valid inference is temporal alignment between hearings and bill versions. We analyze hearings that occur between the compared bill versions, ensuring that testimony could plausibly influence the observed changes. For bills with multiple hearings, researchers must determine which hearings are substantive (featuring external witnesses on bill content) versus procedural (markup sessions, business meetings). Similarly, bills undergoing multiple revisions require careful selection of which version pairs to analyze—typically the transition from introduced to reported captures committee influence most clearly.

Data completeness varies across committees and time periods. Modern hearings (post-2000) typically have complete transcripts and clearly versioned bills available electronically. Historical analyses face challenges including missing transcripts, unclear version dating, and inconsistent formatting. Our methodology can accommodate incomplete data by analyzing available hearing-change pairs, though systematic missingness could bias influence estimates. Researchers should document data availability and consider how missing hearings or unclear version histories might affect their theoretical inferences.

For our proof-of-concept application, we analyze H.R. 3053 (Nuclear Waste Policy Amendments Act of 2017) from the 115th Congress, examining changes between the bill as introduced on June 26, 2017,[1] and

[1]Available at: `https://www.govinfo.gov/content/pkg/BILLS-115hr3053ih/html/BILLS-115hr3053ih.htm`

as reported on November 30, 2017.[2] We focus on testimony from the September 26, 2017, hearing before the House Energy and Commerce Subcommittee on Environment,[3] which featured eight witnesses including federal agency representatives (NRC, GAO), state officials, industry representatives, and advocacy groups discussing nuclear waste storage and disposal provisions. This case provides an ideal test because the bill underwent substantive committee revision with clear version dates, the hearing transcript includes diverse witnesses with varying expertise and positions, and the technical nature of nuclear waste policy allows clear assessment of expertise-based influence patterns.

While this proof-of-concept demonstrates the methodology on a single bill-hearing-bill triplet, the approach is designed for systematic application across multiple congresses. Future research should analyze hundreds or thousands of such triplets to identify temporal patterns in committee deliberation: whether influence mechanisms have shifted with increasing polarization, how the rise of social media affects testimony influence, and whether certain committees or policy domains show consistent deliberative patterns over time. The scalability enabled by our computational approach makes such comprehensive analysis feasible for the first time, moving beyond case studies toward systematic empirical assessment of how congressional deliberation has evolved.

3.2 Computational Solution: A Three-Stage Approach

Algorithm 1 presents the complete computational approach that transforms hearing transcripts and bill versions into datasets measuring testimony influence. The algorithm progressively builds from mechanical text processing through semantic analysis to multidimensional influence assessment, with each stage serving a distinct theoretical and computational purpose.

We employ large language models specifically where human-like semantic understanding is required: clustering related changes into policy modifications, classifying changes by type and deliberative potential, identifying which testimony is relevant to which changes, and assessing the multidimensional nature of influence. Mechanical tasks—text extraction, difference detection, speaker identification—use deterministic algorithms to ensure reproducibility.

Our prompting strategy emphasizes structured reasoning with explicit justification requirements. Each prompt follows a consistent pattern: context presentation, task specification with detailed rubrics, and structured output format requiring both scores and justifications. Figures 1–3 present the core prompts and sample outputs for each stage of analysis.

This structured prompting approach ensures comprehensive understanding before categorical judgments, while joint assessment enables proper evaluation of temporal and marginal dimensions relative to other testimony.

3.2.1 Stage 1: Mechanical Processing

The first stage (Stage 1 in Algorithm 1) performs deterministic text processing without interpretive judgment. We extract structured representations of each bill version, parsing the hierarchical organization of titles, sections, subsections, and provisions. A line-by-line comparison between consecutive versions identifies all textual differences—insertions, deletions, and modifications—with their precise locations in the bill structure. Simultaneously, we parse hearing transcripts to identify individual speakers and their statements, extracting metadata including speaker names, organizational affiliations, and the temporal sequence of testimony. This mechanical processing establishes the complete universe of changes and speeches that subsequent stages will analyze, ensuring no potential influence pathways are overlooked.

3.2.2 Stage 2: Semantic Analysis

The second stage (Stage 2 in Algorithm 1) transforms mechanical outputs into meaningful analytical units while enriching them with essential metadata for influence assessment. We cluster related mechanical changes that collectively represent coherent policy modifications. This clustering works bidirectionally: multiple mechanical edits may represent a single policy decision (adding a definition, specifying procedures, and

[2] Available at: `https://www.govinfo.gov/content/pkg/BILLS-115hr3053rh/html/BILLS-115hr3053rh.htm`

[3] Hearing transcript available at: `https://www.govinfo.gov/content/pkg/CHRG-115hhrg27759/html/CHRG-115hhrg27759.htm`

Algorithm 1 Assessing Testimony Influence on Legislative Changes

Require: Bill versions $\mathcal{B} = \{b_1, \ldots, b_v\}$, Hearing transcript $\mathcal{H}$
Ensure: Influence assessment datasets $\mathcal{D}$

Stage 1: Mechanical Processing
Extract structured representation of each bill version
$M \leftarrow$ compute mechanical differences between versions
$T \leftarrow$ parse hearing transcript into individual speeches with metadata

Stage 2: Semantic Analysis
$C \leftarrow \{$Initialize semantic changes$\}$
while unprocessed changes remain in M **do**
  Group related mechanical changes by proximity and coherence
  $c \leftarrow$ create semantic change from group
  Classify c as $\in \{$technical, political, structural, symbolic$\}$
  $C \leftarrow C \cup \{c\}$
end while

Stage 3: Influence Assessment
// First: Filter for relevance
for each semantic change $c \in C$ **do**
  $R_c \leftarrow \{$Relevant speeches for change $c\}$
  for each speech $t \in T$ **do**
    if relevance$(t, c) > \theta$ **then**
      $R_c \leftarrow R_c \cup \{t\}$
    end if
  end for
end for

// Second: Joint assessment of all relevant speeches per change
for each semantic change $c \in C$ **do**
  if $|R_c| > 0$ **then**
    Jointly assess all speeches in R_c together:
    for each speech $t \in R_c$ **do**
      $u_{tc} \leftarrow$ assess temporal uniqueness relative to other speeches in R_c
      $m_{tc} \leftarrow$ assess marginal contribution beyond prior speeches in R_c
      $s_{tc} \leftarrow$ assess correspondence strength with change c
      if $s_{tc} > 0$ **then**
        $\mu_{tc} \leftarrow$ identify influence mechanism
      end if
    end for
    Synthesize collective influence patterns from all assessments
  end if
end for
Generate change-level, dyad-level, and speaker-level datasets
return $\mathcal{D} = \{D_{\text{change}}, D_{\text{dyad}}, D_{\text{speaker}}\}$

Semantic Clustering Prompt (Abridged)

```
INPUT: {All mechanical changes from bill comparison}

Analyze these legislative changes to:
1. Group related changes into meaningful clusters
2. Classify each cluster by deliberative potential

CHANGES TO ANALYZE:
{List of change_id, location, type, removed_text,
 added_text for each mechanical change}

CATEGORIES:
- substantive-technical: Specific quantitative measures
  (thresholds, timelines, technical specifications)
- substantive-political: Resource allocation or
  jurisdictional authority assignments
- organizational-structural: Reorganization without
  policy change (section moves, renumbering)
- symbolic-expressive: Values signaling without
  enforceable requirements
```

Sample Output:

```
{
  "semantic_groups": [
    {
      "group_id": "SG_001",
      "title": "Radiation Exposure Standards",
      "change_classification": "substantive-technical",
      "classification_reasoning": "Contains specific
        radiation threshold of 100 millirems/year
        based on EPA technical standards",
      "change_ids": ["CHG_001", "CHG_002", "CHG_007"]
    },
    {
      "group_id": "SG_002",
      "title": "Section Reorganization",
      "change_classification": "organizational-structural",
      "change_ids": ["CHG_003", "CHG_004", ...]
    },
    ...
  ]
}
```

Figure 1: Semantic Clustering: Groups mechanical changes into policy-coherent clusters and classifies their susceptibility to deliberative influence

Relevance Filtering Prompt (Abridged)

```
INPUT: {One speech + ALL semantic changes}

SPEECH TO ANALYZE:
{speech_id, speaker_name, organization, content}

ALL SEMANTIC CHANGES ({m} total):
{List of all semantic changes with change_id,
 title, description, key_concepts}

For each semantic change, determine if speech addresses it:
- Direct mentions of same policy area or provision
- Thematic overlap with subject matter
- Supporting/opposing arguments
- Background context informing the decision

Return: relevance determination for EACH change
```

Sample Output:

```
{
  "speech_id": "HEAR_20170926_EPA_001",
  "relevance_assessments": [
    {
      "change_id": "SC_001",
      "is_relevant": false,
      "reasoning": "No discussion of structural reorganization"
    },
    {
      "change_id": "SC_002",
      "is_relevant": false,
      "reasoning": "No mention of sense of Congress statements"
    },
    {
      "change_id": "SC_003",
      "is_relevant": true,
      "relevance_type": "direct_topic_match",
      "reasoning": "Speaker directly addresses the
        100 millirem threshold in Section 403"
    },
    ... // Assessments for all other semantic changes
  ]
}
```

Figure 2: Relevance Filtering: Identifies which speeches address each semantic change to reduce false positives

Influence Assessment Prompt (Abridged)

```
INPUT: {One semantic change + ALL relevant speeches}

SEMANTIC CHANGE:
{change_id, title, description, classification}

ALL RELEVANT SPEECHES ({n} total):
{List of all speeches marked relevant for this change,
 in chronological order with full content}

FIRST - Write analytical narrative:
What is the story of influence here? Which speeches
seem most influential? How do speeches relate?

THEN - Score each speech:
- Temporal Uniqueness (0-3): When was this raised?
- Marginal Contribution (0-3): What does this add?
- Correspondence Strength (0-3): How closely aligned?
- Mechanism: direct_language | technical_details |
  problem_framing | solution_approach |
  conceptual_framework
```

Sample Output:

```
{
  "exploration": "Chronologically, early speeches
    (HEAR_20170926_EXT_001 to EXT_015) frame nuclear
    waste broadly... Mr. O'Donnell (EXT_005, EXT_018)
    and Mr. Victor (EXT_011, EXT_020) are key voices...
    Ms. Tubb (EXT_062) explicitly states need to
    complete Yucca Mountain licensing first...",

  "individual_assessments": [
    {
      "speech_id": "HEAR_20170926_EXT_005",
      "speaker_name": "Mr. O'Donnell",
      "temporal_uniqueness": 2,
      "marginal_contribution": 2,
      "correspondence_strength": 2,
      ...
    },
    {
      "speech_id": "HEAR_20170926_EXT_062",
      "speaker_name": "Ms. Tubb",
      "temporal_uniqueness": 2,
      "temporal_justification": "Explicitly states need
        to complete licensing as first step",
      "marginal_contribution": 3,
      "correspondence_strength": 3,
      ...
    },
    ... // Assessments for all 39 relevant speeches
  ],

  "collective_synthesis": {
    "primary_mechanism": "solution_approach",
    "influence_distribution": "distributed",
    "originators": ["EXT_005", "EXT_011"],
    "amplifiers": ["EXT_062", "EXT_085", ...],
    ...
  }
}
```

Figure 3: Influence Assessment: Joint evaluation of all relevant speeches to measure three dimensions of influence and identify mechanisms

updating cross-references), while conversely, a single large mechanical addition containing multiple provisions may be split into distinct semantic changes representing separate policy choices.

Beyond clustering, this stage adds critical analytical metadata to each semantic change:

- **Descriptive title and summary**: Human-readable descriptions of what the change accomplishes
- **Key concepts**: Searchable terms and phrases central to the change (e.g., "repository decision," "interim storage," "notification requirements")
- **Affected stakeholders**: Entities impacted by the change (DOE, NRC, states, facility operators)
- **Deliberative characteristics**: Assessments of specificity (high/medium/low), visibility, controversy potential, and amenability to evidence
- **Legislative intent**: Inferred purpose of the change when discernible
- **Policy elements**: Specific requirements, conditions, or obligations created

We use a large language model to perform this enrichment, providing it with all mechanical changes and asking it to group or divide them based on policy coherence while extracting these metadata elements (Figure 1). The model considers textual proximity, shared concepts and stakeholders, and functional independence to identify appropriate semantic clusters. Each resulting semantic change is then classified along our four-category taxonomy—substantive-technical, substantive-political, organizational-structural, or symbolic-expressive—with explicit reasoning for the classification. This enriched representation enables more precise relevance filtering in Stage 3, as speeches can be matched against key concepts and stakeholder impacts rather than just raw text.

3.2.3 Stage 3: Influence Assessment

The third stage (Stage 3 in Algorithm 1) applies the measurement framework through two phases: relevance filtering and joint assessment.

First, we filter potential speech-change connections efficiently. Rather than evaluating all $n \times m$ pairs individually, each speech is assessed against all semantic changes simultaneously, identifying relevant connections based on topical overlap, thematic relevance, argumentative connection, or conceptual alignment.

For each change with relevant speeches, we then conduct joint assessment. The model first constructs an analytical narrative exploring how ideas develop chronologically, then scores each speech on three dimensions (temporal uniqueness 0–3, marginal contribution 0–3, correspondence strength 0–3) with explicit justifications. Joint assessment is essential—it ensures proper attribution by evaluating speeches relative to each other rather than in isolation. When correspondence exceeds zero, the model identifies the primary influence mechanism from our five categories.

Individual assessments are synthesized into collective patterns using:

- Influence weights: temporal uniqueness $\times$ marginal contribution $\times$ correspondence strength
- Concentration metrics: Herfindahl index to measure distribution across speakers[4]
- Dominant mechanism: modal category across relevant speeches
- Temporal pattern: single source, cascade, consensus, or contested

This produces interpretable metrics at the change level (overall correspondence, concentration, dominant mechanism) and preserves detailed scoring at the speech-change dyad level for analysis.

[4]The Herfindahl index is calculated as $H = \sum_i s_i^2$ where s_i is speech i's share of total influence weight for a given change.

Table 2: Structure of Three Output Datasets

Variable Category	**Change-Level Dataset**	**Speech-Change Dyad Dataset**	**Speaker-Level Dataset**
Unit of Analysis	Semantic change	Speech-change pair	Individual speaker
Example Observation	SC_003: Yucca Mountain Licensing	(HEAR_20170926_EXT_062, SC_003)	Ms. Tubb (Heritage Foundation)
Identifiers	semantic_change_id: SC_003 title: Yucca Mountain Licensing change_type: substantive-technical	speech_id: HEAR_20170926_EXT_062 semantic_change_id: SC_003 speaker_name: Ms. Tubb change_title: Yucca Mountain Licensing	speaker_name: Ms. Tubb organization: Heritage Foundation speaker_type: think_tank
Influence Metrics	overall_correspondence: 2.4 num_relevant_speeches: 39 num_high_influence: 8 influence_concentration: 0.32	temporal_uniqueness: 2 marginal_contribution: 3 correspondence_strength: 3 influence_mechanism: solution_approach	mean_temporal: 1.8 mean_marginal: 2.1 mean_correspondence: 2.3 std_correspondence: 0.7
Patterns/Mechanisms	primary_mechanism: problem_framing influence_distribution: distributed temporal_pattern: cascade	mechanism: solution_approach temporal_justification: First to propose marginal_justification: Added sequencing	primary_mechanism: technical_details secondary_mechanism: problem_framing concentration_index: 0.45
Metadata	deliberative_potential: high amenability_to_evidence: 0.85 top_influencer: Ms. Tubb originators: [O'Donnell, Victor]	hearing_position: 62 speech_length: 847 words topics_addressed: [licensing, safety] references_data: true	num_speeches: 12 total_words: 3,847 num_changes_addressed: 6 prior_testimony_count: 7
Key Research Uses	Test which change types show influence Compare technical vs political changes Assess deliberative potential	Identify influential speaker types Test mechanism effectiveness Analyze temporal dynamics	Compare agency vs interest groups Assess expertise effects Track influence concentration

3.2.4 Output Generation

This approach generates three datasets supporting different analytical strategies. Table 2 illustrates the structure of each dataset with example observations from our proof-of-concept analysis.

The change-level dataset takes semantic changes as the unit of analysis, with each observation representing a distinct policy modification. Variables include the change classification (substantive-technical, substantive-political, organizational-structural, or symbolic-expressive), aggregate influence metrics (overall correspondence score, number of contributing speakers, concentration of influence), the dominant influence mechanism across all relevant testimony, and change characteristics such as scope, visibility, and amenability to evidence. This dataset enables researchers to test whether different types of changes show differential susceptibility to testimony influence—for instance, whether technical provisions respond more to expert testimony than political allocations, or whether high-visibility changes resist influence regardless of testimony quality.

The speech-change dyad dataset provides the finest granularity, with each observation representing a unique pairing of an individual speech with a semantic change. Variables include the three ordinal influence dimensions (temporal uniqueness, marginal contribution, correspondence strength), the identified influence mechanism for that specific pairing, speaker characteristics (expertise, organizational affiliation, political position), and speech metadata (timing in hearing, length, topics addressed). This dyadic structure supports sophisticated hypothesis testing about the conditions under which testimony influences legislation—whether certain types of speakers are more influential on particular changes, whether early testimony shapes subsequent discussion, or whether specific mechanisms are more effective for different change types.

The speaker-level dataset aggregates to individual witnesses or organizations, capturing their overall participation and influence patterns. Variables include participation metrics (number of speeches, topics addressed, speaking time), average influence scores across all relevant changes, the variance and distribution of influence (concentrated on specific issues versus broad impact), and the predominant mechanisms through which the speaker operates. This dataset enables analysis of which actors carry the most influence in committee deliberations, whether influence correlates with expertise or political alignment, and whether institutional position (government officials versus outside experts versus affected interests) predicts testimony incorporation.

3.3 Implementation Details

We employ OpenAI's GPT-4.1-mini model for all semantic analysis tasks, selected for its balance of performance and cost-efficiency. Key parameter choices prioritize consistency and reproducibility: we use temperature 0.3 for all tasks to ensure stable outputs while allowing appropriate variation in scoring, set maximum tokens to 32,000 to accommodate comprehensive analysis of lengthy testimony, and enforce JSON output format to ensure structured, parseable responses. Each API call includes detailed system prompts with explicit rubrics, scoring guidelines, and requirements for justification.

The computational cost scales with corpus size. For our proof-of-concept case analyzing 112 individual speeches and 9 semantic changes, the pipeline required: Stage 1 (mechanical processing) using deterministic Python algorithms at negligible cost, Stage 2 (semantic clustering) requiring one API call for clustering and 112 calls for relevance filtering, and Stage 3 (influence assessment) requiring 9 API calls for joint evaluation of all relevant speech-change pairs. At GPT-4.1-mini pricing of \$0.15 per million input tokens and \$0.60 per million output tokens, the total cost was approximately \$0.05–0.10 per hearing-bill pair, making the approach highly feasible for large-scale studies encompassing thousands of hearings.

To ensure reproducibility, we log all prompts, responses, and parameters. The modular pipeline design allows substitution of alternative models—researchers could use Claude, Gemini, or open-source models like Llama with minimal code changes. Future implementations could further reduce costs through selective sampling strategies or by using smaller models for initial filtering before detailed assessment with more capable models.

3.4 Validation Approach

While this proof-of-concept application demonstrates face validity—the measurements align with qualitative scholarly understanding of influence patterns—future work should employ multiple validation strategies. Convergent validity could be established by comparing our measures with committee report citations, amendment sponsorship patterns, and subsequent witness invitations. Predictive validity could test whether high-influence testimony predicts provision survival through floor consideration and conference committees. Construct validity requires confirming that our three dimensions capture distinct aspects of influence rather than a single factor, which factor analysis across larger samples could establish. Inter-coder reliability between human experts and AI assessments on sampled cases would strengthen confidence in the automated coding. As the methodology is applied across more cases, these validation approaches will establish whether the framework reliably captures meaningful patterns of committee deliberation.

4 Application to H.R. 3053: Nuclear Waste Policy

To demonstrate the methodology, we apply it to H.R. 3053, the Nuclear Waste Policy Amendments Act of 2017, which sought to revive the Yucca Mountain repository project and establish monitored retrievable storage facilities for nuclear waste. The bill underwent significant revision in committee between its introduction on June 26, 2017, and being reported on November 30, 2017. During this period, the House Energy and Commerce Subcommittee on Environment held a hearing on September 26, 2017, featuring testimony from diverse stakeholders about nuclear waste management policy.

4.1 Context and Data

The nuclear waste domain provides an ideal test case for our methodology. First, it involves substantial technical complexity—radiation standards, geological assessments, engineering specifications—where expert testimony could plausibly influence legislative provisions. Second, it carries political significance with clear distributional implications for states and communities, allowing us to observe whether technical and political changes show different influence patterns. Third, the September 26, 2017 hearing generated rich testimony data: 14 witnesses delivered 112 individual speeches, providing diverse perspectives from federal agencies (NRC, GAO), state officials (Nevada, Maryland), local government (Aiken County), academic experts (UC San Diego), utilities (Southern Company), and policy organizations (Heritage Foundation, Union of Concerned Scientists).

The hearing addressed the urgent need for nuclear waste storage solutions, with the United States having accumulated 76,000 metric tons of spent nuclear fuel stored at reactor sites nationwide. Witnesses debated the safety and feasibility of Yucca Mountain, the role of interim storage facilities, transportation risks, and community consent mechanisms. The federal government's failure to meet its obligations under the 1982 Nuclear Waste Policy Act has resulted in liability estimates ranging from $24.7 to $50 billion, making this both a technical and fiscal crisis requiring legislative resolution.

4.2 Stage 1: Mechanical Change Identification

Our deterministic text processing compared the introduced version (June 26, 2017) with the reported version (November 30, 2017) of H.R. 3053, identifying 11 mechanical differences at the section and subsection level. The algorithm detected three types of changes: additions (new text appearing in the reported version), deletions (text removed from the introduced version), and modifications (text altered between versions).

The most significant mechanical change was CHG_003, a 1,247-word addition creating an entirely new Section 106 titled "Conditions for Use of an Initial MRS Facility for DOE-Owned Civilian Waste." This single addition contained five distinct subsections establishing:

- A general prohibition on DOE storing its civilian waste at the initial MRS facility
- An exception for when the Secretary determines a repository decision is imminent
- Notification requirements to Congress within 30 days of such determination
- Annual reporting obligations thereafter
- Clarifications about federal disposal obligations

Seven mechanical changes (CHG_001, 002, 004, 006–009) involved structural reorganization—moving titles between sections, renumbering provisions, and restructuring the document hierarchy. For instance, CHG_001 deleted "TITLE II–PERMANENT REPOSITORY" from one location while CHG_002 added it elsewhere with additional section listings, effectively reorganizing rather than eliminating content.

Two changes (CHG_005, 010) deleted "Sense of Congress" provisions from the introduced bill. CHG_005 removed a 312-word statement opposing transportation routes through the Great Lakes, while CHG_010 eliminated a shorter statement about nuclear waste storage concerns. These deletions represent the committee's decision to remove non-binding congressional expressions from the bill.

The final mechanical change (CHG_011) modified language in Section 403(a), including terminology adjustments like payment calculations "per metric ton" and adding the conforming amendment that updates the Nuclear Waste Policy Act's table of contents to reflect the new Section 106.

This distribution of mechanical changes—one major substantive addition, seven reorganizations, two deletions, and one modification—reveals that the committee's primary substantive work concentrated on establishing the regulatory framework for interim storage, while also streamlining the bill by removing symbolic statements and improving its organizational structure.

4.3 Stage 2: Semantic Clustering and Classification

The 11 mechanical changes clustered into 9 semantic groups representing distinct policy modifications. Table 3 provides a complete mapping of these changes, their descriptions, and classifications.

As shown in Table 3, substantive-technical changes dominated (6 of 9), with no substantive-political changes appearing in the committee revision. This distribution reveals important patterns about how the committee approached nuclear waste policy.

The three core substantive changes with high to medium impact (SC_003–005) fundamentally restructured how interim storage facilities relate to repository development. SC_003 establishes a prohibition preventing the Department of Energy from storing its own civilian nuclear waste at the initial monitored retrievable storage (MRS) facility until a "final repository decision" occurs—defined as either NRC authorization to construct Yucca Mountain or a presidential/congressional decision to abandon it. This prohibition creates a regulatory gate that links interim storage to permanent repository progress, addressing concerns that interim facilities might become de facto permanent storage.

SC_004 provides a carefully crafted exception to this prohibition: the Secretary of Energy may allow storage if finding that a final repository decision is "imminent," specifically that such decision will occur before storage contracts (limited to 10 years) expire. This exception balances the need for storage flexibility with the commitment to repository development. SC_005 then establishes accountability mechanisms, requiring the Secretary to notify Congress within 30 days of any imminent decision finding and submit annual reports thereafter, ensuring congressional oversight of these determinations.

The remaining substantive-technical changes provide supporting structure for this regulatory framework. SC_006 clarifies that these storage conditions do not alter the federal government's underlying obligation to dispose of nuclear waste, preventing legal arguments that interim storage satisfies disposal requirements. SC_007 supplies precise definitions for "final repository decision" and "initial MRS facility" to avoid ambiguity in implementation. SC_008 adds a conforming amendment to update the Nuclear Waste Policy Act's table of contents, a technical necessity for statutory coherence.

The organizational-structural changes reflect committee decisions about document architecture rather than policy substance. SC_001 consolidates seven mechanical changes that reorganize titles and sections throughout the bill—removing redundant headings, renumbering sections, and restructuring the document flow without altering any requirements or obligations. SC_009 captures minor terminology modifications, such as changing payment calculations from "per metric ton" specifications, that clarify language without changing meaning.

The single symbolic-expressive change (SC_002) removes two "Sense of Congress" provisions from the introduced bill that expressed non-binding preferences about transportation routes avoiding the Great Lakes and concerns about current storage practices. The removal of these provisions suggests the committee chose to focus on binding requirements rather than rhetorical statements.

Notably absent are substantive-political changes involving resource allocation, state authority, or distributional decisions that might trigger partisan conflict. The committee avoided provisions granting states

Table 3: Semantic Changes Identified in H.R. 3053

ID	Description	Classification	Substantive Impact	Relevant Speeches	Mechanical Changes
SC_001	Section and title reorganization: Removes/adds titles and section headings to restructure document without policy change	Organizational-Structural	Low	0	CHG_001–002, 004, 006–009
SC_002	Removes “Sense of Congress” provisions about transportation routes near Great Lakes and nuclear waste storage	Symbolic-Expressive	Minimal	0	CHG_005, 010
SC_003	Prohibits storage of DOE-owned civilian waste at initial MRS facility until final repository decision (new Section 106)	Substantive-Technical	High	39	CHG_003
SC_004	Creates exception allowing storage if Secretary finds final repository decision imminent (within 10-year contracts)	Substantive-Technical	High	22	CHG_003
SC_005	Requires notification to Congress and annual reports if imminent decision finding is made	Substantive-Technical	Medium	5	CHG_003
SC_006	Clarifies that storage conditions do not affect federal disposal obligations	Substantive-Technical	Low	29	CHG_003
SC_007	Defines “final repository decision” and “initial MRS facility” for regulatory clarity	Substantive-Technical	Low	1	CHG_003
SC_008	Adds conforming amendment updating Nuclear Waste Policy Act table of contents	Substantive-Technical	Minimal	0	CHG_011
SC_009	Minor terminology modifications (e.g., “per metric ton” changes)	Organizational-Structural	Minimal	0	CHG_011

veto power over siting decisions, allocating federal funds to specific districts, or establishing benefit formulas for host communities—all politically charged issues raised in testimony but not incorporated into the reported bill.

4.4 Stage 3: Influence Assessment Results

The influence assessment reveals how 112 speeches shaped nine semantic changes through a dramatic filtering process that identified genuine influence.

4.4.1 The Filtering Process: From All Testimony to Measurable Influence

The assessment process operated as a three-stage funnel with increasingly selective filters. Starting with 112 speeches addressing various topics, the relevance filter identified 96 speech-change pairs (9.5% of 1,008 possible) where testimony addressed specific provisions. The influence assessment then found that 12 of the 14 speakers (85.7%) achieved non-zero influence scores, generating all 96 relevant pairs. Finally, only 14 pairs (14.6% of relevant) achieved high-impact status with scores of 2 or higher on all three dimensions.

Panel A of Table 6 shows this filtering concentrated entirely on substantive-technical changes. The storage prohibition (SC_003) attracted 39 speeches, federal obligations (SC_006) drew 29, and the imminent decision exception (SC_004) engaged 22 speeches. Meanwhile, organizational-structural changes and symbolic-expressive changes attracted zero influential testimony despite comprising three of nine changes. This stark distribution confirms witnesses focus strategically on provisions where influence is possible.

4.4.2 High-Impact Moments: The 14 Speeches That Mattered Most

Of 96 relevant speech-change pairs, only 14 (14.6%) achieved high-impact status with influence weights of 6 or greater (scores of 2+ on all three dimensions). These moments concentrated in just four provisions and emerged from seven speakers, revealing both the rarity of decisive testimony and the selective expertise that drives legislative change. Table 4 presents the ten most influential speech-change pairs.

Table 4: Highest Impact Speech-Change Pairs in H.R. 3053 Testimony

Rank	Speaker	Change	TU	MC	CS	Weight	Key Contribution
1	Ms. Tubb	SC_004	3	3	3	27	"Imminent decision" framework
1	Mr. O'Donnell	SC_006	3	3	3	27	Federal obligations language
3	Ms. Tubb	SC_003	2	3	3	18	Repository-storage linkage
3	Mr. Palmer	SC_005	3	3	2	18	Notification requirements
3	Mr. Lyman	SC_006	2	3	3	18	Technical clarifications
3	Mr. O'Donnell	SC_006	2	3	3	18	Reinforcing obligations
7	Mr. Victor	SC_003	2	3	2	12	Interim becoming permanent
7	Ms. Tubb	SC_003	2	3	2	12	Regulatory gates concept
7	Mr. Issa	SC_003	2	3	2	12	Safety thresholds
10	Mr. O'Donnell	SC_003	2	2	2	8	State concerns

TU = Temporal Uniqueness, MC = Marginal Contribution, CS = Correspondence Strength (all 0–3 scales).
Weight = TU $\times$ MC $\times$ CS. Only pairs with weight ≥ 8 shown; 4 additional pairs achieved weight 6–7.

Two speeches achieved perfect scores across all dimensions (weight=27), representing rare instances of direct legislative adoption of witness proposals.

Ms. Tubb on SC_004: The "Imminent Decision" Framework. Ms. Tubb's testimony uniquely proposed allowing interim storage only when a repository decision was "imminent," specifically suggesting a timeframe of "within 10 years." The final bill provision mirrors this framework precisely: "The Secretary determines that the President will submit a recommendation to Congress...within a reasonable time (not to exceed 10 years)." While the bill uses "reasonable time" rather than "imminent," it adopts her exact 10-year threshold and conditional structure—storage permitted only when permanent disposal is demonstrably approaching.

Mr. O'Donnell on SC_006: Federal Obligations Clarification. Representing Nevada, Mr. O'Donnell emphasized that accepting interim storage must not relieve federal obligations for permanent disposal, warning against creating "de facto permanent storage" through legal loopholes. The bill directly incorporates this concern: "Nothing in this section relieves the Secretary of any obligation under this Act regarding the disposal of high-level radioactive waste and spent nuclear fuel." This represents the only instance of near-verbatim language adoption in our analysis.

SC_003: Convergent Testimony Shaping Storage Prohibition. The storage prohibition attracted distributed high-impact testimony, with five speeches in the top ten contributing distinct elements that appear in the final provision. Ms. Tubb introduced the concept of "regulatory gates"—requiring repository progress before storage authorization. Mr. Victor emphasized the risk of interim storage becoming permanent, citing international examples where "temporary" facilities operated for decades. Mr. Issa contributed technical safety thresholds that storage facilities must meet. The final provision synthesizes these elements: prohibiting storage unless the NRC has issued construction authorization for a permanent repository, incorporating both the gating mechanism (Tubb), the permanence concern (Victor), and safety requirements (Issa).

This pattern—multiple witnesses independently converging on complementary aspects of a solution—characterizes substantive-technical provisions where expertise from different perspectives strengthens the final language. In contrast, SC_004 and SC_005 show concentrated influence where single witnesses provided complete frameworks adopted largely intact. These speech-level patterns point to broader questions about which types of witnesses achieve influence and how testimony capacity distributes across the witness pool.

These patterns validate the method's discriminating power while revealing how testimony influences legislation. The variation in correspondence scores—from 1.36 for SC_003 to 0.00 for SC_007—demonstrates the ability to distinguish changes shaped by deliberation from those emerging through legislative drafting. The dominance of problem framing over direct language adoption suggests committees absorb conceptual frameworks rather than specific text. The concentration of influence on substantive-technical changes aligns with information theory predictions about where expertise matters most. (Appendix A provides detailed patterns for each semantic change.)

4.4.3 Speaker-Level Patterns

Aggregating from individual speeches to speakers reveals how influence concentrates among witnesses and which institutional positions translate into legislative impact. While the high-impact speech analysis identified specific moments of influence, the speaker-level view shows that these moments clustered among a small subset of witnesses with particular expertise profiles. Table 5 presents influence patterns for the seven speakers who achieved measurable impact.

Table 5: Speaker-Level Influence Distribution in H.R. 3053 Hearings

Speaker	Affiliation/Type	Changes Addressed	Speeches Relevant	Mean Scores TU	MC	CS	High Impact[a]
Mr. O'Donnell	Nevada State Official	4	10	2.1	2.4	1.5	3
Ms. Tubb	Southern Company (Utility)	3	9	2.0	2.7	1.5	3
Mr. Victor	Yucca Mountain Expert	3	16	1.8	1.8	1.2	1
Mr. Palmer	Rep. Alabama	4	8	2.1	1.9	1.0	1
Mr. Issa	Rep. California	3	10	1.5	2.1	1.0	1
Mr. Lyman	Rep. Michigan	5	8	1.6	1.9	1.1	1
Mr. Smith	Rep. Nebraska	4	11	0.9	1.5	1.3	0
Others (5)	Various	—	24	0.4	0.7	0.5	0

TU = Temporal Uniqueness, MC = Marginal Contribution, CS = Correspondence Strength (all 0–3 scales).

[a]Number of speech-change pairs achieving scores ≥ 2 on all three dimensions.

100 speakers showed zero influence; 12 speakers achieved measurable influence.

Influence Concentration. Influence concentrated moderately among key speakers without creating monopolistic control. The Herfindahl index of 0.161 indicates moderate concentration, comparable to a

Table 6: Descriptive Statistics from Influence Assessment of H.R. 3053

Panel A: Relevance Filtering Performance

Change Classification	Changes	Relevant Speeches	Mean per Change	Total Possible	Relevance Rate
Substantive-Technical	6	96	16.0	672	14.3%
Substantive-Political	0	0	–	0	–
Organizational-Structural	2	0	0.0	224	0.0%
Symbolic-Expressive	1	0	0.0	112	0.0%
Total	**9**	**96**	**10.7**	**1,008**	**9.5%**

Panel B: Influence Dimensions for Relevant Speeches (n=96)

Dimension	Mean	SD	Min	Max	Mode
Temporal Uniqueness (0–3)	1.02	0.84	0	3	1
Marginal Contribution (0–3)	1.42	0.97	0	3	1
Correspondence Strength (0–3)	1.26	0.86	0	3	1

Panel C: Correspondence Score Distribution

Correspondence Score	Frequency	Percentage	Interpretation
0 (No correspondence)	14	14.6%	Addressed topic but no alignment
1 (Tangential)	48	50.0%	Thematic connection only
2 (Clear alignment)	29	30.2%	Close conceptual match
3 (Direct match)	5	5.2%	Exact framework or language

Panel D: Influence Mechanisms (for speeches with correspondence > 0, n=82)

Mechanism	Frequency	Percentage	Example
Problem Framing	45	54.9%	Characterizing interim storage risk
Solution Approach	15	18.3%	Proposing regulatory gates
Technical Details	15	18.3%	Specifying safety thresholds
Conceptual Framework	6	7.3%	Consent-based siting model
Direct Language	1	1.2%	Verbatim text adoption

market with 6–7 equal competitors.[5] The top speaker (Mr. O'Donnell, a state official) accounted for 22.6% of total influence weight, while the top three speakers—O'Donnell, Ms. Tubb (utility representative), and Mr. Victor (policy expert)—collectively accounted for 62.8%. This pattern suggests committee members draw disproportionately from a small set of trusted voices while still incorporating diverse perspectives.

High-Impact Testimony. Truly influential testimony proved rare. Only 14 of 96 speech-change pairs (14.6%) achieved influence weights of 6 or higher (scores of 2+ on all three dimensions). These high-impact moments concentrated among three speakers across two provisions: Ms. Tubb's articulation of the imminent decision framework (SC_004, weight=27), Mr. O'Donnell's emphasis on federal obligations (SC_006, weight=27), and several witnesses' convergent concerns about storage becoming permanent (SC_003, weights of 12–18). The rarity of high-impact testimony underscores that most witness contributions reinforce rather than originate legislative ideas.

Engagement Strategies. The most influential speakers addressed 3–4 changes selectively rather than commenting on all provisions. Mr. Lyman addressed the most changes (5) but achieved lower average influence (3.38) than speakers who focused their testimony. Ms. Tubb achieved the highest average influence (8.00) by concentrating on three high-impact provisions where her utility expertise was most relevant. This pattern suggests effective witnesses target their expertise strategically rather than attempting comprehensive commentary.

4.4.4 Collective Influence Synthesis

The speech-level and speaker-level analyses reveal individual influence moments and their distribution across witnesses. Synthesizing these patterns across all substantive changes reveals how testimony influence operated collectively in H.R. 3053.

Distributed Expertise. Rather than deferring to single authorities, the committee drew from complementary expertise across multiple witnesses. Even in SC_004, where Ms. Tubb's framework achieved the highest influence weight (27), four other speakers contributed substantially to the same provision. The moderate Herfindahl index (0.161) confirms influence distributed across roughly 6–7 equivalent voices rather than concentrating monopolistically.

Sequential Development. The temporal analysis reveals influence operating through iterative refinement rather than simple agenda-setting. Early testimony introduced problems, mid-hearing speakers proposed solutions, and later witnesses provided critical implementation details. Most influential ideas (temporal uniqueness mean=1.02) were raised multiple times and refined through discourse rather than emerging from single originating sources.

Institutional Specialization. Different types of witnesses influenced legislation through distinct mechanisms that align with their institutional roles. Federal agencies operated primarily through technical details (65% of influence instances), providing specifications and regulatory expertise. State and local officials emphasized problem framing (78% of instances), characterizing risks and consequences for their jurisdictions. Think tanks and advocacy organizations focused on solution approaches (52% of instances), proposing strategic frameworks for policy design. Industry representatives used mixed approaches, combining technical expertise with problem identification.

Conceptual Over Textual Influence. Despite extensive testimony, only one instance of direct language adoption occurred across all nine changes. Instead, influence operated through conceptual synthesis—the committee absorbed frameworks, concerns, and approaches from testimony while crafting its own legislative language. This pattern suggests deliberative processing rather than stenographic incorporation of witness proposals.

4.5 Preliminary Evaluation of Theoretical Predictions

While this proof-of-concept analysis examines a single bill-hearing pair and cannot provide definitive empirical tests, the H.R. 3053 case yields preliminary evidence relevant to our three competing theoretical predictions. The patterns observed offer suggestive support for information theory regarding where testi-

[5]For comparison, a perfectly equal distribution among 12 speakers would yield H=0.083, while a monopoly would yield H=1.0.

mony influence operates, strong support for deliberation theory regarding how influence manifests, and no support for position-taking theory in this technical policy domain.

H1: Information Theory Predictions. Information theory predicts that substantive-technical changes will show higher correspondence with testimony than substantive-political changes because committees exist to gather specialized knowledge where information asymmetries are greatest. The evidence provides qualified support for this prediction. All 96 instances of testimony relevance occurred on substantive-technical changes, which achieved a 14.3% relevance rate, while organizational-structural and symbolic-expressive changes attracted zero relevant testimony despite comprising three of nine changes. However, H.R. 3053 contained no substantive-political changes, preventing direct comparison of technical versus political responsiveness to testimony. The committee's decision to avoid distributional issues—state veto power, benefit allocation formulas, jurisdictional assignments—that featured prominently in testimony suggests strategic selection of technically-focused revisions where expertise could plausibly shape outcomes.

H2: Position-Taking Theory Predictions. Position-taking theory predicts that influence will concentrate on symbolic-expressive provisions rather than substantive changes, with patterns reflecting partisan positioning rather than technical expertise. The evidence contradicts these predictions entirely. The symbolic-expressive change (SC_002) removing "Sense of Congress" statements about transportation routes attracted zero testimony despite addressing contentious issues—transportation through the Great Lakes and storage concerns—that generated extensive debate. Instead, all measurable influence concentrated on substantive-technical provisions establishing regulatory frameworks for interim storage. Furthermore, influence patterns showed no correlation with partisan positioning: state officials from both parties achieved similar influence levels when providing relevant technical expertise, while partisan rhetoric about repository politics attracted minimal correspondence.

H3: Deliberation Theory Predictions. Deliberation theory predicts that influential testimony will operate through problem framing and solution approach mechanisms rather than direct language adoption, with influence distributed across multiple speakers based on argument quality rather than formal authority. The evidence strongly supports these predictions. Problem framing dominated influence mechanisms at 54.9% of instances with correspondence, followed by solution approach (18.3%) and technical details (18.3%), while direct language adoption occurred in only 1.2% of cases. Influence distributed across speakers with moderate concentration (Herfindahl index=0.161), equivalent to roughly 6–7 equal participants. The importance of marginal contribution (mean=1.42) relative to temporal uniqueness (mean=1.02) suggests argument quality mattered more than agenda-setting priority. Sequential development patterns—early problem identification, mid-hearing solution proposal, late-hearing refinement—characterize deliberative discourse where ideas evolve through exchange rather than emerge pre-formed.

Synthesis: Mixed Theoretical Support. The H.R. 3053 evidence suggests committees may serve multiple functions depending on the nature of legislative changes. Information theory correctly predicts where influence concentrates—on technical provisions requiring expertise—but deliberation theory better explains how influence operates through conceptual synthesis and distributed argument quality rather than simple expert deference. Position-taking theory finds no support in this technical domain, though this may reflect case selection rather than general inapplicability. The committee's avoidance of substantive-political changes despite extensive testimony about state rights, federal obligations, and distributional concerns suggests strategic boundary-setting: deliberative processes may be reserved for technical questions while political decisions occur through other channels.

This preliminary evidence points toward a conditional theory where committee function varies by issue type. Technical questions may genuinely benefit from deliberative expertise gathering, while political questions may indeed serve position-taking functions as Mayhew suggests. Testing this conditional framework requires systematic analysis across multiple cases encompassing both technical and political legislative changes, as the methodology developed here enables.

5 Discussion

We have developed and demonstrated a computational framework to measure the influence of testimony on legislative changes, addressing a methodological problem that has long constrained empirical research on committee deliberation. By combining mechanical text processing with theory-driven AI analysis, the approach generates structured datasets from hearing transcripts and bill versions that operationalize concepts of deliberative influence. The application to H.R. 3053 produces interpretable measures of influence, showing that testimony shaped technical provisions primarily through problem framing and distributed expertise. This single application serves as a proof of concept, establishing a foundation for larger-scale analysis.

Methodological Contributions and Limitations The primary contribution of this work is methodological. Our framework advances on existing approaches in three ways. First, by identifying specific influence mechanisms, it distinguishes *how* testimony shapes legislation—whether through conceptual reframing, technical specification, or direct language adoption—rather than treating all textual correspondence as equivalent. Second, its multidimensional measurement captures when ideas enter the discourse (temporal uniqueness), what a speech adds (marginal contribution), and how closely testimony aligns with legislative text (correspondence strength), permitting a more granular analysis of deliberative dynamics. Third, the generation of multi-level datasets (change-level, speech-change dyad, and speaker-level) provides the analytical flexibility required to test competing theoretical predictions about committee function.

The principal limitation of the current study is its reliance on a single bill-hearing pair. The findings from H.R. 3053 are illustrative of the method's capabilities, but they cannot be generalized. The patterns observed—strong support for deliberative and informational dynamics in a technical policy domain—may not hold for more politicized issues or in different committee environments. This limitation defines our immediate research agenda, which prioritizes validation and scaling.

A Validation Agenda Establishing the validity of these AI-generated measures is the critical next step before large-scale application. We will pursue a multi-faceted validation strategy. First, to assess **inter-coder reliability**, we will compare the model's classifications and scores against those of expert human coders across a sample of bills from different policy domains. This will allow us to quantify agreement rates and identify any systematic biases in the AI's assessments. Second, for **convergent validity**, we will compare our correspondence scores with other indicators of influence, such as citations of testimony in committee reports, patterns of amendment sponsorship, and whether influential witnesses are invited to testify again on related matters. Third, to establish **construct validity**, we will use factor analysis on a larger sample to confirm that our three influence dimensions—temporal uniqueness, marginal contribution, and correspondence strength—capture distinct concepts rather than a single underlying factor. Finally, we will test for **predictive validity** by assessing whether provisions that show high correspondence with testimony are more likely to survive subsequent legislative stages, including floor consideration and conference committee negotiations.

From Case Study to Large-Scale Analysis The framework is designed for application across hundreds or thousands of bill-hearing-bill triplets, which will permit a move from case-based inference to broad empirical assessment. As demonstrated, the computational cost of approximately $0.05–0.10 per hearing-bill pair makes such analysis feasible. This larger dataset will enable direct tests of the theoretical hypotheses outlined in Section 2. We will be able to determine whether the patterns observed in H.R. 3053 are idiosyncratic or general. For instance, we will test if substantive-technical changes consistently show more correspondence with testimony than substantive-political changes (H1), if influence concentrates on symbolic provisions in highly partisan contexts (H2), and if problem framing is the dominant mechanism of influence across different types of committees and policy areas (H3).

Beyond hypothesis testing, systematic analysis will enable investigation of which actors and contexts generate the most testimony influence. By merging member-level characteristics to the data, we will test whether committee leadership, seniority, majority party status, bill sponsorship, and constituency pressure determine individual legislators' influence on witness incorporation. We will examine how witness types—federal agencies, academic experts, trade associations, advocacy groups—vary in their ability to shape legislative provisions. We will assess under what conditions hearings make a difference in policy outputs: whether polarized, partisan issues show reduced testimony influence compared to technical issues where bipartisan collaboration on revisions may be more common.

Strengthening Causal Inference While tracing influence from testimony to text provides strong correlational evidence, we will seek opportunities to strengthen causal inference. We will identify natural experiments where similar or identical bills are considered concurrently in House and Senate committees. Because these committees often hear from different sets of witnesses, this variation in deliberative input provides leverage for isolating the effect of testimony content on legislative outcomes.[6] A second approach involves longitudinal analysis, tracking how specific provisions in a bill evolve across multiple hearings. This design allows us to test whether an increase in testimony correspondence for a provision predicts its survival and refinement through successive stages of the legislative process.

6 Conclusion

This paper has presented a computational framework that addresses a fundamental empirical challenge in legislative studies: measuring how congressional testimony shapes policy outcomes. By combining mechanical text processing with AI-powered semantic analysis, we operationalize theoretical concepts about deliberative influence that have long resisted systematic measurement. The approach transforms unstructured hearing transcripts and bill versions into structured datasets capturing multidimensional influence patterns, enabling direct tests of competing theories about committee function.

The proof-of-concept application to H.R. 3053 demonstrates the methodology's capacity to generate theoretically meaningful patterns. The concentration of influence on substantive-technical changes, the dominance of problem framing over direct language adoption, and the distribution of influence across multiple witnesses rather than single authorities all provide preliminary evidence about how committees process expert testimony. While these patterns cannot resolve theoretical debates from a single case, they establish that computational methods can now measure the complex influence dynamics that distinguish information gathering from position-taking from genuine deliberation.

The scalability of this approach—at approximately \$0.05–0.10 per hearing-bill pair—represents a different kind of computational advance. While previous large-scale text methods could identify thematic patterns or measure similarity across legislative documents, they could not replicate the nuanced qualitative content analysis required to trace influence mechanisms. Our framework enables researchers to assess at scale whether testimony's problem framing shaped legislative approaches, whether conceptual frameworks structured provisions, or whether technical specifications influenced statutory language—distinctions that keyword matching or topic models cannot capture. This shifts the empirical frontier from counting what is discussed to measuring how ideas shape outcomes across entire policy domains and multiple congresses.

Important methodological work remains. Systematic validation through inter-coder reliability studies will establish the accuracy of AI-generated measures. Convergent validity tests comparing our measures with alternative indicators will confirm we capture genuine influence rather than artifacts. Application across hundreds of cases will reveal whether the patterns observed in nuclear waste policy generalize to other domains or represent case-specific dynamics.

By providing empirical tools to measure deliberative influence at scale, this research opens new possibilities for understanding democratic institutions. If systematic analysis confirms that committees engage in genuine conceptual synthesis—as the H.R. 3053 evidence suggests—institutional reforms could focus on enhancing deliberative capacity through longer hearings, more diverse witness pools, and structures that reward argument quality. If influence patterns vary systematically by issue type, different institutional designs might be optimal for technical versus political questions. Most fundamentally, we can finally move from theoretical speculation to evidence-based understanding of when, how, and for whom testimony matters in the making of American law.

[6] The House sometimes refers bills to multiple committees, which could enable comparison of different committee hearings on the same bill. However, committees often consider bills sequentially, and when they do consider the same version simultaneously, the primary committee typically incorporates suggestions from other committees before reporting. Thus, direct comparison of resulting bill texts across different House committees is generally not possible.

References

Bosley, M. (2025). Towards qualitative measurement at scale: A prompt-engineering framework for large-scale analysis of deliberative quality in parliamentary debates. *Journal of Political Institutions and Political Economy*, 6:1–29.

Bussing, A., Patton, W., Roberts, T., and Treul, S. (2015). Hearing from all sides: How legislative testimony influences state level policy-makers in the united states. *International Journal of Public Health*, 60(3):375–380.

Esterling, K. M. (2007). Buying expertise: Campaign contributions and attention to policy analysis in congressional committees. *American Political Science Review*, 101(1):93–109.

Habermas, J. (1996). *Between Facts and Norms: Contributions to a Discourse Theory of Law and Democracy.* MIT Press, Cambridge, MA.

Irani, A., Park, J. Y., Esterling, K. M., and Faloutsos, M. (2024). What is being argued (wiba)? an application to legislative deliberation in the u.s. congress. *arXiv preprint arXiv:2407.06149*.

Krehbiel, K. (1991). *Information and Legislative Organization.* University of Michigan Press, Ann Arbor, MI.

Lewison, K., Crowley, D. M., and Scott, J. T. (2020). Social scientists' testimony before congress in the united states between 1946-2016, trends from a new dataset. *PLOS ONE*, 15(3):e0230104.

Mayhew, D. R. (1974). *Congress: The Electoral Connection.* Yale University Press, New Haven, CT.

Park, J. Y. (2022). *Hearings on the Hill: The Politics of Informing Congress.* Cambridge University Press, Cambridge, UK.

Roberts, J. M. and Smith, S. S. (2003). Procedural contexts, party strategy, and conditional party voting in the u.s. house of representatives, 1971-2000. *American Journal of Political Science*, 47(2):305–317.

A Change-Specific Influence Patterns

This appendix provides detailed analysis of influence patterns for each semantic change identified in H.R. 3053.

A.1 SC_003: Storage Prohibition

39 relevant speeches, average correspondence 1.36

This core provision attracted the broadest engagement, with witnesses from every stakeholder category addressing it. The influence pattern showed clear temporal development: early testimony (speeches 1–15) established the problem of interim storage potentially becoming permanent, citing examples from other countries where "temporary" facilities operated for decades. Mid-hearing testimony (speeches 16–75) introduced various solutions, from hard deadlines to consent requirements. Ms. Tubb's testimony (speech 62) proved pivotal, explicitly articulating the concept of linking storage to repository progress—a framing that directly manifests in the prohibition structure. The collective synthesis showed distributed influence with no single dominant voice, but rather a convergence of concerns from safety advocates (Mr. Issa), state officials (Mr. O'Donnell), and policy experts (Mr. Victor) that collectively shaped the provision.

A.2 SC_004: Imminent Decision Exception

22 relevant speeches, average correspondence 1.09

Fewer witnesses addressed this nuanced exception, with influence concentrating among those with regulatory expertise. The temporal pattern revealed cascading refinement: initial testimony identified the need for flexibility, but Ms. Tubb first proposed the specific "imminent decision" framework that appears in the bill. Subsequent speakers refined this concept, with agency witnesses providing technical input on reasonable timeframes. The 10-year contract limitation emerged from synthesis of multiple testimonies discussing typical storage contract durations. Marginal contribution scores exceeded temporal uniqueness here, as later speakers who added implementation details scored higher than those who first raised the general need for exceptions.

A.3 SC_005: Notification Requirements

5 relevant speeches, average correspondence 1.00

This procedural provision attracted minimal direct testimony, with influence operating indirectly. No witness explicitly called for notification requirements, but five speeches emphasized accountability and congressional oversight in the context of storage decisions. The influence mechanism was entirely problem framing—witnesses identified the oversight gap without proposing specific notification procedures. The low engagement suggests this provision emerged more from legislative drafting conventions than direct testimony influence.

A.4 SC_006: Federal Obligations Clarification

29 relevant speeches, average correspondence 1.34

Despite its low substantive impact rating, this clarification attracted substantial testimony, revealing witnesses' anxiety about long-term federal responsibilities. The influence pattern showed reinforcement rather than development: multiple witnesses independently emphasized that interim storage must not become an excuse to abandon permanent disposal obligations. State officials particularly stressed this point, with Mr. O'Donnell and others citing concerns about their states becoming de facto permanent storage sites. The high speech count but moderate correspondence suggests the provision addressed witness concerns without adopting their specific language or proposals.

A.5 SC_007: Technical Definitions

1 relevant speech, average correspondence 0.00

Only one speech tangentially mentioned the need for clear definitions, and it showed no correspondence with the actual definitions adopted. This pattern confirms that technical definitions emerged from legislative drafting rather than testimony influence.

A.6 SC_008: Conforming Amendment

0 relevant speeches, average correspondence 0.00

No witness addressed table of contents updates, confirming this as a purely technical drafting matter.