IA4FriLex: Enhancing The Legislative Consultation Process With AI

Talia B. Kimber [1] **Ornella Vaccarelli** [2] **Thomas Christin** [1] **Thomas Kadelbach** [3] **Beat Wolf** [2]

Abstract

Legislative consultation procedures are a core component of participatory law-making but require public administrations to process large volumes of heterogeneous and predominantly unstructured submissions under strict procedural constraints. This task is particularly demanding in practice and often constitutes a bottleneck in legislative workflows.

To overcome this challenge, we present IA4FriLex, an AI-assisted pipeline designed to support the processing and synthesis of consultation submissions through a structured and legally grounded workflow. Built exclusively on open-source software and Large Language Models (LLMs), the system automates well-defined stages of consultation handling while preserving human oversight and legal responsibility. IA4FriLex produces standardized, department-ready consultation reports aligned with established cantonal administrative practices.

The system has been implemented and deployed in collaboration with the Cantonal administration of Fribourg and evaluated on four real cantonal legislative cases, covering both completed and ongoing consultations. Results show that IA4FriLex reliably generates high-quality first-pass syntheses and reduces consultation report preparation time by at least 80% compared to fully manual drafting.

These findings demonstrate that carefully constrained, on-premise deployments of LLM-based systems can effectively support legislative consultation processes, offering a scalable and institutionally compatible approach to AI-assisted law-making.

1. Introduction

Legislative consultation procedures constitute a recognized instrument of participatory law-making in democratic legal systems, allowing public authorities to gather and assess reasoned opinions on draft normative acts prior to their adoption. In Switzerland, in particular, consultation procedures are a structurally embedded and systematic element of the legislative process, reflecting long-standing traditions of consensus-oriented governance and direct democracy. This practice applies not only at the federal level but also at the cantonal level, where draft laws are routinely subjected to consultation among institutional actors, municipalities, associations, and other interested bodies, in accordance with cantonal legislative procedures (swi, 1999; REA).

While the value of comprehensive consultations is undisputed, the practical execution presents significant operational challenges. The input received, often vast and heterogeneous, requires meticulous manual collection, legal interpretation, and synthesis. This process is inherently time-consuming and resource-intensive, frequently creating a bottleneck in the legislative timeline and placing substantial demands on the capacity of cantonal administrations. Under these conditions, the processing of consultation submissions may become susceptible to inconsistencies and inadvertent omissions, even when conducted by experienced legal and administrative experts within the competent cantonal authorities, potentially leading to delayed or partial legal assessment and affecting the coherence of the legislative review process.

For these reasons, Large Language Models (LLMs) have increasingly been applied to the analysis of government documents and citizen-generated content, including public consultation submissions and policy-related feedback (de Lucena et al., 2025). Empirical studies indicate that LLM-based approaches can improve public understanding of complex policy documents by generating accessible summaries and explanations (Yun et al., 2024; Wu et al., 2025), as well as model public attitudes and reactions to policy interventions at scale (Lee et al., 2024). Earlier work in Natural

[1]Office of Statistics and Data, Canton of Fribourg, Switzerland [2]Institute of Artificial Intelligence and Complex Systems, HES-SO University of Applied Sciences and Arts Western Switzerland [3]Office of the Secretary General, Department of Economy, Employment, and Vocational Training, Canton of Fribourg, Switzerland. Correspondence to: Talia B. Kimber <talia.kimber@fr.ch>, Ornella Vaccarelli <ornella.vaccarelli@hefr.ch>.

Proceedings of the Swiss AI Days 2026, Martigny, VS & Fribourg, FR, Switzerland PMLR 309, 2026.

Language Processing (NLP) similarly employed text analysis techniques to study government–citizen interactions and participatory processes (Hubert et al., 2018), underscoring the long-standing relevance of computational methods for supporting public consultation and policy analysis. In the legal domain, a growing body of research shows that LLMs can be adapted to perform structurally and semantically demanding tasks. These include the classification of legal interpretations (Dugac et al., 2025), long-context summarization of judicial decisions (Heddaya et al., 2024), and the development of domain-specific benchmarks for evaluating citation-aware, legally grounded text generation (Zhang et al., 2025). Collectively, these studies highlight the potential of LLM-based methods to support legal text analysis while simultaneously emphasizing persistent challenges related to domain specificity, factual reliability, and the need for evaluation frameworks that integrate retrieval, reasoning, and controlled generation (Pipitone & Alami, 2024; Louis et al., 2024).

Although these developments point to significant opportunities for improving the efficiency of legal information processing, their application to legislative and regulatory contexts, particularly to consultation materials, requires careful methodological design. Legislative texts and consultation submissions are characterized by legal paraphrasing, implicit references to norms, and heterogeneous argumentative structures, making unconstrained generative use inappropriate. Effective deployment therefore necessitates structured output constraints and safeguards to mitigate inconsistency and error in legally consequential settings (Dahl et al., 2024; Bresciani et al., 2024).

In the Swiss context, the use of AI in public administration and the legal domain has become an explicit area of institutional development and applied research. Federal policy documents and guidance emphasize that AI systems deployed by public authorities must comply with principles of legality, transparency, data protection, proportionality, and human oversight, with existing data protection law directly applicable to AI-supported processing (Federal Data Protection and Information Commissioner (FDPIC), 2025; Federal Council of Switzerland, 2025).

At the level of legal research and practice, recent work has focused on adapting natural language processing techniques to Swiss legal corpora, including multilingual benchmarks for the automatic summarization of Federal Supreme Court decisions (Rolshoven et al., 2024). In parallel, professional guidance and legal technology initiatives document the controlled adoption of AI tools in Swiss legal practice, with particular attention to local deployment, confidentiality, and compliance with professional secrecy obligations (LexTech Institute, 2025; K-flash, 2025). In the legislative and parliamentary domain, language technologies have been explored to support access to, retrieval of, and navigation within large collections of legal and legislative texts (Vaccarelli et al., 2025). Together, these developments illustrate a growing recognition in Switzerland that AI can support legal and legislative workflows when embedded within institutionally grounded, legally constrained, and transparent processes.

IA4FriLex (AI for Fribourg's Legislative Text) builds on this momentum through the concrete implementation of an AI-assisted pipeline specifically designed to support legislative consultation procedures. The system has been developed and deployed as a pilot within the Cantonal administration of Fribourg, where it has been used in real consultation processes to assist cantonal services in the analysis and synthesis of consultation submissions (see Figure 1).

Unlike approaches limited to document retrieval or exploratory analysis, IA4FriLex operationalizes the end-to-end transformation of heterogeneous consultation inputs into a structured, standardized, and department-ready legislative report, aligned with established administrative reporting practices. The pipeline integrates open-source LLMs within a controlled and auditable processing framework, enabling automated summarization and structuring while preserving human oversight.

Empirical use within the cantonal administration has demonstrated a substantial reduction in processing time for consultation materials, thereby alleviating a recurrent efficiency bottleneck in the legislative workflow. By relying exclusively on open-source components and on-premise deployment, IA4FriLex further addresses institutional requirements related to transparency, data protection, and customizability, making it suitable for sustained use in a public-sector legal context.

2. Methodology & Implementation

2.1. Pipeline Description

IA4FriLex implements a structured, legally informed pipeline designed to support the transformation of heterogeneous legislative consultation submissions into a standardized consultation report. The methodology underlying the pipeline is grounded in established cantonal administrative practices and has been developed in close collaboration with the Cantonal administration of Fribourg, in particular with the Office of Statistics and Data (SSD, 2025) and the Office of the Secretary General, under the coordination of the Department of Economy, Employment, and Vocational Training (DEEF, 2025). This institutional collaboration informed the methodological design at each stage of the pipeline, ensuring conformity with cantonal procedures, legal standards, and administrative workflows.

The pipeline formalizes the successive stages of consul-

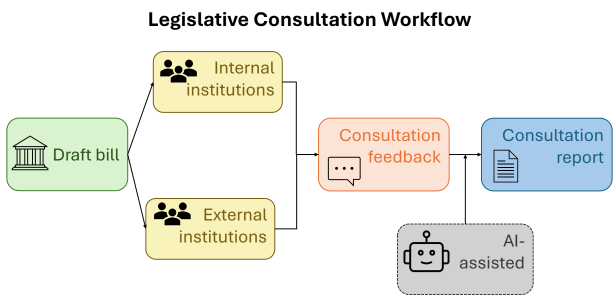

Figure 1. Overview of the legislative consultation process, from draft bill to report generation. IA4FriLex automates key steps of the workflow by generating structured summaries and standardized reports. Consultation feedback typically originates from a broad range of stakeholders, including internal state entities and external actors such as municipalities, regional and intercommunal associations, political parties, professional organizations, and other interested groups.

tation handling, from the reception of submissions to the production of a consolidated report, into a reproducible and auditable workflow. Particular attention has been given to preserving the interpretative role of legal and administrative experts, with automation introduced only for tasks that are repetitive, structurally well-defined, or time-intensive. As illustrated in Figure 2, the system is organized into three principal stages:

1. data collection and normalization,
2. LLM-driven structured summarization, and
3. report generation.

Data sources. The input data consist of institutional and public submissions collected in the context of cantonal legislative consultation procedures conducted by the Canton of Fribourg (Canton de Fribourg, 2025). These submissions relate to draft legislative texts and reflect the diversity of formats and submission practices encountered in real administrative settings. The system supports input files in `.docx`, `.pdf`, and `.msg` formats, which correspond to the predominant channels through which consultation responses are received.

Outlook email files (`.msg`) are systematically parsed to extract both the email body and any attached documents. This design choice reflects established consultation practices, whereby substantive comments may be provided either directly in the body of the message or exclusively within attached files. Processing both components is therefore necessary to ensure the completeness and integrity of each submission and to prevent the inadvertent omission of legally relevant content in subsequent stages of the pipeline. When multiple documents originate from the same entity, they are concatenated into a single submission. In practice, subject-matter experts have not encountered cases where these combined documents contain conflicting statements from the same contributor. However, this mechanism ensures that all inputs are systematically captured should such cases arise. Further details on the composition of the consultation corpus and its use for evaluation are provided in Section 3.

Normalization. All admissible files are converted to Markdown to obtain a uniform text representation suitable for subsequent pipeline steps. Language detection is performed on each Markdown document to inform prompt handling in multilingual cases.

Because Switzerland is a multilingual state with four national languages (German, French, Italian, and Romansh) (Federal Department of Foreign Affairs (FDFA), 2020), and because cantons operate in different primary languages, consultation responses frequently arrive in multiple languages. This linguistic diversity is both a constitutional feature and a practical reality of legislative workflows. As a result, submissions may mix languages across documents and senders. Automatic language detection is therefore a critical preprocessing step. It identifies the source language before prompting the LLM and enables accurate translation. In IA4FriLex, all summaries are translated into French to produce a final report in a single, consistent language.

LLM usage and outputs. Summarization is performed by open-source Large Language Models (LLMs) served via ollama (ollama). Model selection and decoding parameters (e.g., temperature) are configurable; by default, `qwen3:14b` (Yang et al., 2025) is used for remote endpoints, and the temperature is set to 0.8.

Each consultation submission is summarized into a constrained JSON object that follows a predefined schema aligned with established cantonal consultation reporting practices. The default schema includes a *sender* field for identifying the authoring entity of the submission; *general observations* for transversal remarks applicable to the draft law as a whole; *observations law structure* for comments addressing the internal organization or coherence of the draft; *observations report* for feedback related to the explanatory message accompanying the draft legislation; and *miscellaneous comments* for remarks that do not fall under the preceding categories. In addition, article-specific fields *article X* are defined for $X \in [1, A_{\max}]$, where $A_{\max}$ is configurable and set to 100 by default, allowing comments that explicitly or implicitly refer to individual provisions to be captured in a structured manner. This schema supports the systematic extraction and organization of both general and provision-specific feedback, including comments expressed through legal paraphrasing or indirect references to the draft text.

The choice of `qwen3:14b` is driven by practical and methodological considerations. First, Qwen3 offers strong performance on multilingual tasks, including German, French, and Italian, making it well-suited for Swiss consultation documents. Second, the model provides robust long-context handling and stable JSON-constrained generation, both essential for producing legally structured summaries without hallucinated fields or formatting deviations. Third, as an open-source model, Qwen3 supports on-premise or privacy-preserving deployments, a key requirement when handling sensitive administrative data. Finally, its computational footprint strikes a balance between summary quality and inference efficiency, enabling scalable processing of large consultation datasets without excessive hardware demands.

Report generation. The document-level JSON summaries are then aggregated into a consolidated consultation report organized into the following sections:

1. list of senders,
2. general observations,
3. observations on the law structure,
4. article-wise comments,
5. observations on the explanatory message, and
6. miscellaneous comments.

The resulting organization reflects the standard format used in consultation syntheses for multiple legislative projects within the Cantonal administration of Fribourg and has been validated in collaboration with the Office of the Secretary General of the Department of Economy, Employment, and Vocational Training to ensure comprehensive coverage of procedurally and legally relevant aspects. The final report is generated in Markdown format and converted into an official `.docx` document compatible with cantonal administrative standards.

Sender categorization. To facilitate coherent analysis and improve comparability across submissions, IA4FriLex groups responses according to predefined sender categories (state institutions; communes and regional associations; political parties; project-relevant partners; other associations and groups). Such categorization is particularly useful in legislative consultations, where similar stakeholders often share institutional roles, legal responsibilities, or policy perspectives. Grouping responses from comparable entities allows the resulting report to highlight patterns, convergences, and divergences within and across stakeholder groups, making it easier for reviewers to identify consensus positions, sector-specific concerns, or systematic differences in feedback. This structured organization also mirrors existing administrative practices, ensuring that the final report remains aligned with the expectations and workflows of cantonal authorities.

Standardization of inputs and outputs. An important insight from the development and evaluation of IA4FriLex concerns the need for greater standardization in the consultation workflow. Working with highly heterogeneous submissions, ranging from email texts to complex PDF attachments, highlights the operational benefits of introducing more uniform input formats. The project shows that defining a standardized consultation form or a recommended set of fields improves the quality, structure, and completeness of incoming data, while reducing ambiguity during preprocessing and LLM summarization.

In parallel, the project also reveals the value of producing a unified and transversal output format for consultation reports across all state offices. The generated report follows a pre-defined, standardized structure that can be used uniformly across administrative departments, ensuring comparability and coherence. To avoid selective biases, the LLM is explicitly instructed not to exclude any information present in the response documents. In cases of uncertainty, the directive is to faithfully retranscribe the original text rather than infer, hallucinate, or guess missing content. This design ensures that the resulting summaries and the final report remain exhaustive, capturing all remarks and observations without omission. Moreover, every statement included in the consolidated report is traceable to its source: each comment is systematically attributed to the corresponding responding entity, enabling straightforward verification by referring back to the original submission. This built-in attribution supports transparency and allows users to validate any assertion in the report against its primary document. By construction, the pipeline processes every comment across all documents, thereby mitigating risks of selection bias while supporting consistent, department-wide reporting practices.

2.2. Implementation

Preprocessing. File-type verification ensures only `.docx`, `.pdf`, and `.msg` are processed. For duplicate basenames with differing extensions, non-PDF versions are preferred for downstream text extraction, because `.docx` files generally contain richer structural information, such as headings, lists, and explicit paragraph boundaries, which can be extracted more reliably than from `.pdf` documents.

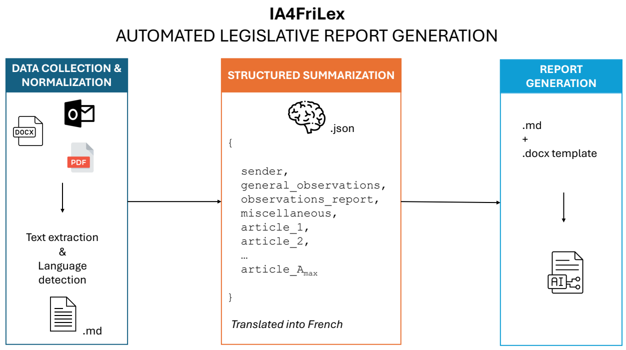

Figure 2. Overview of the IA4FriLex pipeline. Heterogeneous and potentially multilingual consultation inputs (`.docx`, `.pdf`, `.msg`) enter a normalization stage that performs text extraction into Markdown and language detection. Each normalized document is then summarized by an LLM into a *constrained JSON* structure; summaries are translated into French to ensure a single-language output. Finally, all JSON summaries are aggregated into a standardized `.md` report and rendered into a department-specific `.docx` template for immediate administrative use.

PDFs, by contrast, often lose semantic structure during conversion and may introduce layout artifacts (e.g., broken lines, misplaced hyphens, or reordered text), increasing the risk of errors in subsequent LLM summarization. Prioritizing non-PDF sources therefore improves extraction accuracy and reduces noise in the normalized Markdown representation.

Known non-substantive files can be excluded via a configurable list. Text extraction uses Docling (Auer et al., 2024) for `.pdf`/`.docx` and extract-msg (extract msg) for emails; outputs are stored as Markdown. Language detection is performed using langdetect (langdetect).

Summary and report generation. Summary settings, such as language model, temperature, and article count, are stored in configurable JSON files. For each normalized document, the LLM is prompted to produce a JSON object conforming to the template.

All per-document JSON objects are loaded and concatenated field-wise into the target report sections. Sender ordering is implemented via a deterministic, rule-based classifier. The aggregate is rendered to Markdown and converted to `.docx`. To ensure that the generated consultation reports are immediately usable within cantonal administrative workflows, IA4FriLex supports the integration of institution-specific Word templates during report generation.

In collaboration with the Cantonal administration of Fribourg, a dedicated `.docx` template has been designed and made available to the involved services for use in real consultation procedures concerning two legislative projects. The structure and formatting of this template were defined in accordance with established cantonal norms and refined based on feedback from the primary institutional beneficiaries, notably the Office of the Secretary General of the Department of Economy, Employment, and Vocational Training. Beyond ensuring formal conformity, the template has also been conceived to facilitate both human use and automated processing. Its structure guides contributors toward a clear and consistent organization of observations, thereby supporting correct completion by users while simultaneously enabling reliable downstream processing by the IA4FriLex pipeline. This dual objective of usability for contributors and structural clarity for automated analysis has been a key design criterion during template development.

Looking ahead, the possibility of introducing assisted or semi-automated completion of consultation templates, supported by Large Language Models, has been identified as a potential avenue for future work. The goal would be to support contributors, i.e., the entities responding to the consultation (administrations, municipalities, associations, etc.), during the drafting of their submissions. Rather than imposing a rigid form, an assisted template would provide light, contextual guidance (e.g., suggested sections) to help contributors structure their remarks and ensure that key aspects of the draft law are addressed, while still allowing them to express their positions in their own terms. This approach aims to find a balance between fully free-form submissions, which can be unfocused or incomplete, and strict forms that may limit the nuance required in legal or policy feedback, thereby improving both the quality and consistency of contributions without reducing contributors' autonomy.

Runtime and environment. The application is implemented in Python with dependency management via uv (uv), and supports flexible deployment across diverse environments. The full pipeline can run in an institutional datacenter or entirely on a local workstation, without relying on external cloud services; an essential requirement for public administrations seeking data sovereignty, confidentiality, and regulatory compliance. LLM inference is managed via ollama, allowing models to be served locally on CPU-only systems as well as on more capable infrastructure when available. Lightweight models, such as `deepseek-r1:1.5b` (DeepSeek-AI et al., 2025), are primarily used for development, testing, and rapid iteration, and can be executed on commodity hardware with limited computational resources. For production use within real consultation workflows, larger models (e.g., `qwen3:14b`) are deployed to ensure higher-quality summaries, greater linguistic robustness, and improved handling of legally complex submissions. By supporting fully on-premise execution,

Table 1. Overview of the consultation datasets used in the four legislative use cases, including the number of submitted documents and the average document length (in tokens) for each project.

Legislative project	Year	Number of documents	Mean number of tokens
Tourism	2022	44	2079.9
Media Access	2024	21	1142.9
Professional Training	2025	46	1103.1
Healthcare Training	2025	20	1747.0

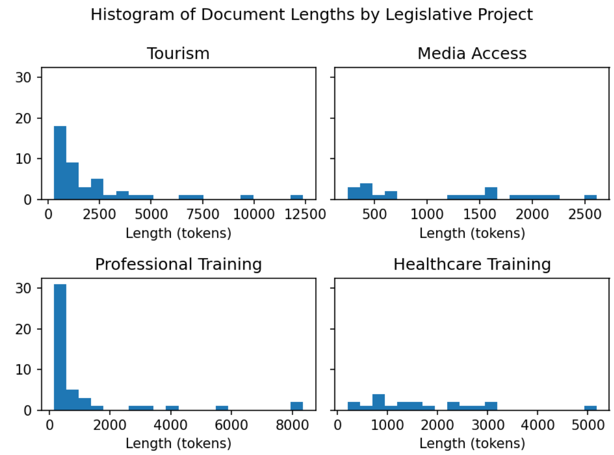

Figure 3. Distribution of document lengths (in tokens) across four legislative projects. Each subplot shows the frequency histogram of individual response document lengths for one law: Tourism, Media Access, Professional Training, and Healthcare Training.

from local personal computers to secured institutional datacenters, the IA4FriLex pipeline ensures privacy-preserving processing of sensitive consultation materials while accommodating the heterogeneous infrastructures of cantonal administrations.

3. Results and Evaluation

3.1. Use cases

IA4FriLex has been evaluated on four cantonal legislative texts selected to reflect both retrospective and prospective consultation scenarios, using real consultation data from the Cantonal administration of Fribourg. The selected use cases cover different types of legislative interventions (revisions of existing laws as well as the introduction of new normative frameworks) and were chosen in collaboration with the cantonal administration to ensure representativeness of actual consultation practice.

To situate the evaluation results and provide a sense of scale, Table 1 summarizes the number of submitted documents per legislative project as well as the average document length in tokens. In addition, Figure 3 presents the distribution of document lengths for each law, illustrating the heterogeneity of the consultation corpora.

For the retrospective evaluation, two legislative texts for which the consultation procedure had already been completed were selected:

- the Law on Tourism (2022) (loi, 2022), representing a substantial revision of the existing cantonal tourism legislation, which entered into force on January 1, 2022;
- the Youth Access to Media Act (adopted in early 2024), which established a new legislative framework rather than amending a pre-existing act (loi, 2024).

For both laws, complete consultation corpora were available, including all public and institutional feedback as well as the official consultation report manually drafted by the administration. This enabled a direct comparison between AI-generated outputs and human-produced consultation reports under realistic conditions.

To complement this retrospective analysis, IA4FriLex has been also applied prospectively to two legislative projects initiated in 2025, for which no consultation synthesis yet existed at the time of analysis:

- the Professional Training Act (LFP, 2025), a reform expected to generate substantial public feedback;
- the Ordinance on the Promotion of Training in the Field of Healthcare (ord, 2025).

In these prospective cases, IA4FriLex has been used to generate the first AI-assisted consultation reports within ongoing legislative processes. This allowed the evaluation of the system not only as a retrospective analytical tool, but as an operational support instrument integrated into real legislative workflows, thereby providing evidence of its applicability and robustness in practical administrative settings.

3.2. Evaluation Method

To assess IA4FriLex, two complementary validation approaches were employed, ensuring both retrospective and prospective evaluation of report quality. First, the *LLM-as-judge* protocol (Lin et al., 2022) has been applied to legislative projects for which a human-authored consultation report was available, enabling a structured comparison between the AI-generated and reference reports. Second, for draft bills without an existing report, expert reviewers qualitatively assessed the AI-generated outputs, focusing on their completeness, coherence, and practical usability within administrative workflows.

Metrics. The evaluation relied on two categories of metrics that jointly capture efficiency and substantive fidelity:

- **Time saved:** We compared the end-to-end processing time of the AI pipeline with the estimated duration of a fully manual drafting process.
- **Summary quality:** We assessed the faithfulness of the AI-generated report using indicators adapted from comparative evaluation frameworks. These included (i) *coverage*, measuring the extent to which observations present in the human report also appear in the AI report, and (ii) *non-contradiction*, ensuring that no AI-generated statements conflict with those in the reference report. The accuracy of the reproduced list of senders has been also evaluated to confirm whether the AI-generated report faithfully reflects the set of entities that submitted feedback. We emphasize that non-contradiction with the human report is a consistency check and does not by itself guarantee factual correctness with respect to the primary submissions. Measuring false positives (unsupported additions) requires explicit source-grounded verification, which we leave to future work via statement-level sampling and annotation.

Qualitative failure taxonomy. Given the constrained use of LLMs in IA4FriLex, potential and practitioner-reported error modes can be organized into a limited set of qualitative categories. These include:

- unsupported additions, where a report statement cannot be traced to the source submission;
- semantic drift or over-generalization, where paraphrasing alters legal scope or modality;
- misplacement errors, in which correct content is assigned to an incorrect article or report section;
- attribution errors, where content is linked to the wrong responding entity;
- omissions due to over-compression, where excessive summarization removes legally relevant detail;
- surface-level linguistic artifacts, including translation inaccuracies or minor typographical issues introduced during text extraction or multilingual processing, without affecting substantive content.

We do not quantify the frequency of these errors in the present study. Instead, this taxonomy clarifies the types of hallucination-related and summarization failures that may occur and provides a structured basis for future, source-grounded false-positive evaluation.

LLM-as-judge protocol. A separate LLM (`gemma3:12b` (Team et al., 2025)), distinct from the model used for summary generation, was prompted to evaluate alignment between the human report R_h and the AI report R_{ai} under three constraints:

1. **Coverage:** All elements listed in the *General Observations* section of the human report R_h must appear somewhere in the AI-generated report R_{ai}. These elements do not need to be located in the same section; they may appear under article-specific comments or in the miscellaneous section. This criterion ensures that no important observations are omitted from the AI report. The LLM-as-judge evaluates coverage by assigning a score from 1 to 5, where 5 indicates that all elements are present and 1 indicates that none are included.
2. **Non-contradiction:** All elements contained in the *General Observations* section of the AI-generated report R_{ai} must not contradict any element in the human report R_h. This criterion ensures that the automated synthesis remains consistent with the reference analysis and does not introduce conflicting interpretations. The LLM-as-judge assigns a score from 1 to 5, where 5 indicates that no contradictions are present and 1 indicates multiple conflicting statements.
3. **Sender completeness:** The AI-generated report R_{ai} should reproduce the same set of senders as the human report R_h, without omissions or additions. To quantify this, we computed the Jaccard index (Jaccard, 1901) between the two sender lists, defined as the size of their intersection divided by the size of their union. A score of 1 indicates perfect overlap (identical sets), while lower values reflect discrepancies in sender identification.

The choice of `gemma3:12b` as the independent evaluation model is motivated by its strong performance in reasoning, judgement, and comparative assessment tasks, as documented in its technical report (Team et al., 2025). Unlike models optimized primarily for generation, Gemma3 is designed to produce stable, instruction-following outputs and to apply consistent criteria when comparing two structured texts. This makes it well suited for the role of an external judge. Moreover, using a model architecturally and parametrically distinct from the summarization model (`qwen3:14b`) reduces the risk of shared biases or correlated failure modes influencing the evaluation. Employing an independent family of LLMs therefore strengthens the methodological robustness of the LLM-as-judge protocol and supports more reliable assessments of coverage, contradiction, and sender completeness.

3.3. Results

Time efficiency. Across the evaluated legislative cases, IA4FriLex demonstrated substantial reductions in processing time. For the Professional Training Act, for example, the administration received 46 distinct consultation responses in multiple formats, languages, and levels of structure. Running the complete IA4FriLex pipeline on this dataset required approximately 40 minutes, including document normalization, LLM-based summarization, and report assembly. Subject matter experts estimated that producing an equivalent first-draft report manually would require several days of concentrated work, leading to a time reduction of at least 80%. These savings reflect the system's ability to handle large volumes of heterogeneous documents in a fully automated manner, while human reviewers focus primarily on verification and interpretation rather than mechanical synthesis.

Quality assessment. The LLM-as-judge evaluation revealed strong alignment between the AI-generated report and the corresponding human-authored reference report. The AI output reproduced the majority of key observations, achieved high coverage, and produced no detected contradictions relative to the human analysis (Figure 4).

For instance, in the revision of the Tourism Act, the coverage score was $4/5$ (80%). According to the LLM-as-judge, the AI report included the majority of essential elements from the human report, though with varying levels of detail, which prevented a perfect score. In contrast, the Youth Access to Media bill achieved full coverage (1.00), demonstrating that all key observations were captured. Regarding non-contradiction, both reports received a perfect score of 1.00, confirming that no conflicting statements were introduced and that the AI-generated syntheses remained consistent with the human analyses.

Sender completeness has been assessed using the Jaccard index (intersection over union) to compare the sets of contributors listed in each report. For the Tourism Act, the Jaccard score was $39/45 = 0.87$, while for the Youth Access to Media bill, it was $17/26 = 0.65$, reflecting high and moderate agreement, respectively. Notably, neither score reached 1 due to minor omissions and erroneous additions in the human reference lists, attributable to manual error rather than deficiencies in the AI pipeline. By enforcing standardized processing, IA4FriLex mitigates such inconsistencies, enhancing both accuracy and traceability in sender identification.

Taken together, these results indicate that IA4FriLex is capable of generating exhaustive first-pass syntheses that remain faithful to the original submissions.

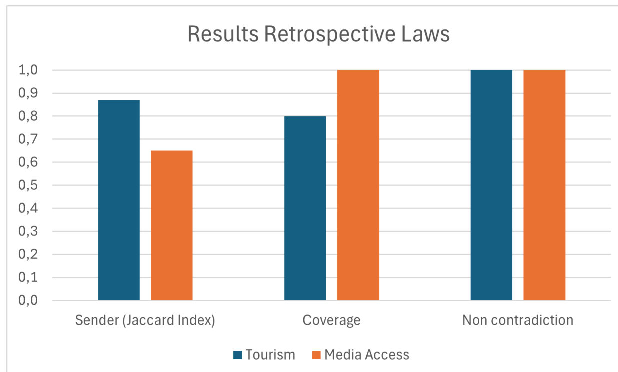

Figure 4. Evaluation of AI-generated reports for two retrospective laws (Tourism Act and Media Access Act), using an LLM-as-judge approach. For the Tourism Act, coverage is 0.80 and non-contradiction is 1.00; for the Media Access Act, both coverage and non-contradiction reach 1.00, indicating complete alignment with the human-written reports. Sender agreement, measured by the Jaccard index, is 0.87 and 0.65 respectively; the fact that these scores are below 1 is explained by minor errors in the human reference lists, rather than omissions by the AI pipeline.

Stakeholder feedback. Administrative staff involved in the evaluation reported a significant reduction in workload and expressed confidence in the structure and completeness of the generated summaries. They highlighted that the standardized format facilitated rapid review and comparison across departments. At the same time, they emphasized the continued importance of human oversight, particularly for interpreting nuanced legal arguments, validating complex or ambiguous statements, and preparing the final version intended for decision-making.

3.4. Validation

The evaluation confirms that IA4FriLex can reliably generate exhaustive first-pass consultation reports and significantly accelerate the early stages of legislative analysis. However, the results also highlight important boundaries regarding the intended use of the system. The AI-generated reports are well-suited for exploration, pre-analysis, and the production of an exhaustive report that compiles and structures all comments without selective omission. This aligns with the core objective of the pipeline: generating a standardized, comprehensive summary that can serve as a solid basis for human review. By contrast, the reports are not intended to replace expert-validated outputs for decision-making, public communication, or official publication. While the evaluations show high content coverage and no detected contradictions, the process still depends on human supervision for interpretation, legal nuance, and final editorial judgment. In summary, IA4FriLex should be understood as a tool for generating an exhaustive, AI-assisted baseline report, which supports analysts by reducing mechanical workload while preserving the need for professional oversight.

4. Conclusion

This paper presented IA4FriLex, an AI-assisted system designed to support legislative consultation procedures through the structured processing and synthesis of consultation submissions. Developed and deployed in collaboration with the Cantonal administration of Fribourg, the system has been applied to real cantonal legislative projects, demonstrating its feasibility and relevance under operational public-sector conditions.

Evaluation across four legislative cases shows that IA4FriLex reliably generates high-quality first-pass syntheses that capture the full range of stakeholder feedback, while preserving the interpretative role and legal responsibility of human experts. In operational use, the system reduced the time required for the compilation and structuring of consultation material by $\sim 80\%$, thereby allowing legal and administrative experts to concentrate on substantive interpretation, validation, and policy analysis. The system formalizes consultation outputs using a standardized JSON representation aligned with established cantonal consultation structures and intended to map directly to an institutionally approved Word template. Although widespread adoption of this template by all consultation stakeholders will require a transition period, the approach already improves consistency, traceability, and comparability across legislative projects while mitigating the risk of inconsistency and inadvertent human error.

A key contribution of this work lies in its institutional and legal grounding. IA4FriLex has been designed to operate fully on-premise rather than relying on external or third-party cloud services, either within secured administrative infrastructures or on local workstations, ensuring compliance with public-sector requirements related to data protection, confidentiality, and data sovereignty. The exclusive use of open-source software and models further enhances transparency, auditability, and long-term maintainability, which are essential considerations in legislative and administrative contexts.

Beyond its immediate application in the Cantonal administration of Fribourg, IA4FriLex illustrates how carefully constrained uses of Large Language Models can support participatory law-making without undermining legal accountability or democratic principles. The modular architecture of the system enables adaptation to other cantonal or state-level settings and provides a concrete pathway toward scalable, AI-assisted consultation processes.

Future work. During the project, a recurring challenge was the heterogeneity of consultation responses. Because submissions arrive in diverse formats, structures, and levels of formality, the resulting variability can limit the quality of automatic processing and lead to avoidable ambiguities. This raises an important avenue for further development: the design of a digital consultation response form potentially enhanced with AI assistance that would guide respondents toward a more structured, uniform, and complete submission. Such a tool could improve data quality at the source, reduce preprocessing complexity, and strengthen the coherence of downstream analyses produced by IA4FriLex.

Moreover, to strengthen future evaluations, we aim to implement a standardized evaluation instrument enabling users to assess completeness, faithfulness to submissions, usability, and source grounding. Future deployments should rely on structured feedback to enable more rigorous and comparable assessments.

In addition, an extension based on a chatbot-style interface powered by a Retrieval-Augmented Generation (RAG) architecture (Lewis et al., 2020) could enable interactive querying consultation responses. Administrative staff would be able to ask natural-language questions and receive precise, context-aware answers drawn from the original submissions. Such an interactive layer would enhance accessibility, support rapid information retrieval, and complement the standardized reporting pipeline by providing a dynamic exploration tool for legislative feedback.

Acknowledgements

The authors thank Martial Clément for his valuable insights on the applicability of the tool to state-level projects, and Quoc Bao Luu for sharing his expertise in software development. We also thank the Office of the Secretary General of the Department of Economy, Employment, and Vocational Training for providing the data used in this study and for their support throughout the project. We are grateful to the Department of Health and Social Services for their constructive evaluation, and to the Department of Institutions, Agriculture, and Forestry for additional data.

Limitations

IA4FriLex was evaluated on a limited number of legislative consultations within a single cantonal context, which may constrain its generalizability beyond similar administrative settings. The system also faces inherent challenges in determining the appropriate level of detail in summaries: it must balance faithful retranscription of legally relevant content with necessary compression, and both omission and over-verbosity remain possible despite schema constraints. Furthermore, the heterogeneity of input documents—across formats, languages, and structural conventions—can introduce extraction errors and inconsistencies that propagate into the summaries. From an operational perspective, the quality and speed of summarization depend on access to adequate computational resources. While the pipeline can run on CPUs, high-quality processing with larger open-source models is significantly more efficient on GPU-equipped servers, which may limit adoption for organizations lacking such infrastructure. Finally, although the system generates first-pass syntheses, expert review remains essential for validating legal nuance, interpreting ambiguous statements, and preparing materials for official decision-making.

Ethical Statement

This study used only administrative documents provided by public authorities and contained no personal or sensitive data. Expert feedback reflected professional assessments and did not constitute human subjects research.

References

Règlement sur l'élaboration des actes législatifs (real). https://bdlf.fr.ch/app/fr/texts_of_law/122.0.21. Accessed: 2025-12-15.

An extremely fast python package and project manager. https://docs.astral.sh/uv/. Accessed: 2025-12-15.

Swiss Federal Constitution, Article 147 (Consultation). https://www.fedlex.admin.ch/eli/cc/1999/404/en#art_147, 1999. SR 101; accessed 2025-12-16.

Loi sur le tourisme, banque de données de la législation fribourgeoise (bdlf). https://bdlf.fr.ch/app/fr/texts_of_law/951.1, 2022. Accessed: 2025-12-15.

Loi sur l'accès des jeunes aux médias, banque de données de la législation fribourgeoise (bdlf). https://bdlf.fr.ch/app/fr/texts_of_law/954.1/, 2024. Accessed: 2025-12-15.

Avant projet de révision totale de la loi sur la formation professionnelle. https://www.fr.ch/etat-et-droit/legislation/, 2025. Accessed: 2025-12-15.

Ordonnance sur l'encouragement de la formation dans le domaine des soins. https://www.fr.ch/etat-et-droit/legislation/, 2025. Accessed: 2025-12-15.

Auer, C., Lysak, M., Nassar, A., Dolfi, M., Livathinos, N., Vagenas, P., Ramis, C. B., Omenetti, M., Lindlbauer, F., Dinkla, K., Mishra, L., Kim, Y., Gupta, S., de Lima, R. T., Weber, V., Morin, L., Meijer, I., Kuropiatnyk, V., and Staar, P. W. J. Docling technical report, 2024. URL https://arxiv.org/abs/2408.09869.

Bresciani, P., Palmirani, M., et al. Constitutional opportunities and risks of ai in the law-making process. *Federalismi.it*, 2:1–18, 2024. URL https://cris.unibo.it/handle/11585/953858.

Dahl, M., Magesh, V., Suzgun, M., and Ho, D. E. Large legal fictions: Profiling legal hallucinations in large language models. *Journal of Legal Analysis*, 16(1):64–93, 06 2024. ISSN 2161-7201. doi: 10.1093/jla/laae003. URL https://doi.org/10.1093/jla/laae003.

de Lucena, D. C., Souza, E., Albuquerque, H. O., Félix, N., Oliveira, A. L., and de Carvalho, A. C. Performance analysis of llms for abstractive summarization of brazilian legislative documents. *Conference on Digital Government Research*, 26, May 2025. doi: 10.59490/dgo.2025.969. URL https://proceedings.open.tudelft.nl/DGO2025/article/view/969.

DEEF. Direction de l'économie, de l'emploi et de la formation professionnelle. https://www.fr.ch/deef, 2025. Accessed: 2025-12-15.

DeepSeek-AI, Guo, D., Yang, D., and et al. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning, 2025. URL https://arxiv.org/abs/2501.12948.

Dugac, G. et al. Classifying legal interpretations using large language models. *Artificial Intelligence and Law*, 2025. doi: 10.1007/s10506-025-09447-9. URL https://link.springer.com/article/10.1007/s10506-025-09447-9.

extract msg. https://pypi.org/project/extract-msg/. Accessed: 2025-12-15.

Federal Council of Switzerland. AI regulation: Federal Council to ratify Council of Europe Convention. https://www.news.admin.ch/en/nsb?id=104110, 2025. Accessed 2026-01-08.

Federal Data Protection and Information Commissioner (FDPIC). Current data protection legislation is directly applicable to AI. https://www.edoeb.admin.ch/en/update-current-legislation-directly-applicable-ai, 2025. Accessed 2026-01-08.

Federal Department of Foreign Affairs (FDFA). Four languages, four cultures – one country, 2020. URL https://www.eda.admin.ch/eda/en/fdfa/fdfa/aktuell/newsuebersicht/2020/09/mehrsprachigkeit-1.html. Accessed: 2025-12-17.

Heddaya, M., MacMillan, K., Malani, A., Mei, H., and Tan, C. Casesumm: A large-scale dataset for long-context summarization from u.s. supreme court opinions, 2024. URL https://arxiv.org/abs/2501.00097.

Hubert, R. B., Estevez, E., Maguitman, A., and Janowski, T. Examining government-citizen interactions on twitter using visual and sentiment analysis. In *Proceedings of the 19th Annual International Conference on Digital Government Research: Governance in the Data Age*, dg.o '18, New York, NY, USA, 2018. Association for Computing Machinery. ISBN 9781450365260. doi: 10.1145/3209281.3209356. URL https://doi.org/10.1145/3209281.3209356.

Jaccard, P. Étude comparative de la distribution florale dans une portion des alpes et du jura. *Bulletin de la Société Vaudoise des Sciences Naturelles*, 37:547–579, 1901.

K-flash. Switzerland's Legal Solutions with AI. https://k-flash.ch/2025/05/30/switzerlands-legal-solutions-for-ai/, 2025. Accessed 2026-01-08.

langdetect. https://pypi.org/project/langdetect/. Accessed: 2025-12-15.

Lee, S., Peng, T.-Q., Goldberg, M. H., Rosenthal, S. A., Kotcher, J. E., Maibach, E. W., and Leiserowitz, A. Can large language models estimate public opinion about global warming? an empirical assessment of algorithmic fidelity and bias. *PLOS Climate*, 2024. doi: 10.1371/journal.pclm.0000429. URL https://doi.org/10.1371/journal.pclm.0000429.

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W.-t., Rocktäschel, T., Riedel, S., and Kiela, D. Retrieval-augmented generation for knowledge-intensive nlp tasks. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, NIPS '20, Red Hook, NY, USA, 2020. Curran Associates Inc. ISBN 9781713829546.

LexTech Institute. AI in Swiss Legal Practice: Balancing innovation with professional obligations. https://www.lextechinstitute.ch/lintelligence-artificielle-dans-la-pratique-contractuelle-des-avocat%C2%B7e%C2%B7s-suisses/?lang=en, 2025. Accessed 2026-01-08.

Lin, S., Hilton, J., and Evans, O. Truthfulqa: Measuring how models mimic human falsehoods, 2022. URL https://arxiv.org/abs/2109.07958.

Louis, A., van Dijck, G., and Spanakis, G. Interpretable long-form legal question answering with retrieval-augmented large language models. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(20):22266–22275, Mar. 2024. doi: 10.1609/aaai.v38i20.30232. URL https://ojs.aaai.org/index.php/AAAI/article/view/30232.

ollama. https://ollama.com/. Accessed: 2025-12-15.

Pipitone, N. and Alami, G. H. Legalbench-rag: A benchmark for retrieval-augmented generation in the legal domain, 2024. URL https://arxiv.org/abs/2408.10343.

Rolshoven, L., Rasiah, V., Brügger Bose, S., Stürmer, M., and Niklaus, J. Unlocking legal knowledge: A multilingual dataset for judicial summarization in switzerland. *arXiv*, 2024. URL https://arxiv.org/abs/2410.13456.

SSD. Service de la statistique et de la donnée. https://www.fr.ch/deef/ssd, 2025. Accessed: 2025-12-15.

Team, G., Kamath, A., Ferret, J., and et al. Gemma 3 technical report, 2025. URL https://arxiv.org/abs/2503.19786.

Canton de Fribourg. Consultations cantonales en cours. https://www.fr.ch/cha/actualites/consultations-cantonales-en-cours, 2025. Accessed: 2025-12-15.

Vaccarelli, O., Brenot, E., De Salis, E., Reis, H. M., Kucera, J., Meyer, P., Ghorbel, H., and Hennebert, J. Parlementaire. *AI days HES-SO'25*, 2025, 2025.

Wu, S., Xiong, Y., Cui, Y., Wu, H., Chen, C., Yuan, Y., Huang, L., Liu, X., Kuo, T.-W., Guan, N., and Xue, C. J. Retrieval-augmented generation for natural language processing: A survey, 2025. URL https://arxiv.org/abs/2407.13193.

Yang, A., Li, A., Yang, B., Zhang, B., Hui, B., Zheng, B., Yu, B., Gao, C., Huang, C., Lv, C., Zheng, C., Liu, D., Zhou, F., Huang, F., Hu, F., Ge, H., Wei, H., Lin, H., Tang, J., Yang, J., Tu, J., Zhang, J., Yang, J., Yang, J., Zhou, J., Zhou, J., Lin, J., Dang, K., Bao, K., Yang, K., Yu, L., Deng, L., Li, M., Xue, M., Li, M., Zhang, P., Wang, P., Zhu, Q., Men, R., Gao, R., Liu, S., Luo, S., Li, T., Tang, T., Yin, W., Ren, X., Wang, X., Zhang, X., Ren, X., Fan, Y., Su, Y., Zhang, Y., Zhang, Y., Wan, Y., Liu, Y., Wang, Z., Cui, Z., Zhang, Z., Zhou, Z., and Qiu, Z. Qwen3 technical report, 2025. URL https://arxiv.org/abs/2505.09388.

Yun, L., Yun, S., and Xue, H. Improving citizen-government interactions with generative artificial intelligence: Novel human-computer interaction strategies for policy understanding through large language models. *PLOS ONE*, 2024. doi: 10.1371/journal.pone.0311410. URL https://doi.org/10.1371/journal.pone.0311410.

Zhang, K., Yu, W., Dai, S., and Xu, J. Citalaw: Enhancing llm with citations in legal domain, 2025. URL https://arxiv.org/abs/2412.14556.