Who Speaks Matters: Authority-Aware Multi-View RAG over Italian Parliamentary Proceedings*

Mirko Tritella[1], Riccardo Pozzi[1], and Matteo Palmonari[1]

University of Milano-Bicocca, Milan, Italy
riccardo.pozzi@unimib.it

Abstract. Parliamentary proceedings are a primary record of democratic deliberation, yet their volume and fragmentation make multi-perspective access difficult for citizens, journalists, and researchers. Applying Retrieval-Augmented Generation (RAG) to parliamentary transcripts introduces three specific risks: dominance of the most frequent speakers, inability to weight speakers according to topical expertise, and citation misattribution in politically sensitive text. We present ParliamentRAG, a RAG system for the Italian Chamber of Deputies that addresses these risks jointly. Its core contribution is a topic-dependent authority model that estimates each speaker's authority as a function of the current query, combining interpretable components such as profession, education, and previous interventions. Given a user query, the system retrieves relevant speech chunks, identifies topic-relevant experts across parliamentary groups, and generates a summary synthesizing their perspectives, accompanied by supporting quotations. ParliamentRAG is evaluated against Google NotebookLM on 15 policy topics via a two-level protocol combining automated metrics and blind A/B human evaluation by six domain experts. The system achieves higher coverage across political groups (0.97 vs. 0.95), perfect quotation faithfulness (1.00 vs. 0.95), and stronger expert preferences on source-related dimensions, while NotebookLM remains stronger on prose-oriented dimensions.

Keywords: Retrieval-Augmented Generation · Parliamentary NLP · Expert Finding · Knowledge Graphs · Quotation Faithfulness · Multi-Perspective Summarization.

1 Introduction

Parliamentary debates constitute a dense documentary record of democratic deliberation: every floor speech, legislative act, and committee intervention expresses the position of an elected representative on matters of public concern. At the time of our experiments (February 2026), the ongoing XIX Legislature of

* Accepted at the In-Use Track of the 25th International Semantic Web Conference (ISWC 2026).

the Italian Chamber of Deputies (October 2022–present) had already produced 608 plenary sessions, 6,010 debates, and 40,416 individual speeches.

Manual navigation of such a corpus is excessively costly, and keyword-based institutional search portals return isolated documents rather than synthesized multi-perspective answers. Furthermore, a representative's position on a given topic typically emerges across multiple sessions and distinct debates rather than within a single intervention, compounding the difficulty of reconstruction.

Journalists, researchers, civil society organizations, and citizens increasingly require tools for exploring parliamentary activity in a structured and transparent way: not isolated documents, but representative, multi-perspective summaries that preserve attribution and traceability to the original speeches. The diverse use cases and target users of such a system are discussed in Section 2.

Large Language Models (LLMs) are well suited for synthesis, but when applied naively to political text they exhibit three risks in civic-information settings: (i) they over-represent the most vocal actors, reflecting the distributional asymmetry of the corpus rather than the actual spread of positions in the debate [39]; (ii) they flatten all speakers to content-based similarity, ignoring the metadata that distinguishes a topical expert from a tangential commenter; and (iii) they hallucinate or misattribute citations at a rate that, although low, is incompatible with journalistic or institutional use [25,29].

In this paper, we present *ParliamentRAG*, a novel RAG application to help both politics professionals (e.g., journalists or analysts) and Italian citizens have a faithful, multi-view, and balanced summary of positions expressed by the ten parliamentary groups of the Italian Chamber of Deputies [11] during parliamentary debates on a specific topic. Knowledge Graphs (KG) enter the design at three levels: all data are stored as a KG because of the relational structure of the domain (e.g., Members of Parliament (MPs) membership of parliamentary groups or MPs signatories for parliamentary acts), graph-based retrieval is a component of the RAG system, and the proposed authority-based ranking depends explicitly on graph information.

Our main objective is to generate synthetic summaries of parliamentary debates while respecting three fundamental principles: (i) **multi-view representation**, ensuring that all parliamentary groups are represented rather than over-relying on frequent speakers; (ii) **authority awareness**, modeling topical authority dynamically so that the most relevant voices emerge per query; and (iii) **quotation traceability**, requiring that every quotation corresponds to a verbatim passage from an official parliamentary transcript.

Figure 1 illustrates an authority-aware, multi-view summary of Italian parliamentary positions on justice reform, including traceable verbatim quotations.

We operationalize these principles into measurable criteria and evaluate ParliamentRAG through a two-level protocol. First, automated metrics assess structural properties of the generated summaries, including coverage and balance of parliamentary groups, citation fidelity, and alignment between cited speakers and the topic-dependent authority model, providing a scalable and reproducible validation of design compliance.

Fig. 1. An example of ParliamentRAG's authority-aware, multi-view summary of the positions of the Italian parliamentary groups on the subject *Justice Reform*. Exact quotations are delimited by « and ». The image presents the first part of a longer answer, shown in the system's English interface; quotations are automatically translated from the original Italian transcripts.

Second, we conduct a blinded expert survey with six domain experts (parliamentary journalists, collaborators, and policy analysts), who evaluate outputs along qualitative dimensions such as clarity, completeness, perceived balance, and overall usefulness, as well as providing pairwise preferences.

We compare ParliamentRAG against NotebookLM, a top-tier system that professionals and citizens may use to explore parliamentary material. ParliamentRAG achieves perfect quotation faithfulness (1.00 vs. 0.95), near-perfect group coverage (0.97 vs. 0.95), and is consistently preferred on source-related dimensions, particularly Source Coverage (5.7:1 preference ratio). While NotebookLM is preferred on prose-quality aspects, overall satisfaction is comparable, indicating that ParliamentRAG matches a strong baseline while excelling on the dimensions aligned with its architectural design.

A live deployment is available at https://www.parliamentrag.it/ and the source code at https://github.com/Emeierkeio/ParliamentRAG under Apache-2.0 license.

2 ParliamentRAG Use Cases

ParliamentRAG is intended for **journalists, parliamentary collaborators, policy analysts, researchers, civil society organizations, and politically engaged citizens** who require transparent and reliable access to parliamentary proceedings. Typical use cases include:

- **Reconstructing political positions** on specific topics by comparing the viewpoints of parliamentary groups and individual representatives.
- **Monitoring legislative activity** and tracking how debates and policy positions evolve over time.
- **Identifying topic experts** by discovering the representatives who most actively and authoritatively contribute to a given subject.
- **Exploring debates transparently**, allowing users to verify generated statements by inspecting the original parliamentary interventions and comparing multiple perspectives.

3 Related Work

We identify five research threads closely related to the contributions of this work: (i) Natural Language Processing (NLP) and (ii) Knowledge Graphs (KGs) applied to the parliamentary domain, (iii) expert and authority modeling in information retrieval, (iv) fairness- and diversity-aware ranking, and (v) faithful attribution and grounded generation.

To the best of our knowledge, no prior work integrates structured parliamentary graphs with query-dependent authority modeling and multi-view generation within a unified RAG architecture.

Parliamentary NLP. Large parliamentary corpora, such as ParlaMint [15] and the Italian IPSA corpus [18], have supported tasks such as stance detection,

topic modeling, and speaker profiling. Ideological-scaling and embedding-based approaches estimate corpus-level political positions [40,37], but do not support per-query retrieval or multi-view synthesis. More recent LLM- and RAG-based systems have improved access to parliamentary material through faceted search for the Maltese Parliament [2] and hybrid vector–KG retrieval for German Bundestag debates [30]. However, these systems do not jointly address the principles central to our work: multi-view representation, authority-aware retrieval, and quotation-level traceability (Section 1).

Parliamentary Knowledge Graphs. Parliamentary Knowledge Graphs and Linked Open Data initiatives have been developed to improve transparency, interoperability, and large-scale analysis of legislative proceedings. Examples include LinkedEP for European Parliament debates [1], ParliamentSampo for Finnish parliamentary records [24], LinkedSaeima for the Latvian parliament [6], and the DemocraSci Knowledge Graph for the Swiss Parliament [38,7]. In Italy, the Chamber of Deputies exposes legislative data through the OCD ontology [9] and the official portal [10]. However, these resources are mainly designed for structured querying, exploration, and data integration, rather than for query-dependent reasoning, multi-view generation, or authority-aware interpretation. Our work addresses this gap by integrating parliamentary data into a RAG architecture for multi-view and authority-aware inference (Section 4).

Expert Finding and Authority. Expert finding formalizes the identification of authoritative individuals from text and metadata [3], distinguishing profile-based from document-based approaches. Link-based ranking [33,26] yields query-independent authority. Source-reliability estimation in RAG [23] assumes a single factually correct answer, an assumption that does not hold in parliamentary debate, where multiple legitimate viewpoints coexist. No prior system computes *interpretable, query-dependent* authority from observable parliamentary features while modeling temporal decay and coalition-crossing invalidation.

Multi-View Retrieval and Fairness. Fairness-aware ranking [47] and diversification methods such as MMR [12] optimize for content or demographic diversity but are not integrated into RAG pipelines for political applications. GraphRAG [14] supports global queries through community summaries, but does not enforce coverage of predefined groups.

Quotation Faithfulness. Attribution frameworks [36], post-hoc systems such as RARR [19], and verification methods [27] improve quotation quality, but the generator remains free to paraphrase the source. Our ParliamentRAG system, instead, provides a *by-construction* guarantee that every quotation is a verbatim substring of a specific source document.

4 Parliamentary Knowledge Graph

ParliamentRAG operates over a knowledge graph (KG) that integrates parliamentary proceedings, metadata, and legislative activity into a unified, Italian-language representation. The underlying data is sourced from the open data infrastructure of the Italian Chamber of Deputies [10], which publishes semanti-

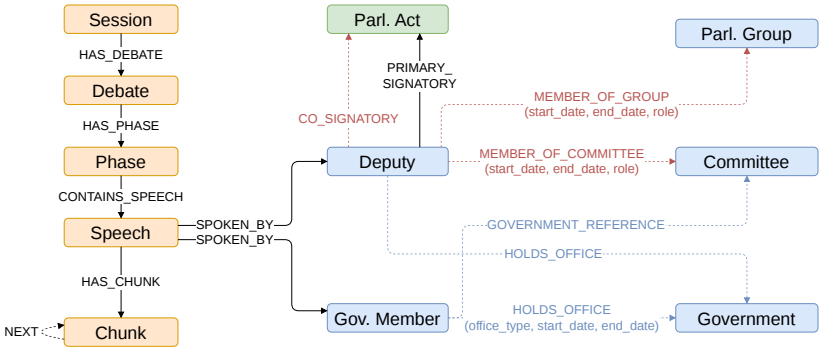

Fig. 2. Knowledge graph schema stored in a single Neo4j instance.

cally structured records in RDF format according to the *Ontology of the Chamber of Deputies* [9]. The dataset includes plenary and committee transcripts, legislative acts with signatories, deputy biographical information, and committee memberships with temporal validity. In this work, we focus on the ongoing legislature.

The RDF data is transformed into a property graph stored in a Neo4j [31] instance, which serves as the sole data layer for the entire system. This architectural choice is motivated by three factors: (i) the need to represent relationships with properties (e.g., temporal validity such as `start_date` and `end_date` on `MEMBER_OF_GROUP` and `MEMBER_OF_COMMITTEE`; see Figure 2), which are naturally handled in a property graph model; (ii) the integration of vector similarity search within the same system, avoiding the overhead of a separate vector store; and (iii) the ability to combine graph traversal, keyword search, and dense retrieval within a unified query infrastructure.

During transformation, all entities and relationships from the source ontology are preserved. At the time of our experiments (February 2026), the graph included 387 deputies (out of 400 elected members; 13 never intervened in plenary debate and thus have no associated speeches), 64 government members (only those who delivered at least one speech during a Chamber session), 10 parliamentary groups, 80 committees, 608 sessions, 6,010 debates, 6,515 phases, 40,416 speeches, and 27,576 legislative acts; the deployed system is continuously updated with new proceedings. The graph is further enriched with information not present in the RDF source: the educational background and profession of MPs, extracted from their short biographical descriptions on the Chamber's website, are embedded and stored as node properties (`education_embedding`, `profession_embedding`). Semi-structured parliamentary data—debate and session transcripts published in XML format [8]—are parsed to extract the hierarchical structure of proceedings and integrated into the graph alongside institutional

data. The resulting graph contained, at that date, 232,755 nodes and 488,487 relationships across 13 node labels and 15 relationship types.

The overall schema is shown in Figure 2. Parliamentary proceedings are organized as a hierarchical structure Session → Debate → Phase → Speech → Chunk, mirroring the institutional workflow. Speeches are segmented into 151,073 chunks, which serve as the atomic retrieval units. Chunking preserves sentence boundaries, each chunk storing its textual content, a 1,536-dimensional dense embedding (text-embedding-3-small), and its position within the speech. Every chunk text is, by an invariant enforced at construction time, an exact substring of its parent speech, allowing retrieved evidence to be deterministically traced back to verbatim source text. Sequential chunks are linked by `NEXT` edges (110,657 links), supporting context-window expansion during generation.

Raw transcripts contain both substantive content and procedural elements (e.g., voting announcements or formal statements), which are removed through lightweight preprocessing. Exact traceability is preserved by construction: since every chunk is an exact substring of its (cleaned) speech text, original passages can be deterministically located in the parliamentary record for quotation grounding.

Beyond the textual hierarchy, the graph captures the political and institutional context of speakers. Deputies are linked to parliamentary groups and committees through time-qualified relationships (505 group memberships, 1,515 committee memberships), allowing affiliation to be resolved at any point in time—a capability essential for coalition-crossing detection in the authority scoring (Section 5). They are also connected to legislative acts via `PRIMARY_SIGNATORY` (27,373) and `CO_SIGNATORY` (103,959) edges, providing a structural signal of topic engagement complementary to speech content. Institutional roles (committee president, vice-president, and secretary, as well as government offices) are recorded as temporally qualified properties of the corresponding membership and office relationships, and serve as signals for authority estimation.

The graph supports three types of access: dense semantic search over chunk embeddings (via Neo4j's native vector index), keyword search over act titles (via a full-text index), and structural graph traversal. This unified representation lets the system combine text and structure, supporting retrieval (e.g., traversing from a parliamentary act to its signatories and their speeches), authority modeling (e.g., aggregating a deputy's legislative activity across signed acts), and quotation grounding (i.e., linking each generated quote back to the exact source passage, guaranteed to be a verbatim substring of the source speech).

Beyond integrating heterogeneous parliamentary data, the knowledge graph directly supports the domain-specific retrieval and query-dependent authority modeling mechanisms described in the next section. These mechanisms exploit structured filtering and graph traversal over the graph representation, complementing dense retrieval over parliamentary debates and keyword-based retrieval over legislative documents.

The deployed graph continues to evolve beyond the snapshot used in the experiments. An incremental pipeline ingests new sessions and acts as they are

published and, since July 2026, resolves NER-extracted references in chunks (person names via the spaCy Italian model; Italian legislative citations via domain-specific patterns) into typed relationships: `MENTIONS` edges from chunks to the deputies or government members named in a speech (≈40,500 at the time of writing) and `CITES` edges from chunks to the Chamber bills cited by number, adding an explicit citation structure to the debates. The graph has likewise been extended with roll-call voting records (≈16.9k roll calls with 6.3M individual deputy votes at the time of writing), exposed in the user interface and included in the RDF export. These extensions postdate the evaluation reported in Section 6 and are not used by the evaluated pipeline; their planned uses (e.g., as an additional graph-retrieval signal) are tracked publicly in the repository issue tracker.

Although stored as a property graph, the KG remains interoperable with the Linked Data ecosystem it derives from: canonical URIs of the source datasets are preserved on all mapped entities, and every node and relationship type has a documented alignment to standard vocabularies (OCD, FOAF, the W3C Organization Ontology, SKOS, and PROV-O, with the proceedings hierarchy mirroring the Akoma Ntoso structure). A round-trip exporter serializes the graph back to RDF; the resulting dumps, together with a presentation of the alignment addressed to non-specialist audiences, are published on the open-data page of the live system (`https://www.parliamentrag.it/data`) and archived on Zenodo under a CC-BY 4.0 license (DOI `10.5281/zenodo.21560332`); project-specific ontology terms and minted entity URIs use the persistent namespace `https://w3id.org/parliamentrag/`.

5 ParliamentRAG

ParliamentRAG is a web application built on top of Next.js [45] and FastAPI [35] that provides an interactive interface for exploring Italian parliamentary proceedings through retrieval-augmented generation. The system is publicly accessible at `https://www.parliamentrag.it`.

The interface supports free-text queries over parliamentary debates, returning structured, multi-perspective answers grounded in official transcripts. Beyond search, it provides auxiliary tools for targeted exploration, including filtered search over legislative acts and speeches, authority analysis of deputies on specific topics, and visualization of parliamentary positions.

We now describe the underlying pipeline that transforms a user query into a multi-view, quotation-grounded response, jointly modeling *what is said* and *who is saying it*. The system follows a retrieval–ranking–generation pipeline in which speaker authority, coverage of political groups, and quotation faithfulness are enforced as first-class constraints.

Given a query q, the system (i) retrieves candidate evidence from parliamentary proceedings via a dual-channel strategy, (ii) reranks evidence using a composite score that incorporates query-dependent speaker authority and group coverage, (iii) selects representative experts for each parliamentary group, and

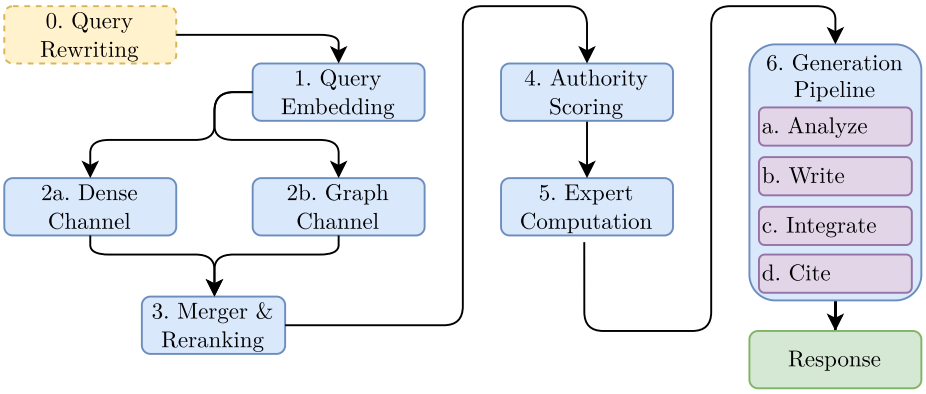

Fig. 3. Processing Pipeline. Steps 2a and 2b are executed in parallel.

(iv) generates a structured multi-view answer with verbatim quotations extracted from source documents. A schema of the pipeline is depicted in Figure 3.

5.1 Dual-channel Retrieval

The retrieval stage must identify speech chunks that are (i) semantically relevant with respect to the user query, (ii) representative of all parliamentary groups, and (iii) reflective of speakers whose expertise is evidenced by legislative activity rather than solely by speech content.

To this end, we adopt a dual-channel approach. A *dense channel* performs vector similarity search over speech chunks, capturing thematically relevant content independently of the speaker. Later, graph traversal is used to resolve the associated speech, speaker, and parliamentary group, also evaluating if the speaker's current group differs from that at the time of the speech. The primary limitation of the dense channel is that it cannot distinguish between a deputy who has spoken tangentially about a topic and one who has actively sponsored legislation in the same domain. The *graph channel* retrieval is based on parliamentary acts. First, acts relevant to the query are selected with hybrid lexical and dense, semantic similarity, then the primary signatories and co-signatories are obtained with graph traversal, as well as their speeches and chunks.

Additionally, when the query can be mapped to a parliamentary committee via a curated keyword-based mapping [41], the retrieval engine additionally privileges chunks from speakers who are members of the relevant committee (e.g., justice or foreign affairs; this information is found in the KG), since committee membership is a strong institutional proxy for domain expertise.

The combination of the two channels yields a more diverse result set, spanning a broader range of political groups and evidence types, than either channel alone.

5.2 Authority-Aware Reranking and Expert Computation

Retrieved evidence is merged by deduplicating items by chunk identifier and retaining the higher score. Each evidence item e is ranked using a weighted score:

$$\text{score}(e) = w_r r(e) + w_d d(e) + w_v v(e) + w_a \, authority(s_e, q) + w_\sigma \sigma(e), \quad (1)$$

where s_e denotes the speaker of e. The weights are set as follows: relevance $w_r = 0.35$, diversity $w_d = 0.15$, coverage $w_v = 0.20$, authority $w_a = 0.05$, and salience $w_\sigma = 0.25$. Relevance is prioritized as the primary retrieval objective, while salience promotes meaningful political content over procedural language. Coverage and diversity ensure balanced representation across parliamentary groups and speakers. Authority models speaker expertise with respect to the query; its weight is the smallest so that it informs ranking without overriding topical relevance, which would risk marginalizing speakers from small parliamentary groups. Authority is computed after retrieval and is also used for expert selection (one top-authority speaker per group), which influences the generation pipeline but not the evidence pool itself.

Authority Score. We define a query-dependent authority score for each speaker by aggregating multiple signals capturing biographical information, institutional roles, and parliamentary activity. The score combines cosine similarity between query embeddings and the embeddings of speaker attributes (profession, education, committees, and roles) with activity-based signals derived from legislative acts and speech interventions, both subject to temporal decay.

Formally, the raw authority score is computed as a weighted sum of heterogeneous components:

$$authority(s, q) = \sum_i w_i \, c_i(s, q), \quad (2)$$

where components include semantic similarity terms, time-decayed counts of parliamentary activity, and role-based priors.

The assigned weights are $w_{\text{profession}} = 0.15$, $w_{\text{education}} = 0.10$, $w_{\text{committee}} = 0.25$, $w_{\text{legislative acts}} = 0.20$, $w_{\text{speech interventions}} = 0.25$, and $w_{\text{institutional role}} = 0.05$. These weights are set empirically based on expert judgment rather than learned from data, keeping the scoring function interpretable and adaptable; the formulation supports future extensions such as personalization, context-specific reweighting, or interactive weight learning [5]. Moreover, all weights—both the reranking weights above and the authority components—are publicly documented in the open-source configuration and directly adjustable by users through the application settings, so the default configuration is a transparent and auditable choice rather than a fixed editorial stance.

Authority signals are time-decayed so that recent parliamentary activity contributes more than older evidence, applying the same decay rate to both legislative acts and speech interventions to prioritize current relevance and avoid overemphasizing past prominence.

Coalition affiliation is time-dependent and resolved at query time, ensuring that only activities within a speaker's current political group are considered, preventing historical coalition shifts from biasing present-day authority estimates.

Expert Computation. At this point, for each parliamentary group $g \in G$, the system selects the speaker with the highest query-dependent authority score among those from g present in the evidence pool. This speaker serves as the group's expert representative: their interventions and cited passages become the primary voice through which that group's position is conveyed in the summary.

5.3 Multi-View Generation Pipeline

As visible in Figure 3, the generation process follows a four-stage pipeline—*Analyze, Write, Integrate, Cite*—that enforces coverage, grounding, and coherence by design. In the first stage (*Analyze*), the input query is split into a set of simple, atomic claims, each linked to the evidence required for every parliamentary group. This ensures that all relevant aspects of the query are explicitly addressed in the following steps.

In the second stage (*Write*), the system produces one section per parliamentary group from a structured *position brief* summarizing the top evidence chunks for that group, ordered by authority score so that the most authoritative speakers are cited first. All groups are treated uniformly, and if no supporting evidence is available, this is stated explicitly instead of generating unsupported content. The third stage (*Integrate*) combines these sections into a single coherent narrative, while preserving alignment with the original content and ensuring that all quotation placeholders are retained.

In the final stage (*Cite*), quotations are resolved deterministically into verbatim passages from the parliamentary record. Every quotation is verified to be an exact substring of the source speech before publication: candidate quotations that are not verbatim are rejected or deterministically remapped onto the original sentences, so the language model can never introduce quotation text of its own. As a result, every quote is directly grounded in source text, preventing hallucinations by construction and ensuring that all claims are supported by verifiable evidence.

6 Experimental Evaluation

The experimental evaluation focuses on the downstream quality of generated responses and assesses whether ParliamentRAG effectively supports the exploration of parliamentary debates and policy positions while satisfying the core design principles introduced in Section 1 (balanced multi-view coverage across parliamentary groups, authority-aware evidence selection, grounded and traceable quotations). More specifically, we address the following research question: *(i) what is the quality of the analyses produced by ParliamentRAG, a RAG architecture explicitly designed around the principles introduced above, in light of these design principles; and (ii) how does this quality compare with that of analyses*

produced by a top-tier general-purpose document-grounded RAG system explicitly instructed to follow the same principles?

For our comparison, we selected Google NotebookLM [20], for the following reasons: we are not aware of other competing systems evaluated and published in peer-reviewed publications; NotebookLM represents a realistic and accessible alternative that end users could readily employ to analyze parliamentary material without developing a custom pipeline, after they autonomously collect the data; it is a highly engineered system backed by the top-tier Gemini 3.0 LLM [22] (see Section 6.1 for the experimental settings of NotebookLM).

To answer our research questions, we combine a response-level automatic analysis with a blind A/B assessment by domain experts on a benchmark of questions: automatic metrics measure the satisfaction of our design principles (coverage balance, quotation faithfulness), while experts' judgments capture perceived quality and overall satisfaction.

6.1 Benchmark and Protocol

Selection of benchmark topic-based questions. A first step consists in identifying topics of interest to guide the actual test. We collected 51 policy topics through a survey involving 17 Italian citizens interested in politics, and reduced them to 15 by selecting topics spanning different levels of perceived polarization and public discussion intensity, as assessed directly in the survey. Topics span eight thematic macro-areas: Justice & Civil Rights, Institutional Reforms, Immigration, Labor & Welfare, Environment & Energy, Technology, Defense & Foreign Policy, Economy & Finance, with varying debate intensity and polarization (Likert 1–5). For each topic we create a query with the following template: "Qual è la posizione dei gruppi parlamentari su {topic}?" (Italian for "What is the position of the parliamentary groups on {topic}?").

NotebookLM experimental setup. ParliamentRAG perforemed retrieval on all 151k chunks extracted from the parliamentary proceedings (February 2026 snapshot). Feeding all proceedings data to NotebookLM, a general-purpose system, is unfeasible and unfair because the system is not engineered to work with this kind and size of data. We therefore create a notebook for each topic by engineering its context to simulate the case where an expert selects proceedings deemed of interest for a topic; observe that this selection step is challenging, as a topic can be addressed across different proceedings, while ParliamentRAG retrieves data from all proceedings. We select $\approx$ 150 relevant + 150 distractor chunks per topic-specific notebook. Chunks are enriched with metadata to provide additional context (speaker, parliamentary group, date, session, debate). The relevant ones are obtained starting from the query with the ParliamentRAG retrieval pipeline, including the reranking, which considers the authority score, while the distractors are obtained from similar topics (same macro-area but different topic). Distractors provide a fairer simulation of retrieval-based context and the presence of non-relevant information in uploaded documents.

Comparison of ParliamentRAG and NotebookLM. A key distinction between the two systems is how the desired design principles are enforced. Par-

liamentRAG implements them directly in its architecture through explicit mechanisms, such as authority-aware ranking and multi-view retrieval across parliamentary groups. Conversely, NotebookLM relies on prompt instructions to encourage similar behaviors, without architectural guarantees; all prompts are available in the code base [42].

The comparison therefore investigates whether structurally enforced constraints can produce more reliable and informative outputs than prompt-level guidance alone. Another difference concerns the underlying LLMs: at the time of the experiments, ParliamentRAG used GPT-4o, while NotebookLM used Gemini 3 (the deployed ParliamentRAG has since been upgraded to the GPT-4.1 model family). Gemini 3 models are more recent (November 2025 [21]) and achieve higher rankings than GPT-4o, released in May 2024 [32], in public benchmarks [28,13].

In summary, both systems are instructed to produce per-group structured responses via equivalent prompts, while NotebookLM is given an engineered context and backed by a stronger LLM. This setup provides a conservative evaluation of ParliamentRAG, focusing on response quality under curated retrieval conditions rather than a fully independent end-to-end retrieval comparison.

Automated Evaluation Protocol. To measure whether the generated answers satisfy the core principles of ParliamentRAG, we first calculate automatic metrics. Coverage is evaluated through two complementary measures. *Groups with Quotation (GQ)* quantifies the fraction of the ten parliamentary groups for which at least one speaker has been quoted in the response, while *Completeness* measures the fraction of groups receiving a dedicated narrative section in the generated analysis. To evaluate quotation reliability, we compute *Quotation Faithfulness*, defined as the fraction of quotations that exactly match substrings of the original parliamentary interventions. Finally, to analyze the behavior of the authority-aware retrieval mechanism, we report the mean authority score (MA) and the corresponding standard deviation (ASD) of the cited speakers under the query-dependent authority model.

Human Evaluation Protocol. Six domain experts, disjoint from the paper authors, were selected among parliamentary journalists, parliamentary collaborators, and policy analysts, with a mean experience of 7.2 years. They independently rated both systems on nine Likert (1–5) dimensions: Answer Quality, Answer Clarity, Answer Completeness, Quotation Relevance, Balance Perception, Balance Fairness, Source Relevance, Source Authority, Source Coverage. Evaluators also provided an overall A/B preference and a 1–5 satisfaction score. The protocol follows the PARADISE framework [46] and the A/B preference format of Chatbot Arena [13]. The evaluation was blind: system identities were hidden and left/right assignment randomized per topic. Two evaluators completed all 15 topics, while other evaluators partially completed the evaluation, yielding a total of $N = 67$ paired evaluations.

Beyond the quantitative ratings, evaluators also assessed individual quotations (e.g., relevance, faithfulness, informativeness, attribution, and potential issues), identified the most authoritative sources among alternatives, and pro-

vided qualitative feedback. Completing the evaluation required approximately two hours per participant. Although ParliamentRAG targets both citizens and professionals, we prioritized domain experts—better positioned to assess political balance, quotation quality, and source authority—favoring an in-depth evaluation on 15 representative topics over a broader but shallower assessment.

We assess differences between ParliamentRAG and NotebookLM using non-parametric paired tests (Wilcoxon signed-rank) with Holm–Bonferroni correction for multiple comparisons. In addition, we calculate Cohen's d as an effect size measure to quantify the magnitude of observed differences.

6.2 Results

Table 1. Automated Evaluation: ParliamentRAG vs. NotebookLM. Mean values (μ) over the 15 topics and difference (Δ). Mean Authority (MA) for NotebookLM is computed from the query-related authority of the MPs mentioned by NotebookLM.

Metric	ParliamentRAG (μ)	NotebookLM (μ)	Δ
Groups with Quotation (GQ)	0.97	0.95	+0.02
Completeness	0.99	1.00	−0.01
Quotation Faithfulness (QF)	1.00	0.95	+0.05
Mean Authority (MA)	0.53	0.52	+0.01

Table 1 reports the automated metrics. ParliamentRAG achieves near-perfect quotation-level group coverage (GQ = 0.97 vs. 0.95 for NotebookLM), which is architecturally guaranteed at generation-time by stratified per-group generation. However, coverage remains bounded by the retrieval stage: if no retrieved chunks are associated with a given political group, that group cannot be represented during generation, explaining why GQ $\neq$ 1.0. On Quotation Faithfulness the gap is structurally significant: ParliamentRAG reaches QF = 1.00 by design, whereas NotebookLM reaches 0.95, meaning approximately 5% of its quotations do not correspond verbatim to the source—an error rate that, in a journalistic context, is corrosive to credibility. Mean authority of cited speakers is modestly higher for ParliamentRAG (0.53 vs. 0.52); since NotebookLM's context was itself selected with the authority-aware ParliamentRAG retrieval pipeline (Section 6.1), a small gap is expected.

Table 2 reports the results of the human evaluation. Overall satisfaction is nearly identical across the two systems (4.24 vs. 4.27), suggesting comparable perceived utility despite substantial architectural differences. However, the dimension-level analysis reveals complementary strengths.

NotebookLM achieves higher ratings on prose-oriented dimensions, namely Answer Quality (4.30 vs. 4.04) and Answer Clarity (4.51 vs. 4.27), and is more frequently preferred on Answer Quality (45% vs. 30%); we return to this asymmetry in Section 7.

Table 2. Human evaluation comparing ParliamentRAG (PRAG) and NotebookLM (NbLM). The left section reports mean Likert ratings (1–5) for each evaluation dimension, while the right section reports the percentage of pairwise preferences in favor of PRAG, ties, and preferences in favor of NbLM.

Dimension	PRAG (μ)	NbLM (μ)	PRAG Pref.	Ties	NbLM Pref.
Answer Quality	4.04	4.30	30%	25%	45%
Answer Clarity	4.27	4.51	22%	48%	30%
Answer Complet.	4.12	4.09	24%	55%	21%
Quotation Relevance	4.37	4.36	24%	54%	22%
Balance Perception	4.66	4.48	21%	72%	7%
Balance Fairness	4.67	4.66	12%	78%	10%
Source Relevance	4.07	3.84	33%	48%	19%
Source Authority	4.21	4.00	30%	57%	13%
Source Coverage	4.64	4.39	25%	70%	5%
Overall Satisfaction	4.24	4.27	31%	42%	27%

By contrast, ParliamentRAG consistently outperforms NotebookLM on dimensions directly related to the objectives of the proposed architecture. In particular, evaluators report higher ratings for Source Relevance (4.07 vs. 3.84), Source Authority (4.21 vs. 4.00), and Source Coverage (4.64 vs. 4.39). Pairwise preferences reinforce this pattern: ParliamentRAG is preferred more frequently on Source Relevance (33% vs. 19%), Source Authority (30% vs. 13%), and especially Source Coverage (25% vs. 5%). Similarly, ParliamentRAG obtains a clear advantage on Balance Perception (4.66 vs. 4.48), although the large proportion of ties (72%) indicates that both systems are often perceived as balanced.

Although none of the observed differences reaches statistical significance after Holm–Bonferroni correction for multiple comparisons, the results exhibit a coherent qualitative pattern. Small but consistent Cohen's d effect sizes favor ParliamentRAG on source- and balance-related dimensions, including Source Relevance ($d = 0.35$), Source Coverage ($d = 0.35$), and Source Authority ($d = 0.28$), while NotebookLM shows advantages on fluency-oriented dimensions such as Answer Quality ($d = -0.31$) and Answer Clarity ($d = -0.29$).

Finally, after completing the per-topic evaluation, respondents were asked whether they would recommend a similar system to their colleagues, with "no" set as the predefined answer. They answered affirmatively in 72% of cases, suggesting that such a system was generally perceived as useful.

7 Discussion

The downstream evaluation reveals complementary strengths between the two systems. ParliamentRAG consistently performs better on source-oriented and balance-related dimensions, including source authority, source coverage, source relevance, and perceived political balance, while NotebookLM achieves higher ratings on fluency-oriented aspects such as answer quality and clarity. This

asymmetry can be explained by ParliamentRAG prioritizing structural guarantees through explicit multi-view retrieval and authority-aware evidence selection. Better fluency and clarity in NotebookLM can be explained either as the outcome of a monolithic generation pipeline or as the result of using a better backbone model. However, the observed complementarity is not symmetric. Fluency limitations can often be mitigated through stronger language models or lightweight rewriting stages, whereas guarantees such as verbatim quotation faithfulness and systematic per-group coverage require architectural support and cannot be reliably enforced through prompting alone. It is also important to recall that NotebookLM was evaluated with a curated, pre-retrieved context, rather than over the full parliamentary corpus as ParliamentRAG was. Thus, the comparison does not test NotebookLM's ability to retrieve evidence from the complete proceedings; in practice, this task would remain substantially more challenging for end users without a dedicated system such as ParliamentRAG.

Among the evaluated properties, quotation faithfulness is the strongest invariant enforced by the architecture. Fabricated or altered quotations are structurally prevented, since every published quotation is verified to be a verbatim substring of the source transcript. Other properties, such as balanced group coverage, should be interpreted as strong design objectives rather than absolute guarantees, as they depend on the availability of relevant retrieved evidence.

Limitations. The evaluation is limited by the small benchmark size (15 topics) and moderate statistical power. In addition, the study compares the complete system against NotebookLM, which is backed by a more powerful LLM, without internal ablation experiments, leaving the contribution of architectural components to future work.

8 Conclusion

We have presented ParliamentRAG, an authority-aware, multi-view RAG system for Italian parliamentary proceedings. On a 15-topic benchmark with expert A/B evaluation, the system matches a strong commercial baseline (NotebookLM) on overall satisfaction while excelling precisely on the dimensions it was designed to target (group coverage, quotation faithfulness, source authority). Directions for future work include personalizing ranking based on users' preferences, developing domain-specific embedding models, extending the system to the Senate and earlier legislatures (data are available from the 2006–2008 legislature onward), and enabling multilingual access to parliamentary debates. The approach also generalizes to foreign parliaments—e.g., the UK Parliament and the European Parliament publish structured data and transcribed debates [44,43,16,17]—enabling cross-country comparative analyses. Another promising direction is efficient local deployment based on small open-weight LLMs, building on scalable knowledge graph traversal [34] and efficient verbatim quotation retrieval [4].

Acknowledgments. We thank the six domain experts who contributed to the evaluation and the 17 respondents to the participatory topic-collection survey. This work was supported by the EU Horizon Europe programme (grant No. 101189771—DataPACT).

Supplemental Material Statement: Source code is released at https://github.com/Emeierkeio/ParliamentRAG under Apache-2.0 license; the repository also contains all generation prompts and the retrieval, reranking, and authority-weight configurations referenced in this paper; the exact code and configuration used to generate the evaluated responses are preserved at the git tag `iswc2026-eval`. The system is publicly deployed at https://www.parliamentrag.it and its public API is also exposed as a Model Context Protocol server (https://mcp.parliamentrag.it). The knowledge graph is available as an RDF dump (Turtle, with individual votes in a separate N-Triples file), archived on Zenodo under a CC-BY 4.0 license (DOI 10.5281/zenodo.21560332; 39 views and 12 downloads at the time of writing) and linked from the open-data page of the live system (https://www.parliamentrag.it/data), which also documents, for non-specialist audiences, the alignment of the graph to standard vocabularies (OCD, FOAF, the W3C Organization Ontology, SKOS, and PROV-O); project-specific terms resolve through the persistent namespace https://w3id.org/parliamentrag/. A single build command regenerates the dump from the updated graph, and the graph itself can be reconstructed end-to-end from the public open data of the Chamber of Deputies (released under CC-BY) using the in-repository pipeline (see Section 4). A semantic description of this work is available in the Open Research Knowledge Graph at https://orkg.org/papers/R1909763; the dataset statistics recorded in the entry are synchronized programmatically (via the ORKG REST API) with the live knowledge graph at every dataset update, so the ORKG description remains aligned with the deployed system rather than reflecting a static snapshot at publication time. Ongoing maintenance and planned extensions of the system and of the knowledge graph are tracked as public issues in the source repository, documenting the evolution of the resource beyond this publication.

Use of Generative AI

Generative AI tools were used during the preparation of this work to assist with language refinement. In particular, these tools were employed to paraphrase selected passages and improve the clarity, fluency, and overall quality of the English text.

References

1. van Aggelen, A., Hollink, L., Kemman, M., Kleppe, M., Beunders, H.: The debates of the european parliament as linked open data. Semant. Web **8**(2), 271281 (Jan 2017). https://doi.org/10.3233/SW-160227, https://doi.org/10.3233/SW-160227
2. Azzopardi, J.: Llms and knowledge discovery in low-resource language parliamentary corpora: The pq dashboard case study. In: Proceedings of the 17th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2025) - Volume 1: KDIR. pp. 159–170. SciTePress

(2025). https://doi.org/10.5220/0013835100004000, https://www.scitepress.org/Papers/2025/138351/138351.pdf
3. Balog, K., Fang, Y., Rijke, M., Serdyukov, P., Si, L.: Expertise retrieval. Foundations and Trends in Information Retrieval **6**, 127–256 (01 2012). https://doi.org/10.1561/1500000024
4. Bevilacqua, M., Ottaviano, G., Lewis, P., Yih, S., Riedel, S., Petroni, F.: Autoregressive search engines: Generating substrings as document identifiers. In: Advances in Neural Information Processing Systems (2022), https://openreview.net/forum?id=Z4kZxAjg8Y
5. Bianchi, F., Palmonari, M., Cremaschi, M., Fersini, E.: Actively learning to rank semantic associations for personalized contextual exploration of knowledge graphs. In: The Semantic Web. pp. 120–135. Springer International Publishing, Cham (2017). https://doi.org/10.1007/978-3-319-58068-5_8
6. Bojārs, U., Darģis, R., Lavrinovičs, U., Paikens, P.: Linkedsaeima: A linked open dataset of latvia's parliamentary debates. In: Acosta, M., Cudré-Mauroux, P., Maleshkova, M., Pellegrini, T., Sack, H., Sure-Vetter, Y. (eds.) Semantic Systems. The Power of AI and Knowledge Graphs. pp. 50–56. Springer International Publishing, Cham (2019). https://doi.org/10.1007/978-3-030-33220-4_4
7. Brandenberger, L., Minder, J., Salamanca, L., Schlosser, S., Gasser, L., Jung, V., Shariat, K., Balode, M., Schmidt-Rohr, A., Babić, L., Perez-Cruz, F., Schweitzer, F.: Democrasci - a parliamentary knowledge graph (4 legislative periods) (0.9.0) [dataset] (2024). https://doi.org/10.5281/zenodo.13920293, https://doi.org/10.5281/zenodo.13920293
8. Camera dei deputati: Debate and session transcripts published in XML. https://documenti.camera.it/apps/commonServices/getDocumento.ashx?sezione=assemblea&tipoDoc=formato_xml&tipologia=stenografico&idNumero=0655&idLegislatura=19, accessed: 2026
9. Camera dei deputati: Ontology of the chamber of deputies: RDF classes. https://dati.camera.it/ocd/classi.rdf, accessed: 2026
10. Camera dei deputati: Portale dei dati aperti della camera dei deputati. https://dati.camera.it, accessed: 2026
11. Camera dei deputati: Xix legislatura – camera dei deputati. https://www.camera.it/leg19/46, accessed: 2026
12. Carbonell, J., Goldstein, J.: The use of mmr, diversity-based reranking for reordering documents and producing summaries. In: Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. p. 335336. SIGIR '98, Association for Computing Machinery, New York, NY, USA (1998). https://doi.org/10.1145/290941.291025, https://doi.org/10.1145/290941.291025
13. Chiang, W.L., Zheng, L., Sheng, Y., Angelopoulos, A.N., Li, T., Li, D., Zhu, B., Zhang, H., Jordan, M.I., Gonzalez, J.E., Stoica, I.: Chatbot arena: an open platform for evaluating llms by human preference. In: Proceedings of the 41st International Conference on Machine Learning. ICML'24, JMLR.org (2024), https://openreview.net/pdf?id=3MW8GKNyzI
14. Edge, D., Trinh, H., Cheng, N., Bradley, J., Chao, A., Mody, A., Truitt, S., Metropolitansky, D., Ness, R.O., Larson, J.: From local to global: A graph rag approach to query-focused summarization (2025), https://arxiv.org/abs/2404.16130
15. Erjavec, T., Kopp, M., Ljubešić, N., Kuzman, T., Rayson, P., Osenova, P., Ogrodniczuk, M., Çöltekin, Ç., Koržinek, D., Meden, K., Skubic, J., Rupnik, P., Agnoloni,

T., Aires, J., Barkarson, S., Bartolini, R., Bel, N., Calzada Pérez, M., Darġis, R., Diwersy, S., Gavriilidou, M., van Heusden, R., Iruskieta, M., Kahusk, N., Kryvenko, A., Ligeti-Nagy, N., Magariños, C., Mölder, M., Navarretta, C., Simov, K., Tungland, L.M., Tuominen, J., Vidler, J., Vladu, A.I., Wissik, T., Yrjänäinen, V., Fišer, D.: Parlamint ii: Advancing comparable parliamentary corpora across europe. Language Resources and Evaluation **59**(3), 2071–2102 (2025). https://doi.org/10.1007/s10579-024-09798-w, https://doi.org/10.1007/s10579-024-09798-w

16. European Parliament: European parliament open data portal. https://data.europarl.europa.eu/en/datasets, accessed: 2026-07-18
17. European Parliament: European parliament plenary debates. https://www.europarl.europa.eu/plenary/en/debates-video.html, accessed: 2026-07-18
18. Frasnelli, V., Palmero Aprosio, A.: There's something new about the Italian parliament: The IPSA corpus. In: Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024). pp. 16037–16046. ELRA and ICCL, Torino, Italia (May 2024), https://aclanthology.org/2024.lrec-main.1394/
19. Gao, L., Dai, Z., Pasupat, P., Chen, A., Chaganty, A.T., Fan, Y., Zhao, V., Lao, N., Lee, H., Juan, D.C., Guu, K.: RARR: Researching and revising what language models say, using language models. In: Rogers, A., Boyd-Graber, J., Okazaki, N. (eds.) Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). pp. 16477–16508. Association for Computational Linguistics, Toronto, Canada (Jul 2023). https://doi.org/10.18653/v1/2023.acl-long.910, https://aclanthology.org/2023.acl-long.910/
20. Google: Google notebooklm. https://notebooklm.google.com/, accessed: 2026
21. Google: Gemini 3: Our most capable model yet. https://blog.google/products-and-platforms/products/gemini/gemini-3/ (2025), accessed: 2026
22. Google NotebookLM: Introducing gemini 3. https://x.com/NotebookLM/status/2002115447425282449 (2025), accessed: 2026
23. Hwang, J., Park, J., Park, H., Kim, D., Park, S., Ok, J.: Retrieval-augmented generation with estimation of source reliability. In: Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing. pp. 34279–34303. Association for Computational Linguistics, Suzhou, China (Nov 2025). https://doi.org/10.18653/v1/2025.emnlp-main.1738, https://aclanthology.org/2025.emnlp-main.1738/
24. Hyvönen, E., Sinikallio, L., Leskinen, P., Drobac, S., Leal, R., La Mela, M., Tuominen, J., Poikkimäki, H., Rantala, H.: Publishing and using parliamentary linked data on the semantic web: Parliamentsampo system for parliament of finland. Semantic Web **16**(1), SW–243683 (2025). https://doi.org/10.3233/SW-243683
25. Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y.J., Madotto, A., Fung, P.: Survey of hallucination in natural language generation. ACM Computing Surveys **55**(12), 138 (Mar 2023). https://doi.org/10.1145/3571730, http://dx.doi.org/10.1145/3571730
26. Kleinberg, J.M.: Authoritative sources in a hyperlinked environment. J. ACM **46**(5), 604632 (Sep 1999). https://doi.org/10.1145/324133.324140, https://doi.org/10.1145/324133.324140
27. Liu, N., Zhang, T., Liang, P.: Evaluating verifiability in generative search engines. In: Findings of the Association for Computational Linguistics: EMNLP 2023. pp. 7001–7025. Association for Computational Linguistics, Singapore (Dec 2023). https://doi.org/10.18653/v1/2023.findings-emnlp.467, https://aclanthology.org/2023.findings-emnlp.467/

28. LMArena: Chatbot arena – LLM leaderboard. https://arena.ai/leaderboard/text (2026), accessed: 2026
29. Maynez, J., Narayan, S., Bohnet, B., McDonald, R.: On faithfulness and factuality in abstractive summarization. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. pp. 1906–1919. Association for Computational Linguistics, Online (Jul 2020). https://doi.org/10.18653/v1/2020.acl-main.173, https://aclanthology.org/2020.acl-main.173/
30. Mosbach, S., Lai, J., Rustagi, K., Tran, D.N., Kraft, M., Bindereif, E., Azzam, M.: Parliamentary debates in the world avatar: A hybrid retrieval-augmented generation system (2025), https://como.ceb.cam.ac.uk/media/preprints/c4e-preprint-338.pdf, preprint
31. Neo4j, Inc.: Neo4j graph database. https://neo4j.com/, accessed: 2026
32. OpenAI: Hello GPT-4o. https://openai.com/index/hello-gpt-4o/ (2024), accessed: 2026
33. Page, L., Brin, S., Motwani, R., Winograd, T.: The PageRank Citation Ranking: Bringing Order to the Web. Tech. rep., Stanford Digital Library Technologies Project (1998), http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.31.1768
34. Pozzi, R., Palmonari, M., Coletta, A., Bellomarini, L., Lehmann, J., Vahdati, S.: Refactx: Scalable reasoning with reliable facts via constrained generation. In: The Semantic Web – ISWC 2025. pp. 290–308. Springer Nature Switzerland, Cham (2026). https://doi.org/10.1007/978-3-032-09527-5_16, https://doi.org/10.1007/978-3-032-09527-5_16
35. Ramírez, S.: Fastapi. https://github.com/fastapi/fastapi, accessed: 2026
36. Rashkin, H., Nikolaev, V., Lamm, M., Aroyo, L., Collins, M., Das, D., Petrov, S., Tomar, G.S., Turc, I., Reitter, D.: Measuring attribution in natural language generation models. Computational Linguistics **49**(4), 777–840 (Dec 2023). https://doi.org/10.1162/coli_a_00486, https://aclanthology.org/2023.cl-4.2/
37. Rheault, L., Cochrane, C.: Word embeddings for the analysis of ideological placement in parliamentary corpora. Political Analysis **28**(1), 112133 (2020). https://doi.org/10.1017/pan.2019.26
38. Salamanca, L., Brandenberger, L., Gasser, L., Schlosser, S., Balode, M., Jung, V., Perez-Cruz, F., Schweitzer, F.: Processing large-scale archival records: The case of the swiss parliamentary records. Swiss Political Science Review **30**(2), 140–153 (2024). https://doi.org/https://doi.org/10.1111/spsr.12590, https://onlinelibrary.wiley.com/doi/abs/10.1111/spsr.12590
39. Santurkar, S., Durmus, E., Ladhak, F., Lee, C., Liang, P., Hashimoto, T.: Whose opinions do language models reflect? In: Proceedings of the 40th International Conference on Machine Learning. ICML'23, JMLR.org (2023), https://openreview.net/pdf?id=7IRybndMLU
40. Slapin, J.B., Proksch, S.O.: A scaling model for estimating time-series party positions from texts. American Journal of Political Science **52**(3), 705–722 (2008). https://doi.org/https://doi.org/10.1111/j.1540-5907.2008.00338.x, https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-5907.2008.00338.x
41. Tritella, M.: Parliamentrag – committee topics configuration. https://github.com/Emeierkeio/ParliamentRAG/blob/main/backend/config/commissioni_topics.yaml (2026), accessed: 2026
42. Tritella, M.: Parliamentrag – evaluation prompts. https://github.com/Emeierkeio/ParliamentRAG/blob/main/docs/Prompts.md (2026), accessed: 2026
43. UK Parliament: Hansard: Parliamentary debates. https://hansard.parliament.uk/, accessed: 2026-07-18

44. UK Parliament: Uk parliament data. https://explore.data.parliament.uk/, accessed: 2026-07-18
45. Vercel, Inc.: Next.js – the react framework for the web. https://nextjs.org/, accessed: 2026
46. Walker, M.A., Litman, D.J., Kamm, C.A., Abella, A.: PARADISE: A framework for evaluating spoken dialogue agents. In: 35th Annual Meeting of the Association for Computational Linguistics and 8th Conference of the European Chapter of the Association for Computational Linguistics. pp. 271–280. Association for Computational Linguistics, Madrid, Spain (Jul 1997). https://doi.org/10.3115/976909.979652, https://aclanthology.org/P97-1035/
47. Zehlike, M., Bonchi, F., Castillo, C., Hajian, S., Megahed, M., Baeza-Yates, R.: Fa*ir: A fair top-k ranking algorithm. In: Proceedings of the 2017 ACM on Conference on Information and Knowledge Management. p. 15691578. CIKM 17, ACM (Nov 2017). https://doi.org/10.1145/3132847.3132938, http://dx.doi.org/10.1145/3132847.3132938