Mapping Reviewable Discretion in Access-to-Information Laws: A Social Computing Framework for Multilingual Legal Text Measurement

Anonymous Author(s)

Anonymous submission for ICSC 2026

Abstract. Access-to-information laws specify request, refusal, appeal, oversight, and remedy clauses. This paper tests whether multilingual semantic embeddings can measure reviewable discretion from statutory text without training on legal-strength scores. It defines *reviewable discretion* through six obligations: request, boundary, reason, challenge, review, and enforcement. Using source-validated access-to-information law texts from 115 sovereign states, the paper builds a five-stage, seven-dimension measure. Qwen V4 embeddings retrieve passages for legal-operation anchors; latent semantic analysis provides a baseline; artifact adjustment and clause audit test validity. The artifact-adjusted index converges with RTI legal coding scores ($r = 0.535$), and dimension-level correlations range from $r = 0.408$ to $r = 0.639$. A stage-level diagnostic shows how score-derived stage coverage orders statutes along a procedural reviewability gradient. The contribution is a reproducible workflow linking embedding scores to statutory clauses, benchmark correlations, artifact adjustment, and audit categories.

Keywords: Social computing · Social applications of large language models · Computational legal measurement · Legal text measurement · Semantic embeddings

1 Introduction

Access-to-information (ATI) statutes are public records of how states make administrative secrecy reviewable. They specify who may request information, when an authority must respond, how a refusal must be justified, where a requester can complain, and what happens when officials obstruct access. These clauses turn discretion into a legal process. They also create a demanding text-measurement problem: the institutional signal lies in operative duties, rights, exceptions, procedures, oversight powers, and remedies rather than in topical vocabulary alone.

This paper develops a multilingual semantic-embedding approach to measuring reviewable discretion in access-to-information (ATI) laws. Reviewable discretion is defined here as a statutory structure that turns administrative choice into

a legally traceable process. In the ATI domain, discretion becomes reviewable when a requester can file a request, the authority must decide within bounded conditions, a refusal must be reasoned, the decision can be challenged, an external body can review compliance, and delay or obstruction can trigger consequences. These components are not decorative clauses. They are the procedural grammar through which a transparency statute makes administrative power observable and contestable.

The empirical object is a corpus of ATI and freedom-of-information statutes from 115 sovereign states. The method combines pretrained multilingual embeddings, theory-anchored semantic anchors, passage segmentation, a latent semantic analysis (LSA) baseline, convergence checks against independently produced RTI legal coding scores, and legal-passage audit. The design follows the measurement logic that a valid indicator begins with a specified construct and documented translation from concept to observation [1]. Text-as-data research adds the requirement that preprocessing, representation, and validation remain visible to the reader [9, 10]. Computational social science contributes the broader commitment to scalable institutional measurement from digital traces with transparent data access, metadata documentation, and auditability [15, 16]. For social computing, the paper treats law as a formal public record and supplies an auditable embedding-based measurement protocol for that class of data.

The paper makes three contributions. First, it specifies reviewable discretion as a text-measurable institutional construct. This moves beyond jurisdictional labels and one-number legal-strength scores by observing the internal structure of statutory text directly. Second, it presents a reproducible measurement protocol for multilingual legal texts: semantic anchors define the target construct, embeddings recover candidate passages, LSA provides an interpretable comparator, and clause audit verifies that high-scoring passages express operative legal content. Third, it reports a stage-level measurement diagnostic. The semantic index converges with RTI legal coding scores, the five stage scores produce a readable procedural gradient, and artifact covariates are documented before benchmark validation.

The relevant unit is a legal passage that may contain a right, duty, exception, procedure, or remedy. The design shows how multilingual embeddings can support this kind of measurement while remaining inspectable: every country score can be traced back to semantic anchors, ranked passages, baseline scores, and audit categories.

This also connects the paper to computational legal studies. Machine-learning systems can model legal behavior when the task, features, and validation target are clearly defined [2, 11]. Recent legal-language benchmarks and domain-adapted encoders sharpen that lesson for legal text [4, 5]. The present paper addresses a different task: measuring the internal structure of public law across countries. A legal text model should travel across time, jurisdictions, and drafting styles without becoming a keyword list for a single legal system. The five-stage reviewability schema is designed for that purpose.

2 Conceptual Target: Reviewable Discretion

2.1 Why Discretion Must Be Measured as Structure

Administrative discretion has two analytically distinct meanings. It can indicate unbounded official choice, but it can also indicate legally necessary flexibility. A statute that eliminates all discretion may be rigid and unusable. A statute that grants discretion without review may create arbitrary power. The institutional question is how statutes make discretion reviewable through requests, reasons, challenges, supervision, and consequences.

Procedural scholarship provides the logic for this construct. Administrative procedures can be instruments of political control because they allocate information, participation, timing, and monitoring costs [18]. Accountability in regulatory governance is a set of answerability and review relations among agencies, oversight bodies, regulated actors, courts, and publics [23]. In this paper, reviewable discretion is an operational measurement construct derived from procedure and accountability theory. The text is expected to reveal whether the legal process creates a trail from request to decision, from decision to challenge, and from challenge to external review or consequence.

The construct also avoids a common measurement shortcut in institutional research. Governance involves administrative capacity, autonomy, and the organization of public authority as well as formal constraints [8]. The measure targets one antecedent legal structure: whether the statute builds a reviewable trail around administrative discretion. This scope gives the indicator its analytic strength by identifying statutory design directly and separating it from implementation performance.

ATI statutes fit this measurement problem. Transparency research treats disclosure as a process in which information must be produced, used, acted upon, and answered by institutions [12, 19]. The statute is the first layer of that process. It sets the legal route through which requesters and oversight institutions can act. A computational measure of ATI statutes should therefore recover where the text locates control.

2.2 Five Stages and Seven Dimensions

The construct is operationalized in five stages. *Entry access* records whether the law enables a person to request information through identifiable channels and reasonable procedural access. *Decision bounds* records whether the authority's response is constrained by deadlines, permissible forms, exception limits, and rules for handling deficient requests. *Challenge channel* records whether the requester can contest refusal, delay, silence, fee decisions, or partial disclosure. *External supervision* records whether a commissioner, court, tribunal, ombudsman, information commission, or other body can review or supervise compliance. *Consequence enforcement* records whether the statute attaches sanctions, remedies, disciplinary consequences, compliance orders, or equivalent consequences to non-compliance.

The five stages are operationalized through seven semantic dimensions. Request procedure, exception boundary, deadline specificity, reason giving, appeal and review, independent oversight, and sanction or remedy are measured separately before being mapped into the stage structure. This distinction matters. The seven dimensions capture semantic content. The five stages capture procedural position in the review chain. A country may have strong request procedures and deadlines but weak external enforcement. Another may have a strong appeal route but thin rules on initial access. A single total legal score can hide these configurations.

The measurement target is the internal design by which a statute organizes official discretion into reviewable form. Table 1 summarizes the construct mapping used before model scoring. The schema can be compared across legal traditions, languages, and statutory styles because it is anchored in functional legal operations rather than in a single jurisdiction's vocabulary.

Table 1. Construct mapping from legal operation to stage score

Stage	Legal operation	Semantic dimensions	Nearest RTI comparator
Entry access	request can be filed and assisted	request procedure	requesting procedure
Decision bounds	authority must decide under bounded rules	exception boundary; deadline; reason giving	refusals; total
Challenge channel	refusal, delay, silence, or fee can be contested	appeal/review	appeals
External supervision	independent body or court can review compliance	independent oversight	total
Consequence enforcement	non-compliance can trigger orders, sanctions, or remedies	sanction/remedy	sanctions

Note. RTI total is used where no direct component matches a stage.

3 Corpus and Measurement Design

3.1 Corpus

The corpus consists of source-validated ATI or freedom-of-information law texts for 115 sovereign states. The country scope follows a sovereign-state sample after entity screening: non-state entities are removed, and each admitted observation has a country identifier, source link, extracted text, and source validation record. The text object is the statute or principal legal text used by the RTI Rating project for legal evaluation.[1] Semantic scores are computed from statutory text

[1] `www.rti-rating.org`.

before RTI enters the analysis. RTI is independent in coding organization and scoring procedure while covering the same legal domain and text objects; it is therefore used as a convergent-validity benchmark, not as a training label or an external test of implementation effects.

Table 2. Corpus and measurement frame

Item	Value	Note
Unit of analysis	country-text cross-section	sovereign-state sample after entity screening
Sample size	115	United Nations member-state scope after non-state exclusions
Text object	access-to-information law texts	RTI / freedom-of-information legal texts
Reviewability stages	5	entry access, decision bounds, challenge channel, external oversight, consequence enforcement
Semantic dimensions	7	request, exception, deadline, reason, appeal, oversight, sanction/remedy
Source-language validation cases	3	pre-RTI checks triggered by extraction, script, or language ambiguity
Qwen model	DashScope text-embedding-v4, 1024 dimensions	normalized embeddings and cosine similarity
Baseline model	TF-IDF and truncated-SVD LSA	interpretable semantic-space comparator

The corpus uses a source-validated legal-text frame. It includes statutes for which legal text could be source-validated, extracted, segmented, and audited. This choice increases construct validity by assigning semantic scores only where the document link, extracted text, and legal identity of the statute are verifiable.

This sample rule is important for social computing research on public records. Statutory corpora can contain amended laws, repealed versions, unofficial translations, scanned PDFs, multilingual appendices, metadata pages, or consolidated texts with unclear legal status. The present pipeline treats document identity as a measurement step. Source validation, text extraction, and entity screening are part of the data-generating process.

3.2 Measurement Pipeline

Figure 1 summarizes the pipeline. First, source-validated ATI legal texts are segmented into 220-word passages with 50-word overlap; shorter extracted texts are

split by character windows when word tokenization is unreliable. Second, semantic anchors are written for each of the seven dimensions. The final anchor table contains eleven reviewability anchors per dimension and ten negative-control anchors. Anchors describe legal operations rather than keywords. For example, a reason-giving anchor refers to a public authority's obligation to explain refusal, delay, partial access, or exemption decisions. A sanction/remedy anchor refers to penalties, compliance orders, disciplinary consequences, or remedies for non-compliance.

Table 3. Anchor design examples

Dimension	Positive reviewability anchor	Negative-control contrast
Request procedure	A requester may submit an information request in writing, electronically, or orally to the designated information officer.	maritime vessel registration and harbor fees
Reason giving	A refusal must be accompanied by written reasons identifying the legal grounds for nondisclosure.	patent and trademark registration
Sanction or remedy	The law specifies sanctions, offences, fines, discipline, or penalties for unlawful refusal or obstruction.	wildlife conservation and habitat permits

Third, Qwen V4 embeddings are generated for both legal passages and anchors through Alibaba Cloud DashScope `text-embedding-v4` on May 25, 2026, with 1024-dimensional normalized vectors. Cosine similarity identifies passages that are semantically close to each dimension. For passage p, dimension d, and anchors A_d, the passage score is $s_{p,d} = \max_{a \in A_d} \cos(e_p, e_a)$. Fourth, the country-dimension score $D_{c,d}$ is the mean of the top five passage scores for country c and dimension d. Fifth, artifact adjustment residualizes the overall semantic index against document length, English function-word share, Roman-script share, and negative-control-anchor similarity. Sixth, the same country corpus is represented through TF-IDF and truncated-SVD LSA as a transparent baseline. Vector-space models remain useful because they offer reproducible semantic representations based on co-occurrence and indexing [22, 25]. LSA supplies a low-parameter dimensional-reduction comparator for the embedding measure [7, 14]. Finally, retrieved passages are audited for operative legal content.

The embedding model supplies passage-level semantic retrieval within a specified measurement design. Sentence-embedding research shows how semantic representations travel across paraphrase boundaries [21]. Multilingual embedding research extends that logic across languages [3, 6]. Sociological embedding research shows that semantic geometry can recover relational meaning in social text [13]. Comparative-politics text analysis emphasizes task-specific validation rather than generic accuracy claims [17]. The present design therefore makes the

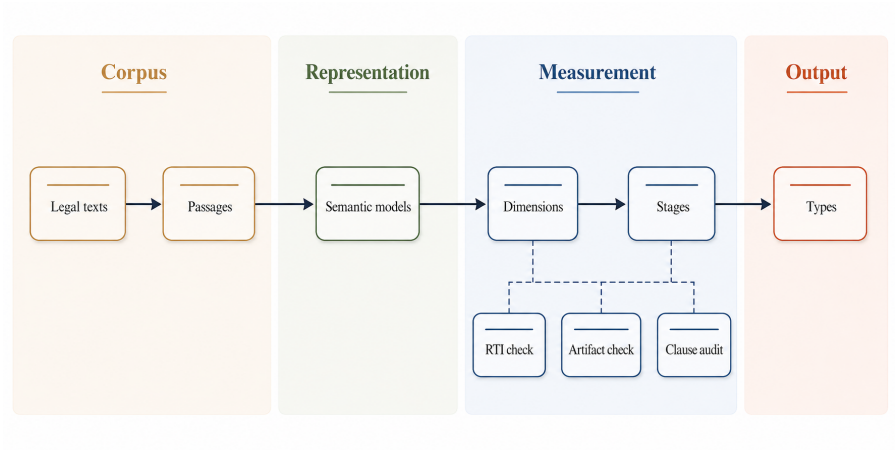

Fig. 1. Measurement architecture linking source-validated texts, passage segmentation, semantic representation, dimension and stage scores, output types, and validation gates.

embedding model accountable to four checks: theoretically specified anchors, LSA comparison, convergence with RTI legal coding benchmarks, and clause-level audit.

The scoring design uses two aggregation levels. At the passage level, each text segment receives similarity scores to each semantic anchor. At the country level, the seven $D_{c,d}$ scores are averaged into an overall index and mapped into five stage scores by the construct table. The reported index is the standardized residual from a linear adjustment for log word count, English function-word share, Roman-script share, and negative-control-anchor similarity. Stage indicators are created from adjusted stage scores: $I_{c,k} = 1$ when stage k reaches the corpus median. Source-language checks were triggered by extraction, script, or language ambiguity before RTI comparison and without using RTI scores. The stage-coverage label is a score-derived diagnostic count: 0–1 stages form lower coverage, 2–3 stages middle coverage, and 4–5 stages higher coverage. The labels describe relative textual coverage in this corpus, not legal sufficiency. Top-k diagnostics test aggregation robustness on the unadjusted retrieval score; validity tests use the artifact-adjusted index. A 45th/50th/55th percentile threshold check preserves ordered mean RTI scores: lower coverage remains near 72.8–73.6, middle coverage near 85.3–89.6, and higher coverage near 96.9–98.0.

A dimension-stage architecture keeps the audit trail visible. A single overall embedding score would collapse request access, exceptions, deadlines, appeals, oversight, and remedies into one number. Under the reported rule, a higher-coverage classification requires supporting passages in the entry, decision, challenge, supervision, and consequence parts of the chain.

3.3 Validity Strategy

The validity strategy has four parts. Following the shared validity standard for qualitative and quantitative measurement [1], construct validity is addressed by defining reviewable discretion as a procedural legal structure and mapping each dimension to a legal operation. Convergent validity is addressed by comparing semantic scores with RTI legal coding benchmarks. Artifact adjustment is documented for document length, English function-word share, script share, and negative-control-anchor similarity; the near-zero post-adjustment correlations are diagnostics of that adjustment, not independent validity evidence. Legal-passage validity is addressed by inspecting retrieved clauses.

The audit rule is restrictive. A high-scoring passage counts as valid evidence when it contains operative legal language: duties, rights, procedures, exceptions, appeal routes, oversight powers, sanctions, or remedies. Headings, table-of-contents lines, definitions alone, translation notes, boilerplate, and broad declarations are excluded from direct evidence. This audit design follows computational grounded theory's insistence that computational pattern discovery must remain connected to substantive interpretation and iterative validation [20].

This is also a response to a broader problem in institutional measurement. Rule-of-law research shows that different measures bearing similar labels can produce different empirical interpretations [24]. The same risk appears in legal-text measurement. A measure called "oversight" may refer to the name of an information commission, its appointment rules, its jurisdiction, its binding-order powers, or its reporting duties. The present approach keeps the label tied to retrieved legal passages and stage definitions, reducing the risk that a variable name carries more institutional content than the evidence supports.

4 Results

4.1 The Semantic Measure Recovers Legal-Text Structure

The first result is convergent validity. The overall semantic reviewability index correlates with RTI total scores at $r = 0.535$ in the 115-country corpus. Because RTI enters only after semantic scoring, this association evaluates convergence with an independently produced coding benchmark for the same legal domain rather than model training fit or implementation effects. The LSA baseline correlation is $r = 0.401$, indicating that Qwen V4 and LSA recover overlapping but not identical semantic structure. LSA captures broad co-occurrence structure, while transformer embeddings better recover legal paraphrase across languages and drafting styles.

Dimension-level correlations with the closest RTI components range from $r = 0.408$ to $r = 0.639$. Independent oversight and deadline specificity show the highest correlations, at $r = 0.639$ and $r = 0.627$. Sanction or remedy has $r = 0.572$, reason giving has $r = 0.554$, appeal or review has $r = 0.463$, request procedure has $r = 0.462$, and exception boundary has $r = 0.408$. These magnitudes are meaningful because the model was not trained on RTI scores. The correlations arise from semantic retrieval of legal passages.

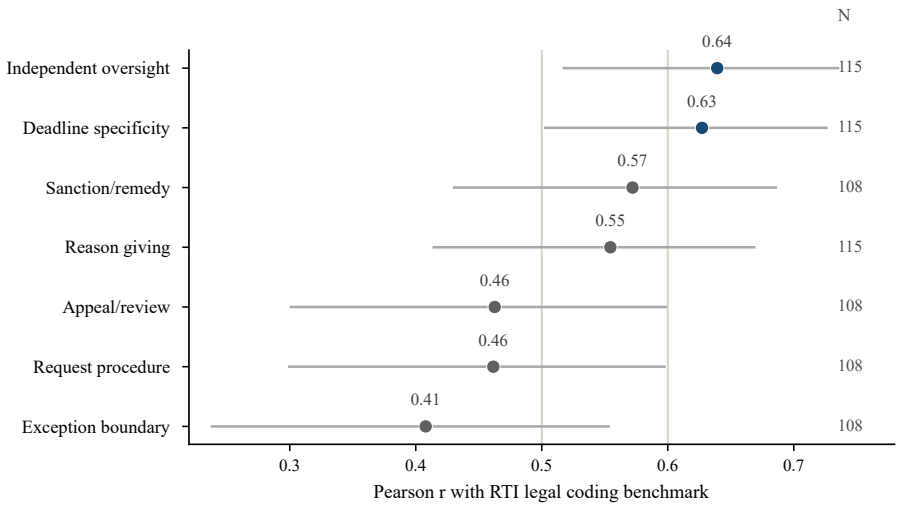

Fig. 2. Dimension-level convergence between semantic scores and RTI legal coding benchmarks. Dots report Pearson r; intervals are approximate 95 percent CIs; labels give country counts.

Table 4. Measurement-validity dashboard

Validity target	Evidence	Result
Construct mapping	pre-specified five-stage and seven-dimension schema	scores are tied to legal operations rather than generic statutory detail
Same-domain convergence	semantic index compared with RTI total after scoring	$N = 115$, Pearson $r = 0.535$
Model baseline	Qwen V4 compared with LSA on the same corpus	$N = 115$, Pearson $r = 0.401$
Dimension validity	seven RTI component checks	Pearson $r = 0.408$–0.639
Artifact adjustment	length, language-script, and negative-control covariates used before validation	post-adjustment correlations with adjustment covariates are near zero by design
Clause audit	retrieved passages checked against operative legal content	1,062 unique candidate snippets; 1,395 retained snippet-dimension annotations

The semantic measure captures statutory reviewability rather than general document richness. It converges with independently produced RTI legal coding scores after artifact adjustment. The post-adjustment near-zero correlations with word count, language-script proxies, and negative-control-anchor similarity document that the adjustment has been implemented; the validity claim rests on construct mapping, RTI convergence, model comparison, threshold stability, and clause audit.

4.2 Stage-Level Interpretability

The five-stage measure also passes an interpretability test. Stage-coverage groups are derived only from semantic stage evidence: zero or one present stage forms lower stage coverage, two or three stages form middle stage coverage, and four or five stages form higher stage coverage. These groups are relative score-derived diagnostics of textual structure, not legal sufficiency judgments. RTI scores are then used as a benchmark to test whether these semantic stage scores recover a legally meaningful ordering.

Figure 3 shows the continuous score layer behind the stage indicators. Each panel holds one procedural stage fixed and plots the country-level semantic stage score against the RTI legal coding score. Because the coverage labels are score-derived, separation along the semantic-score axis is descriptive. The validation evidence comes from the ordered alignment on the RTI benchmark axis and from the threshold check. Higher-coverage countries appear higher on RTI, lower-coverage countries concentrate lower, and middle-coverage countries occupy the middle of the benchmark space. This pattern matters because the stage scores and coverage groups are constructed before RTI enters the analysis.

This stage-level result shows that the pipeline recovers a structured legal signal. The measure distinguishes where reviewability is concentrated within the legal procedure: at entry, decision, challenge, supervision, or consequence. That distinction supplies the comparative leverage needed to study legal design as an institutional structure with internally observable procedural positions.

4.3 What the Model Retrieves

The legal-passage audit shows that high-scoring passages correspond to operative legal clauses. Request-procedure passages commonly specify written, oral, postal, in-person, fax, email, or electronic requests; assistance for persons who need help submitting a request; forms; receipts; transfer rules; and duties to clarify unclear requests. Decision-bound passages specify deadlines, response notices, fee rules, public-interest tests, exception boundaries, and duties to identify whether the information is held. Challenge-channel passages specify appeals, internal review, complaint routes, judicial review, commissioner review, and time limits for challenging refusal or silence. Oversight passages specify the authority, powers, independence, reporting duties, inspection rights, and recommendations or orders of information bodies. Consequence passages specify offences, fines, disciplinary measures, compliance orders, compensation, or remedies.

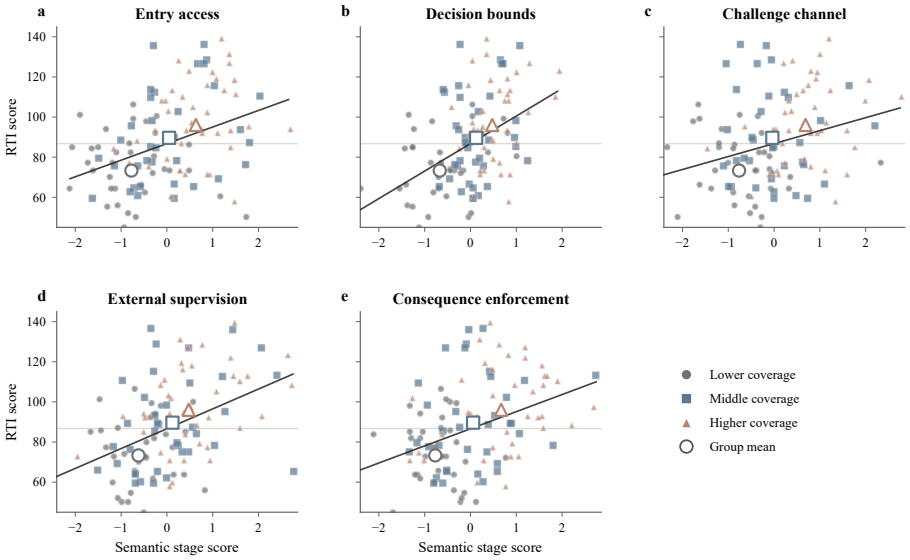

Fig. 3. Continuous stage scores by procedural stage. Points are countries; x-axis reports semantic stage score and y-axis RTI legal coding score. Colors and markers show score-derived stage-coverage groups; hollow markers report group means.

Table 5. Examples from retrieved operative clauses

Stage	Retrieved legal signal	Operative relation
Entry access	"A request can be made in an oral or written form, including electronically"	creates a lawful filing route
Challenge channel	"The requester shall have the right to initiate a complaint"	creates a route for contesting non-disclosure
External supervision	"power ... to summon and enforce the appearance of persons"	gives an oversight body investigatory authority

Table 6. Audit-flow counts for retrieved legal passages

Count level	Meaning for validation	Count	Countries
Unique candidate snippets	retrieved snippets entering audit screening	1,062	108
Operative article/clause annotations	retained evidence with article- or clause-level operation	1,041	96
Operative clause annotations	retained clause-level statutory evidence	249	68
Source-language correction annotations	retained source-language validation evidence	105	3

Counts are reported at two levels: unique retrieved snippets entering audit and snippet-dimension annotations retained as evidence. The audit also identifies predictable exclusion categories: definitions without operative duties, headings, table-of-contents fragments, translation notes, page headers, and PDF extraction noise. The final frame records source-language statutory evidence for three validation cases and preserves audit categories for retrieved snippets.

The audit results matter for theory. A semantic model that retrieves a definition of "public body" measures reviewability when the passage also says who can request, who must decide, how refusal is justified, or how non-compliance is challenged. Similarity to a concept becomes evidence when the passage contains an operative legal relation.

4.4 Model Agreement Across Representations

The Qwen V4 and LSA comparison clarifies the measurement claim. The models agree enough to support a common legal-text signal while preserving useful differences between model families. The overall Qwen score correlates with the LSA baseline at $r = 0.401$. LSA anchors the analysis in a transparent semantic-space representation. Qwen V4 improves multilingual and paraphrase-sensitive retrieval. Their overlap identifies stable reviewability patterns; their divergence marks cases where drafting style, translation, or legal formula requires closer inspection. In a cross-national corpus, translation, document formatting, and statutory drafting can change the surface distribution of words. A baseline model, a residualization step, and a passage audit together show that the reported country-stage scores are supported by multiple checks.

5 Discussion

5.1 Why the Measure Adds Information

The measurement approach differs from existing legal-strength scores in three ways. First, it decomposes legal strength into procedural stages. Two countries can have similar total legal scores while placing reviewability in different parts of the process. One statute may be strong on deadlines and request access but weak on external review. Another may have an information commission but weak rules on reasons and remedies. The five-stage chain makes these differences visible.

Second, each score can be traced back to legal passages, so a reviewer can inspect the clauses that drive a country's score. Third, the approach travels across languages and drafting traditions. Legal-family labels identify broad historical backgrounds, while semantic measurement observes statutory design directly. This is especially useful because legal systems borrow statutory templates, translate model laws, and combine administrative-law vocabularies from different traditions. The method captures textual configurations that historical family labels alone do not observe.

5.2 Methodological Implications for Social Computing

The paper offers a design principle for social computing: large language models and embedding systems become measurement instruments when they operate inside a visible research design. The construct is specified before modeling; anchors are written in institutional language; a simpler semantic baseline is retained; RTI legal coding scores test convergence; and retrieved passages are audited at the clause level.

This principle is particularly important for statutes and legal records. Law contains formulaic language, cross-references, definitions, headings, translation artifacts, and repeated phrases. Embeddings work best when the target is the institutional operation expressed by a clause. Construct specification, audit trails, artifact controls, and convergent benchmarks turn the LLM from a search accelerator into a social measurement instrument. For ICSC's broader social computing audience, this workflow extends social computing from expression and interaction data to formal institutional records.

6 Conclusion

This paper introduces a multilingual semantic-embedding method for measuring reviewable discretion in access-to-information laws. It defines reviewable discretion as a statutory procedure that makes administrative choice requestable, bounded, reasoned, challengeable, externally reviewable, and enforceable. It operationalizes the construct through seven semantic dimensions and five procedural stages, applies the measure to 115 sovereign states, and validates it through RTI convergence, LSA comparison, residualized quality controls, language checks, and legal-passage audit.

The results establish the feasibility and value of the approach. The semantic index shows meaningful convergence with independently produced legal coding benchmarks. The central result is that ATI laws differ in where they place reviewability within the procedural chain. The score-derived stage labels separate lower-, middle-, and higher-coverage patterns, and the retrieved passages correspond to operative legal clauses. These findings support a broader claim for social computing: transformer embeddings can support auditable measurement of statutes when they are tied to explicit constructs, convergent benchmarks, interpretable baselines, and text evidence.

References

1. Adcock, R., Collier, D.: Measurement validity: A shared standard for qualitative and quantitative research. American Political Science Review **95**(3), 529–546 (2001). https://doi.org/10.1017/S0003055401003100
2. Aletras, N., Tsarapatsanis, D., Preoţiuc-Pietro, D., Lampos, V.: Predicting judicial decisions of the european court of human rights: A natural language processing perspective. PeerJ Computer Science **2**, e93 (2016). https://doi.org/10.7717/peerj-cs.93

3. Artetxe, M., Schwenk, H.: Massively multilingual sentence embeddings for zero-shot cross-lingual transfer and beyond. Transactions of the Association for Computational Linguistics **7**, 597–610 (2019). https://doi.org/10.1162/tacl_a_00288
4. Chalkidis, I., Fergadiotis, M., Malakasiotis, P., Aletras, N., Androutsopoulos, I.: LEGAL-BERT: The muppets straight out of law school. In: Findings of the Association for Computational Linguistics: EMNLP 2020. pp. 2898–2904 (2020). https://doi.org/10.18653/v1/2020.findings-emnlp.261
5. Chalkidis, I., Jana, A., Hartung, D., Bommarito, M., Androutsopoulos, I., Katz, D.M., Aletras, N.: LexGLUE: A benchmark dataset for legal language understanding in english. In: Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics. pp. 4310–4330 (2022). https://doi.org/10.18653/v1/2022.acl-long.297
6. Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., Grave, E., Ott, M., Zettlemoyer, L., Stoyanov, V.: Unsupervised cross-lingual representation learning at scale. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. pp. 8440–8451 (2020). https://doi.org/10.18653/v1/2020.acl-main.747
7. Deerwester, S., Dumais, S.T., Furnas, G.W., Landauer, T.K., Harshman, R.: Indexing by latent semantic analysis. Journal of the American Society for Information Science **41**(6), 391–407 (1990). https://doi.org/10.1002/(SICI)1097-4571(199009)41:6<391::AID-ASI1>3.0.CO;2-9
8. Fukuyama, F.: What is governance? Governance **26**(3), 347–368 (2013). https://doi.org/10.1111/gove.12035
9. Gentzkow, M., Kelly, B., Taddy, M.: Text as data. Journal of Economic Literature **57**(3), 535–574 (2019). https://doi.org/10.1257/jel.20181020
10. Grimmer, J., Stewart, B.M.: Text as data: The promise and pitfalls of automatic content analysis methods for political texts. Political Analysis **21**(3), 267–297 (2013). https://doi.org/10.1093/pan/mps028
11. Katz, D.M., Bommarito, M.J., Blackman, J.: A general approach for predicting the behavior of the supreme court of the united states. PLOS ONE **12**(4), e0174698 (2017). https://doi.org/10.1371/journal.pone.0174698
12. Kosack, S., Fung, A.: Does transparency improve governance? Annual Review of Political Science **17**, 65–87 (2014). https://doi.org/10.1146/annurev-polisci-032210-144356
13. Kozlowski, A.C., Taddy, M., Evans, J.A.: The geometry of culture: Analyzing the meanings of class through word embeddings. American Sociological Review **84**(5), 905–949 (2019). https://doi.org/10.1177/0003122419877135
14. Landauer, T.K., Foltz, P.W., Laham, D.: An introduction to latent semantic analysis. Discourse Processes **25**(2–3), 259–284 (1998). https://doi.org/10.1080/01638539809545028
15. Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabási, A.L., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., Jebara, T., King, G., Macy, M., Roy, D., Van Alstyne, M.: Computational social science. Science **323**(5915), 721–723 (2009). https://doi.org/10.1126/science.1167742
16. Lazer, D.M.J., Hargittai, E., Freelon, D., Gonzalez-Bailon, S., Munger, K., Ognyanova, K., Radford, J.: Computational social science: Obstacles and opportunities. Science **369**(6507), 1060–1062 (2020). https://doi.org/10.1126/science.aaz8170
17. Lucas, C., Nielsen, R.A., Roberts, M.E., Stewart, B.M., Storer, A., Tingley, D.: Computer-assisted text analysis for comparative politics. Political Analysis **23**(2), 254–277 (2015). https://doi.org/10.1093/pan/mpu019

18. McCubbins, M.D., Noll, R.G., Weingast, B.R.: Administrative procedures as instruments of political control. The Journal of Law, Economics, and Organization **3**(2), 243–277 (1987). https://doi.org/10.1093/oxfordjournals.jleo.a036930
19. Meijer, A.: Understanding the complex dynamics of transparency. Public Administration Review **73**(3), 429–439 (2013). https://doi.org/10.1111/puar.12032
20. Nelson, L.K.: Computational grounded theory: A methodological framework. Sociological Methods & Research **49**(1), 3–42 (2020). https://doi.org/10.1177/0049124117729703
21. Reimers, N., Gurevych, I.: Sentence-BERT: Sentence embeddings using siamese BERT-networks. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing. pp. 3982–3992 (2019). https://doi.org/10.18653/v1/D19-1410
22. Salton, G., Wong, A., Yang, C.S.: A vector space model for automatic indexing. Communications of the ACM **18**(11), 613–620 (1975). https://doi.org/10.1145/361219.361220
23. Scott, C.: Accountability in the regulatory state. Journal of Law and Society **27**(1), 38–60 (2000). https://doi.org/10.1111/1467-6478.00146
24. Skaaning, S.E.: Measuring the rule of law. Political Research Quarterly **63**(2), 449–460 (2010). https://doi.org/10.1177/1065912909346745
25. Turney, P.D., Pantel, P.: From frequency to meaning: Vector space models of semantics. Journal of Artificial Intelligence Research **37**, 141–188 (2010). https://doi.org/10.1613/jair.2934