From Formal Transparency to Practical Interpretability: WOOLens for Open Government Data

Yelyzaveta Terentieva
University of Amsterdam
Amsterdam, Netherlands
yelyzaveta.terentieva@student.uva.nl

Joris Wechsler
University of Amsterdam
Amsterdam, Netherlands
joris.wechsler@student.uva.nl

Cynthia de Vries
University of Amsterdam
Amsterdam, Netherlands
cynthia.de.vries@student.uva.nl

Teun Jans
University of Amsterdam
Amsterdam, Netherlands
teun.jans@student.uva.nl

Jaap Kamps
University of Amsterdam
Amsterdam, Netherlands
kamps@uva.nl

Abstract

Access to government information is a key condition for democratic accountability, yet formal transparency does not automatically ensure practical usability. Many countries have adopted Freedom of Information Acts, offering unprecedented access to government documents, but their full potential is often not realised. For example, in the Netherlands, Woo dossiers are typically published as merged PDF files with limited structure, making them difficult to navigate. This paper presents WOOLens, an end-to-end pipeline that transforms Woo dossiers into structured, navigable formats through document segmentation, type classification, date extraction, and timeline visualisation. We evaluate two pipeline variants—an OCR-based and a VLM-based approach—and report a Panoptic Quality score of 0.640 for the VLM pipeline compared to 0.458 for the OCR baseline. A usability evaluation complements the technical assessment. Our findings reinforce the argument that transparency must be designed, not merely declared.

Keywords

open government data, Freedom of Information, document processing, page stream segmentation, AI, transparency

ACM Reference Format:
Yelyzaveta Terentieva, Joris Wechsler, Cynthia de Vries, Teun Jans, and Jaap Kamps. 2026. From Formal Transparency to Practical Interpretability: WOOLens for Open Government Data. In *Proceedings of Proceedings of the First International Workshop on Artificial Intelligence and Open Government (AIOG), co-located with the 21st International Conference on Artificial Intelligence and Law (ICAIL 2026).* ACM, New York, NY, USA, 9 pages.

1 Introduction

Access to government information is a key condition for democratic accountability [2]. Increasingly, open government information is available under so-called FOIA or Freedom of Information Acts,[1] offering great potential but also new accessibility challenges [16, Sec. 9.5]. In the Netherlands, this access is regulated through the *Wet open overheid* (Woo), the Dutch freedom of information law, which was introduced to support this access and to enable interested individuals, such as journalists, lawyers, investigators, and concerned citizens, to request government documents [21]. The WooZM platform makes government information released under the Woo searchable, accessible, and reusable on a large scale [20]. However, while this improves access at the level of discovery, individual Woo dossiers often remain difficult to interpret in practice.

Woo dossiers are typically published as large, inconsistently structured merged PDF files that combine different types of material, including emails, letters, scans, screenshots, and messaging records, without clear boundaries, standardised metadata, or chronological order. Although inventory lists are often included, they usually provide only limited information, such as document names or redaction grounds, while lacking essential navigation details, including dates, document types, and page references. The official guidance for civil servants provides only vague instructions and does not set formal requirements for machine-readability or metadata inclusion [12, p. 19]. As a result, users are often forced to read entire dossiers sequentially in order to reconstruct even basic information. This makes the material time-consuming to work through and difficult to interpret in practice.

This paper addresses that gap by developing an end-to-end pipeline that transforms Woo dossiers into a structured and navigable format. The WOOLens pipeline includes document segmentation, document type classification, date extraction, timeline visualisation, and, where relevant, summarisation with the support of LLM tools. By combining these steps, the project aims to make it easier to understand what documents are present in a dossier and how they relate to one another over time.

The rest of this paper is structured as follows. Section 2 discusses related work on transparency of government information. Section 3 details our research design and methodology. Section 4 evaluates the system in terms of document segmentation, classification performance, and usability. We discuss our findings in Section 5 and draw initial conclusions in Section 6.

2 Background and Related Work

In this section, we frame our research by discussing related work on the (lack of) transparency of government information.

[1] E.g., US's https://www.foia.gov/, UK's https://www.legislation.gov.uk/ukpga/2000/36/contents, and EU's https://commission.europa.eu/about/service-standards-and-principles/transparency/freedom-information_en.

This gap between formal transparency and practical usability is well documented in research on open government data. As Gao and Janssen [7] put it, "Nevertheless, data is useless unless it is used to generate benefits from it." Also Cucciniello and Nasi [5] make this point clearly by arguing that formal transparency is not the same as effective transparency. Governments may publish information because they are legally required to do so, but this does not necessarily mean that they provide the kind of information citizens need or are able to use. This distinction between nominal transparency, where information is formally available, and effective transparency, where information can actually be used, is particularly visible in Freedom of Information regimes such as the Dutch Woo. While the law guarantees access to government documents, the way these documents are published often introduces significant practical barriers.

Janssen et al. [10] help explain this gap through the concept of "transparency by design." Instead of understanding transparency as something achieved after publication, they argue that it should be embedded in the design of information systems and administrative processes from the beginning. Systems that do not account for accessibility and interpretability may formally meet transparency requirements while still failing to make information meaningfully usable. In the case of Woo dossiers, the problem is therefore not only access, but also how documents are structured, grouped, and published. When information is released in formats that offer little support for interpretation, transparency remains procedural rather than functional. The result is a persistent gap between formal disclosure and effective public access.

Previous research [17] has addressed parts of this problem, most notably the detection of document boundaries in merged PDF files. Yet this solves only one part of a broader issue. Making Woo dossiers technically accessible is not enough to make them understandable for non-expert users. While previous work has focused primarily on document boundary detection, WOOLens extends this approach into a broader interpretive pipeline that supports not only segmentation, but also navigation, contextualisation, and practical use.

3 Research Design and Methodological Approach

In this section, we will detail our research methodology and the front-end and back-end design of our prototype system.

This project is situated within the broader context of Design Science Research (DSR), which aims to address organisational or societal challenges through the design and subsequent evaluation of IT artefacts [9]. The process outlined by Peffers et al. [14] for conducting DSR projects supported this research project in building the artefact through iterative processes of problem identification, objective definition, artefact design, and evaluation. The problem was identified based on previous research and knowledge by the ICAI OpenGov Lab team,[2] and the exploration of their WooZM dataset by our research team.[3] The initial problem identification was expanded and refined through continuous product development and demonstration to different stakeholders, as well as the team's own experience. Subsequently, our prototype system, called WOOLens, is developed as a research artefact that explores how AI can make Woo dossiers more interpretable in practice. A formal technical and user-centred evaluation is discussed in the next chapter and serves as the starting point for further development avenues for this artefact.

This approach brings together two forms of knowledge. On the one hand, it draws on an understanding of how people work with information, including what makes documents difficult to navigate and what kinds of structure support interpretation. On the other hand, it engages with technical possibilities, namely, what computational methods can do to reorganise and present complex material in a more usable way. WOOLens therefore combines an interpretive concern with usability and access with a technical exploration of how AI-driven processing can support these aims.

3.1 Tool View Compared to the Original PDF Pages

Figure 1 and Figure 2 illustrate how WOOLens reconstructs documents from their raw PDF form into structured, readable interface views. Each comparison shows the rendered output of our tool alongside the corresponding pages from the original published Woo dossier, demonstrating both the accuracy of the reconstruction and the improved readability across different document types.

Figure 1 shows how e-mails are grouped by their reply structure ("RE") and presented in a recognisable e-mail interface. It also includes the corresponding original PDF page from the same dossier, showing how WOOLens transforms the source material into a more readable and structured view.

Similarly, Figure 2 shows chat messages in WOOLens grouped by day within a single conversation view, alongside an original PDF page from the same exchange. In the original dossier, each message appears as a separate record with its own metadata, making the conversational flow difficult to follow without manual reconstruction.

Together, these figures demonstrate how WOOLens transforms fragmented, document-centric source material into coherent, context-aware views that more closely reflect how the original communication took place.

3.2 The Front End Design

Figure 3 shows the front-end design. The user interface takes the form of a single-page web application. The Woo dossier is selected through the search bar or manually updated, and offers customisation and enhanced context selection, such as including an overview list of documents that often accompany Woo dossiers.

Figure 4 gives a detailed view of an email extracted from a processed Woo dossier. The left panel displays the processed documents with type badges, dates, and page counts; a central viewer with tabs for emails, chats, a chronological timeline, and an AI-generated summary; and a right panel showing document statistics, named entities, and redaction codes. Emails and chats are presented in recognisable conversation interfaces, visually distinguishing senders and recipients, with email threads grouping reply chains chronologically. Redactions are marked inline and reference the legal ground for withholding; hovering over a redaction surfaces an AI-generated explanation of what that ground means in practice.

[2]https://opengov.nl/
[3]https://woozm.nl/

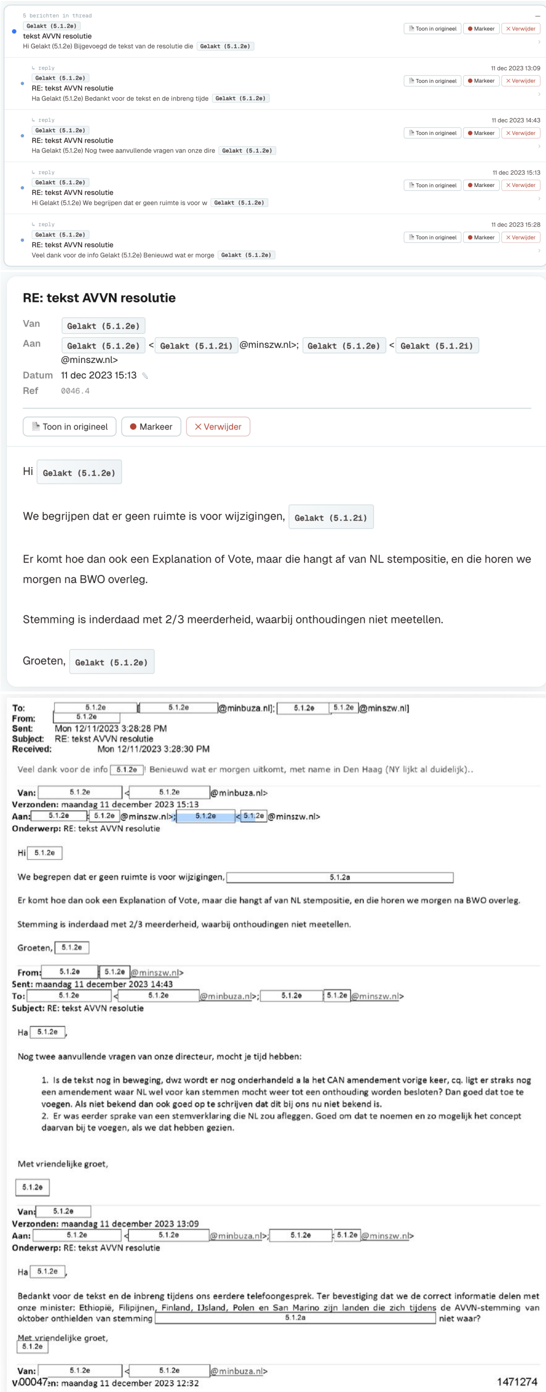

Figure 1: Comparison between the original PDF e-mail material and the reconstructed WOOLens views. Top: e-mail preview. Middle: reconstructed e-mail view in WOOLens. Bottom: original PDF page from the published Woo dossier.

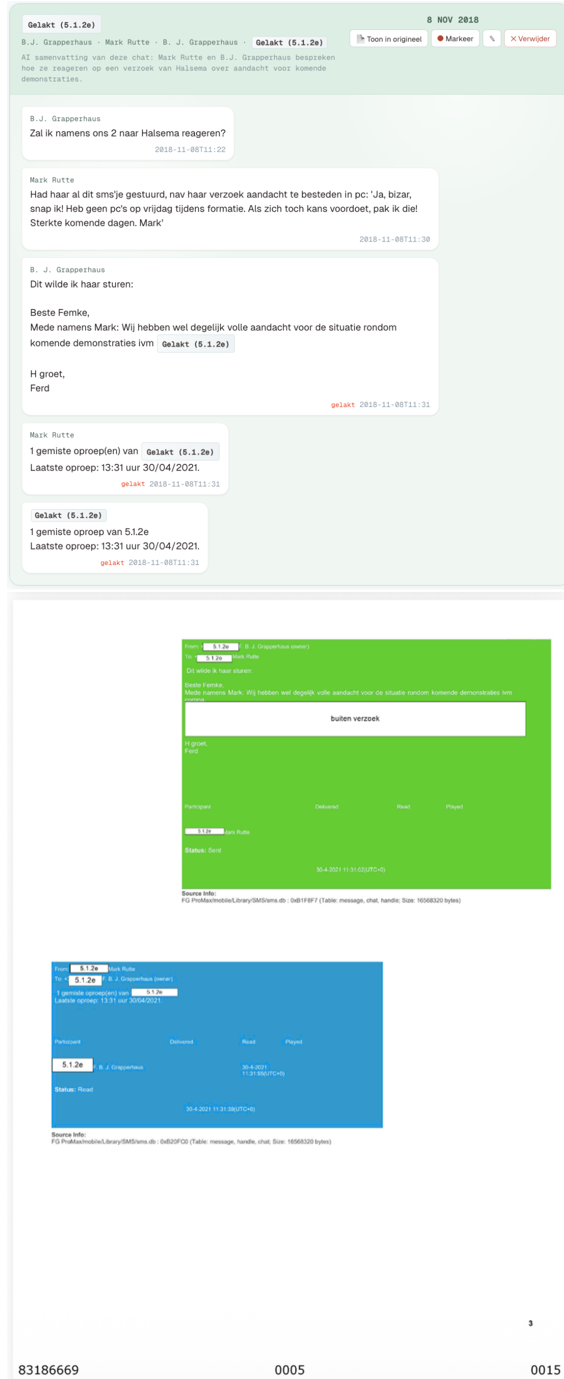

Figure 2: Comparison between the reconstructed WOOLens chat view and the original PDF page from the published Woo dossier. Top: WOOLens interface. Bottom: original PDF page.

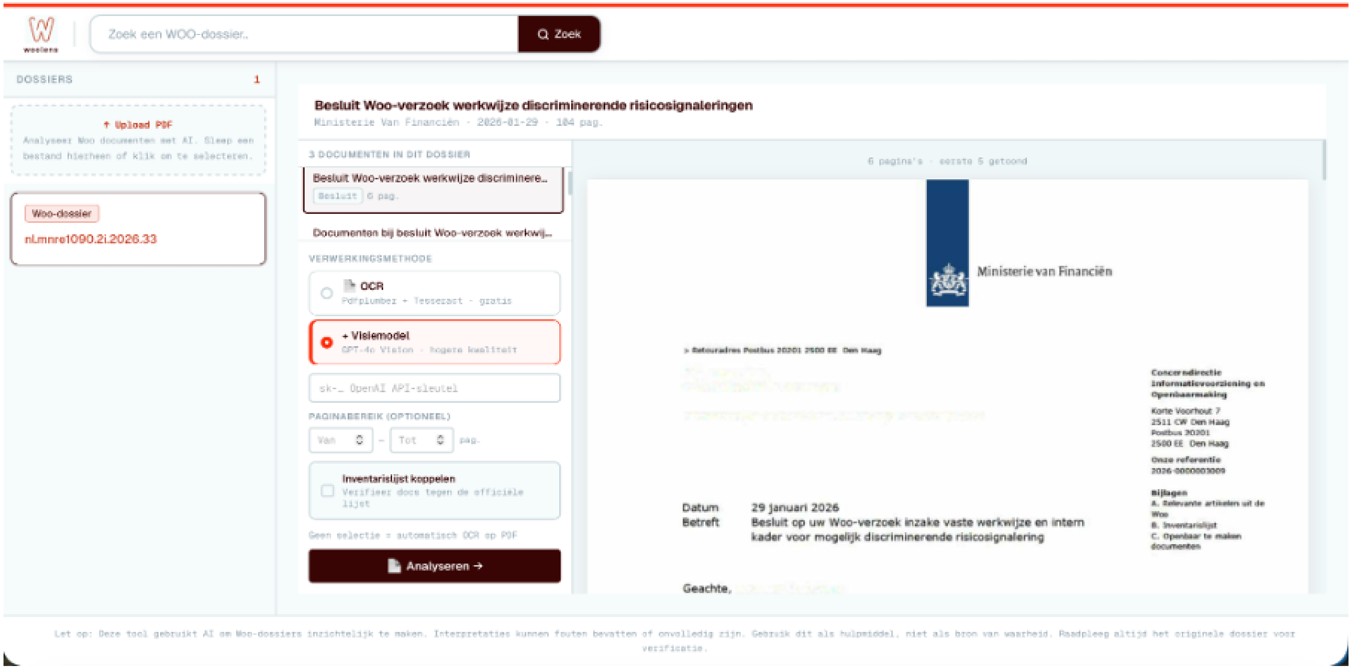

Figure 3: WOOLens interface: dossier upload and pipeline selection.

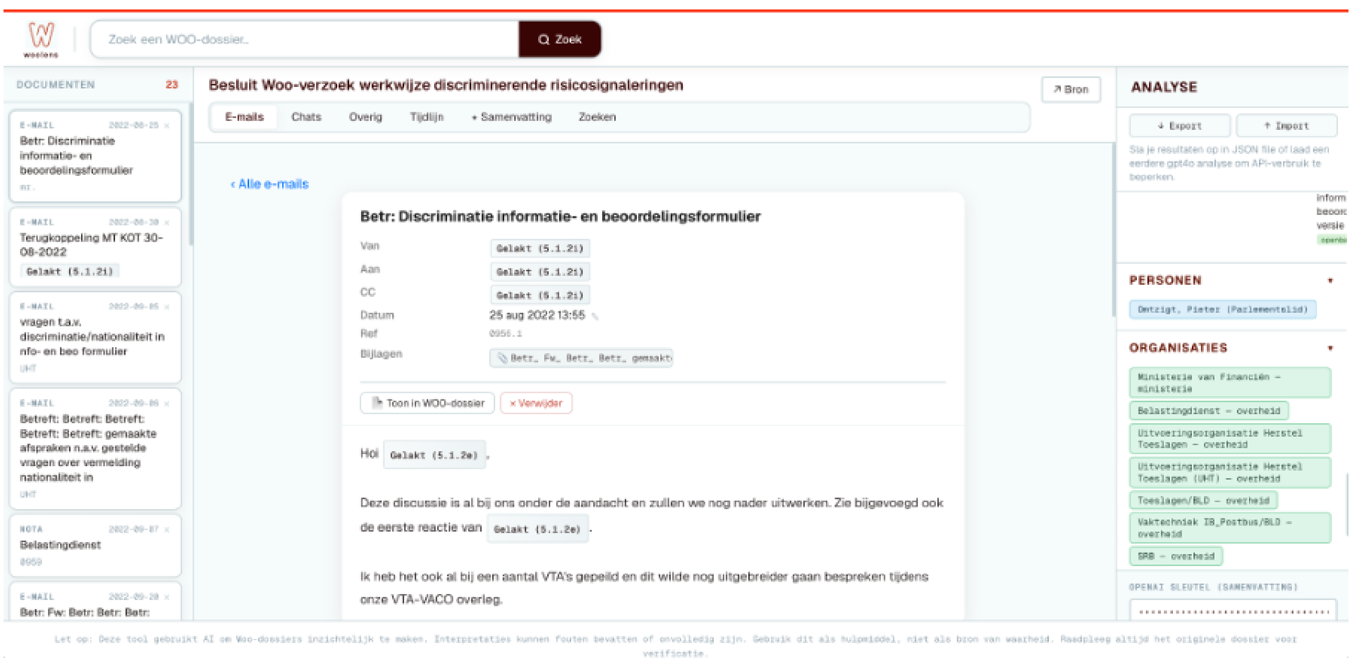

Figure 4: Email detail view with redaction markers.

Figure 5 shows the extracted timeline overview of a Woo dossier. The timeline tab presents documents in chronological order with an optional AI-generated summary per time period. A full-text search bar allows users to filter across all documents by keyword. The right panel surfaces persons and organisations identified across the dossier, each accompanied by an AI-generated contextual explanation. A transparency score indicates the proportion of redacted

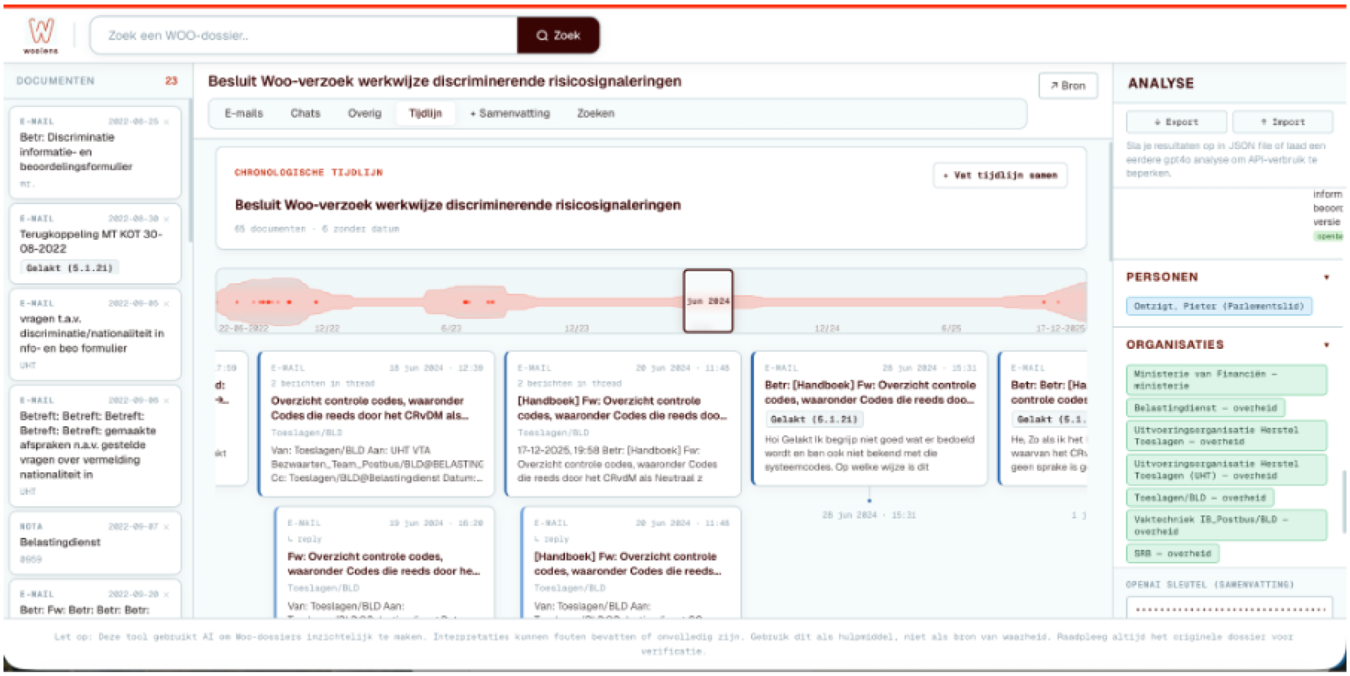

Figure 5: Chronological timeline view of processed documents.

versus accessible content across the dossier. Every identified element, whether an email, chat, document, or entity, includes a direct link to its location in the original PDF, allowing users to verify findings against the source material. The interface supports export and import of processed dossiers as JSON, enabling users to save and reload their work without reprocessing and limit API usage. The frontend was developed with the assistance of Claude Code [4], Anthropic's AI-powered coding assistant, using Visual Studio Code as the development environment, which was used to generate and iterate on the HTML, CSS, and JavaScript output.

3.3 The Back End Design

The WOOLens tool extracts the documents and metadata within a Woo dossier in one of two separate pipelines: one locally run OCR-based pipeline and one VLM-based pipeline. Both pipelines create the same structured output and visualisation, but differ in performance and cost of use, which is further discussed under 'Component Evaluation'. The OCR-based pipeline relies on rules-based determinants of document segmentation and type identification. It primarily uses the existing embedded text layer, and alternatively extracts text using Tesseract (v5) if a page is image-based. This pipeline relies heavily on scanning for four-digit document stamps on some Woo dossiers, which are present on page corners. If no stamps are detected, the pipeline segments documents based on within-document page counters or e-mail headers. Finally, document types are identified using keyword-pattern scoring on the first page of identified documents.

The VLM-based pipeline uses a combined multi-pass approach with GPT-4o[4] and GPT-4o-mini[5] models. The pre-processing pass scans the pages using Tesseract OCR for stamps, similar to the OCR-based pipeline, as well as performing an analysis of document codes on a sample of eight pages per stream, which then informs the prompt in subsequent passes. The first pass runs each page through GPT-4o with great detail to extract text, document codes, boundary signals and document categories. The second pass only activates in documents where no consistent stamps are found and feeds a per-page summary to GPT-4o-mini and returns document boundary decisions based on a variety of signals. The final pass performs full e-mail extraction from documents classified as e-mails, including reply-chain structure and e-mail headers to visualise e-mails effectively.

4 Document Segmentation and Classification Performance Evaluation

In this section, we detail the system-centred evaluation of the back end in terms of document segmentation and document type classification quality, as well as user-centred evaluation of the system's overall usability.

4.1 Component Evaluation

Evaluating a system like WOOLens requires more than a single performance metric. Drawing on Ratner and Thylstrup [15]'s framework of ecologies of evaluation, we approach evaluation not simply as a final validation step, but as a structured inquiry into how the

[4]https://developers.openai.com/api/docs/models/gpt-4o
[5]https://developers.openai.com/api/docs/models/gpt-4o-mini

Table 1: Descriptive Statistics of the Evaluation Dossiers based on ground truth annotation

	Max.	Min.	Median	Average
Number of Pages	217.0	14.0	58.0	70.7
Number of Documents	99.0	6.0	25.0	30.0
Document length in pages	19.0	1.0	1.0	2.4

system performs across different dimensions that matter to different users and contexts. In practice, this means that the evaluation addresses two related questions: first, whether the pipeline does what it is technically designed to do, and second, whether the resulting interface actually helps users navigate a Woo dossier more effectively.

As demonstrated by Van Heusden et al. [17], splitting concatenated PDF streams into their respective documents is a foundational step for the usability and retrieval of FOIA request data. This is especially relevant for the Dutch freedom of information regime under the Wet open overheid (Woo), where disclosure often takes the form of large PDF dossiers rather than separate, consistently described records. In such cases, the legal act of disclosure does not by itself provide users with a navigable document collection, because document boundaries, dates, types, and sequences still have to be reconstructed. This is further complicated by the limited data standardisation across ministries, municipalities, and other institutions subject to the Woo law [18]. This task is referred to as Page Stream Segmentation (PSS) and has been studied in various national contexts using different computational techniques, with recent results showing promising performance for models that combine textual and visual features [17]. In addition, fine-tuned decoder-based LLMs have shown promising results on text-based PSS tasks compared to encoder models, yet vision-language models (VLMs) remain unexplored [8]. Due to time constraints and to explore these research branches, a combination of GPT-4o and GPT-4o-mini was used to test whether a VLM can improve PSS across a set of highly heterogeneous documents. The multi-pass design combined visual page analysis with text-based boundary detection and was continuously fine-tuned to fit common Woo dossier types. A detailed pipeline overview can be found in the Appendix A. This performance is then compared to a deterministic OCR-based pipeline using pdfplumber and Tesseract V5 to extract text from documents and split based on regex-based stamp codes that are the most common document indicators on Woo dossiers.

To evaluate the technical performance of the PSS task, a ground-truth dataset was manually annotated from 10 Woo dossiers. The dossiers were chosen based on temporal and structural variety, in order to reflect a wide variety of document types that are the main difficulty for PSS in open government data. Therefore, the dossiers were chosen from the most frequently occurring government bodies that appear in the WooZM database at the time of writing and released within the past five years, to reflect the most up-to-date document structures. An overview of the sample is shown in Table 1. The evaluation metrics Panoptic Quality (PQ), Segmentation

Table 2: Document Segmentation and Document Type Classification results of OCR and VLM-based Pipelines

	OCR-based				VLM-based			
Dossier	RQ	SQ	PQ	Type	RQ	SQ	PQ	Type
A	0.828	0.990	0.819	0.600	0.825	0.969	**0.799**	**0.867**
B	0.462	1.000	0.462	0.222	0.421	0.861	**0.363**	**0.778**
C	0.222	0.833	0.185	0.583	0.960	0.979	**0.940**	**1.000**
D	0.105	1.000	0.105	0.500	0.400	0.778	**0.311**	**0.833**
E	1.000	1.000	**1.000**	0.200	0.857	0.983	0.843	**0.782**
F	0.000	0.000	0.000	1.000	0.476	0.920	**0.438**	**0.625**
G	0.000	0.000	0.000	0.533	0.750	0.939	**0.704**	**0.933**
H	0.240	0.844	0.203	0.550	0.387	0.884	**0.342**	**0.850**
I	0.835	0.981	0.820	0.677	0.951	0.996	**0.947**	**0.889**
J	1.000	0.984	**0.984**	**0.774**	0.735	0.973	0.716	0.935
Average	0.469	0.763	**0.458**	**0.564**	0.676	0.928	**0.640**	**0.849**

Quality (SQ), and Recognition Quality (RQ) discussed by Van Heusden et al. [17] were used, as they provide measures of both the quantification and the precision of document segmentation.

Table 2 shows the results of the OCR-based and VLM-based pipeline. The comparison between the OCR and VLM pipeline presents a clear picture. While the page-level segmentation is of secondary interest in PSS tasks, the document segmentation shows a clear improvement through the VLM-based pipeline, with a PQ score of 0.640 compared to 0.458. While the OCR-based pipeline scored very well on some dossiers that had a clear stamp structure that it was instructed to detect, it performed poorly on documents that had varying document stamps or no stamps at all. Secondly, document type identification was performed much better by the VLM-based pipeline with an accuracy of 85% of documents correctly identified, compared to 56% in the OCR pipeline.

Comparing these scores to other benchmarks is difficult, as the research design differs substantially from our intended use case. In particular, studies such as Van Heusden et al. [17] and Wiedemann and Heyer [19] train and test models with a single document stream, which limits generalizability across domains. The drop in performance observed in Van Heusden et al. [17]'s study when the models were tested on a new stream exemplifies the heterogeneity in practice of FOIA requests and signals a need for a more general approach that is not reliant on learned features.

4.2 Usability Evaluation

To evaluate whether WOOLens makes Woo dossiers meaningfully more navigable in practice, a user study was conducted using a mixed-methods approach. Quantitative usability was measured through the System Usability Scale (SUS), a ten-statement scale originally developed by John [11] and later validated with score benchmarks by Bangor et al. [1], allowing results to be interpreted against an established standard. This was complemented by a think-aloud protocol, a method introduced by Ericsson and Simon [6] in which participants verbalise their reasoning during tasks, providing qualitative insight into where the interface supports or confuses users. Participants were recruited through random and anonymous

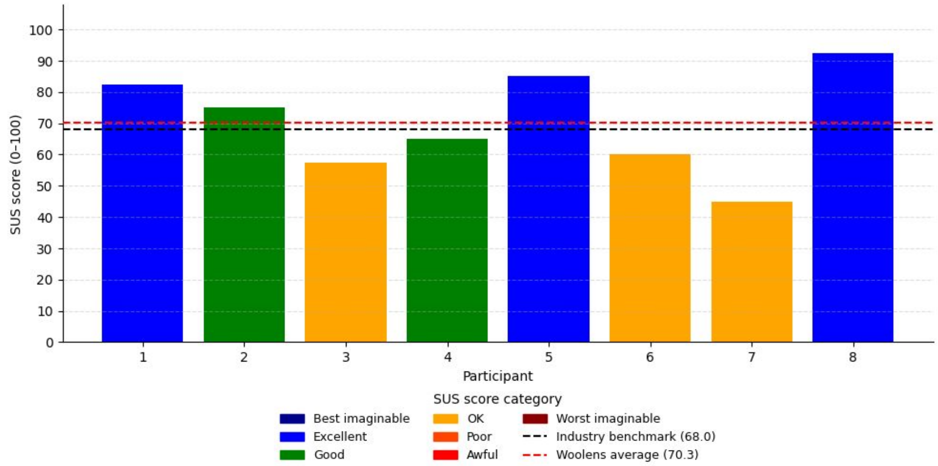

Figure 6: SUS scores per participant with industry benchmark.

selection and represented a deliberate range of prior familiarity with Woo requests, from no knowledge to some experience with official documents. Each session consisted of six tasks of increasing complexity, from first impressions and locating specific communications to timeline interpretation, document identification, and redaction analysis. By combining standardised measurement with behavioural observation, the evaluation aims to capture both overall usability and the specific points where added structure helps users make sense of a dossier they would otherwise face as an undifferentiated PDF.

Eight participants completed the usability evaluation. While eight participants is a limited sample for generalisation, it is consistent with established practice in exploratory usability studies, where five to eight users are typically sufficient to identify the majority of usability issues. The SUS scores ranged from 45.0 to 92.5 ($M = 70.3$, $SD = 16.1$), placing the overall group mean at the lower boundary of the "Good" adjective rating according to Bangor et al. [1]. Individual scores varied considerably: three participants reached the "Excellent" range (P-01: 82.5; P-05: 85.0; P-08: 92.5), two fell within the "Good" range (P-02: 75.0; P-04: 65.0), and three scored "OK" (P-03: 57.5; P-06: 60.0; P-07: 45.0). In Figure 6, SUS scores were interpreted using the adjective rating scale proposed by Bangor et al. [1, Fig. 11], in which mean scores correspond to qualitative labels ranging from 'Worst Imaginable' to 'Best Imaginable', with an average industry benchmark of 68.0 [13]; the WOOLens tool scores slightly above with $M = 70.3$.

Task performance varied by task type. Task 3, which asked participants to locate internal email correspondence related to the Dutch government's position on UNGA resolutions concerning the Israel-Palestine conflict [3], was completed successfully by six of eight participants, suggesting that the email interface was broadly navigable. Task 5, which involved locating a specific position document, was completed by four participants, with two completing it only partially due to a broken document preview, indicating a technical bug. Across all task-related measures, participants achieved a mean task performance score of $M = 0.84$ out of 1.0, indicating generally high task success despite the variation observed across individual task types. Figure 7 presents each participant's usability profile as a radar chart, showing the average of their SUS score, task performance, and prior Woo familiarity.

Taken together, the evaluation results suggest that WOOLens holds clear practical value for users engaging with Woo dossiers, as reflected in a high recommendation intent across the group ($M = 9.1/10$). This finding is particularly notable, given that no comparable tool currently exists for navigating Dutch open government disclosures, underscoring a genuine unmet need in the domain. At the same time, the evaluation identified specific areas where usability can be meaningfully improved. Think-aloud observations revealed recurring friction points, most notably the lack of discoverability for the redaction hover interaction, the unclear left sidebar statistics panel, and the absence of onboarding guidance. Participants consistently indicated that a short introductory video or structured onboarding flow would substantially lower the barrier to entry, particularly for users with limited prior familiarity with Woo dossiers. The timeline view and AI-generated summary were consistently praised as the most valuable features, confirming that the core design decisions around structured navigation and contextualisation are well-founded. Resolving the identified technical issues, including the broken document preview that affected Task 5 completion, and implementing targeted interface improvements based on participant feedback would be expected to raise the mean SUS score meaningfully in subsequent iterations, moving the tool more firmly into the "Good" to "Excellent" range and strengthening its practical utility for journalists, researchers, and engaged citizens alike.

5 Discussion and Future Work

The most significant limitation of the current system is the cost-effectiveness of a vision-based model for scanned documents on which this artefact relies. This makes the pipeline computationally expensive and, at this stage, constrains its practical use for processing large volumes of dossiers in batch, especially for a general audience. WOOLens works well as a case-based tool, but it is

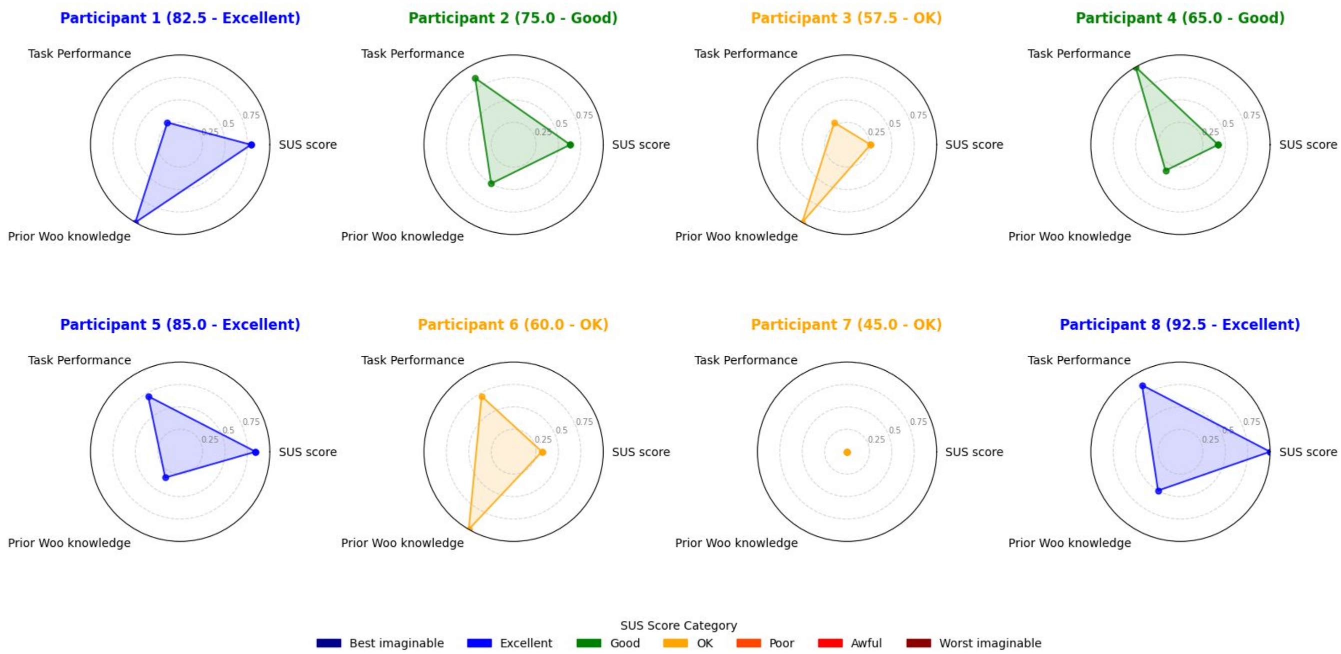

Figure 7: Radar Charts: Participant Usability Profiles.

not yet a scalable infrastructure. However, earlier research on PSS models shows possibilities for fine-tuning the model with annotated data, which may increase the quality of PSS further and could partially justify the increased cost of such a model. Furthermore, the technical evaluation reflects the time constraints given to this project and introduces limitations to the generalizability of our results. Testing our model on more common benchmarks, such as OpenPSS or Tobacco800, could provide beneficial insights into the quality of our pipeline. Furthermore, techniques that were used in previous research, such as fine-tuning the model, active learning, or combining models, were not in the scope of this research but may substantially improve results.

Another limitation lies in the raw data itself. The quality of any automated pipeline is constrained by the quality of the documents it processes. Heavily redacted pages, inconsistent formatting, and poor scan resolution all reduce what the system can extract, regardless of the model used. This points to a broader issue: the structural problems in Woo disclosures are not only a matter of how documents are presented to the public, but also of how they are produced and managed internally. Better upstream data practices would make tools such as WOOLens substantially more effective. This point about the importance of upstream data quality is also supported by Gao and Janssen [7], who stress that AI applications remain highly dependent on data quality, including accuracy, completeness, availability, and timeliness. Although some AI systems can still produce useful outputs from low-quality datasets, weak input data increases the likelihood of error and limits the reliability of the results. In other words, raw data quality remains a central condition for meaningful AI-supported interpretation. This is particularly relevant in the context of Woo dossiers, where the material is often incomplete, visually degraded, or difficult to process consistently. At the same time, the need for better upstream practices does not make downstream interpretability tools redundant. Many Woo dossiers have already been released as opaque, heavily redacted document collections, and individuals still need ways to work with this material. WOOLens should therefore be understood not as a substitute for better document management or publication standards, but as an intervention in the current state of disclosure. In a broader sense, one of the project's central ambitions is to highlight the importance of raw data quality in government document disclosure. A logical follow-up study would therefore be to develop more concrete guidelines for how government institutions prepare, structure, and publish documents so that they are more usable for both human readers and downstream computational processing.

6 Conclusion

This paper addresses the gap between formal transparency and practical usability in Dutch Woo disclosures. By developing the WOOLens system, we demonstrated how automated processing can convert unstructured dossiers into a more interpretable, usable format. Through document segmentation, type classification, date extraction, and timeline visualisation, WOOLens helps users make better sense of the information made public. In this way, the system supports a form of transparency that is not only formal but also usable in practice.

The usability evaluation showed that the interface has clear practical value. Participants achieved a mean SUS score of 70.3, placing WOOLens at the lower boundary of the "Good" adjective rating, and a mean task performance score of 0.84 out of 1.0, indicating that users were generally able to navigate the system successfully across different task types. At the same time, the evaluation also identified clear areas for improvement, including the discoverability of redaction information, the clarity of the statistics panel, and the absence of onboarding guidance.

These findings provide clear directions for the next iteration of the tool. Taken together, these findings reinforce the argument that transparency cannot be reduced to the release of information alone. It also depends on whether that information can be accessed, interpreted, and used. Designing for transparency, rather than treating it as an afterthought, is therefore essential if legal disclosure is to support practical accountability.

Acknowledgments

This research was conducted as part of an embedded research project for the Master's in Cultural Data and Artificial Intelligence program at the University of Amsterdam. The authors are supported by ICAI (AI for Open Government Lab).

Jaap Kamps is also supported by the Netherlands Organization for Scientific Research (NWO NWA # 1518.22.105), and the University of Amsterdam (AI4FinTech program). Views expressed in this paper are not necessarily shared or endorsed by those funding the research.

Disclosure of Interests The authors have no competing interests to declare that are relevant to the content of this article.

Disclosure of Generative AI use During the preparation of this work, the authors used *Grammarly* and *Claude Code* in order to: **Grammar and spelling check** and **Paraphrase and reword**, and for **Coding Assistance**.

After using these tools/services, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] Aaron Bangor, Philip T. Kortum, and James T. Miller. 2008. An empirical evaluation of the system usability scale. *International Journal of Human–Computer Interaction*, 24, 6, (July 29, 2008), 574–594. doi:10.1080/10447310802205776.

[2] Monika Bauhr and Marcia Grimes. 2017. Transparency to curb corruption? concepts, measures and empirical merit. *Crime, Law and Social Change*, 68, 4, (Nov. 2017), 431–458. doi:10.1007/s10611-017-9695-1.

[3] [n. d.] Besluit op woo-verzoek totstandkoming van nederlandse positiebepaling bij AVVN-resoluties over conflict tussen israël en palestijnse gebieden – ministerie van buitenlandse zaken (2024). WooPID — wooverheid. Retrieved Apr. 15, 2026 from https://pid.wooverheid.nl/?pid=nl.mnre1013.2i.2024.65.

[4] 2026. Claude code by anthropic | AI coding agent, terminal, IDE. Claude. Retrieved Apr. 10, 2026 from https://claude.com/product/claude-code.

[5] Maria Cucciniello and Greta Nasi. 2014. Transparency for trust in government: how effective is formal transparency? *International Journal of Public Administration*, 37, 13, (Nov. 10, 2014), 911–921. doi:10.1080/01900692.2014.949754.

[6] Karl Anders Ericsson and Herbert Alexander Simon. 1993. *Protocol analysis: verbal reports as data*. (rev. ed., 4. pr ed.). *A Bradford book*. The MIT Press, Cambridge, Mass. 443 pp. ISBN: 978-0-262-05047-0 978-0-262-55023-9.

[7] Yingying Gao and Marijn Janssen. 2020. Generating value from government data using AI: an exploratory study. In *Electronic Government*. Gabriela Viale Pereira, Marijn Janssen, Habin Lee, Ida Lindgren, Manuel Pedro Rodríguez Bolívar, Hans Jochen Scholl, and Anneke Zuiderwijk, (Eds.) Springer International Publishing, Cham, 319–331. ISBN: 978-3-030-57599-1. doi:10.1007/978-3-030-57599-1_24.

[8] Hunter Heidenreich, Ratish Dalvi, Rohith Mukku, Nikhil Verma, and Neven Pičuljan. 2024. Large language models for page stream segmentation. Version Number: 1. (2024). doi:10.48550/ARXIV.2408.11981.

[9] Alan R. Hevner, Salvatore T. March, Jinsoo Park, and Sudha Ram. 2004. Design science in information systems research. *MIS Q.*, 28, 1, (Mar. 1, 2004), 75–105.

[10] Marijn Janssen, Ricardo Matheus, Justin Longo, and Vishanth Weerakkody. 2017. Transparency-by-design as a foundation for open government. *Transforming Government: People, Process and Policy*, 11, 1, (Mar. 20, 2017), 2–8. doi:10.1108/TG-02-2017-0015.

[11] Brooke John. 1996. SUS: a 'quick and dirty' usability scale. In *Usability Evaluation In Industry*. (0th ed.). Patrick W. Jordan, B. Thomas, Ian Lyall McClelland, and Bernard Weerdmeester, (Eds.) CRC Press, (June 11, 1996), 207–212. ISBN: 978-0-429-15701-1. doi:10.1201/9781498710411-35.

[12] Ministerie van Binnenlandse Zaken en Koninkrijksrelaties. 2024. Rijksbrede instructie voor het behandelen van woo-verzoeken - publicatie - rijksoverheid.nl. Last Modified: 2024-07-16T11:24. (July 16, 2024). Retrieved Apr. 10, 2026 from https://www.rijksoverheid.nl/documenten/publicaties/2022/04/01/rijksbrede-instructie-voor-het-behandelen-van-woo-verzoeken.

[13] James R. Lewis. 2018. The system usability scale: past, present, and future. *International Journal of Human–Computer Interaction*, 34, 7, (July 3, 2018), 577–590. doi:10.1080/10447318.2018.1455307.

[14] Ken Peffers, Tuure Tuunanen, Marcus A. Rothenberger, and Samir Chatterjee. 2007. A design science research methodology for information systems research. *Journal of Management Information Systems*, 24, 3, (Dec. 2007), 45–77. doi:10.2753/MIS0742-1222240302.

[15] Helene Friis Ratner and Nanna Bonde Thylstrup. Ecologies of evaluation: sites and registers of machine learning evaluation in a public sector context. Public Lecture, University of Amsterdam, (Mar. 18, 2026).

[16] Johanne R. Trippas et al. 2025. Report from the 4th strategic workshop on information retrieval in lorne (SWIRL 2025). *SIGIR Forum*, 59, 1, 1–68. doi:10.1145/3769733.3769739.

[17] Ruben Van Heusden, Jaap Kamps, and Maarten Marx. 2024. OpenPSS: an open page stream segmentation benchmark. In *Linking Theory and Practice of Digital Libraries*. Vol. 15177. Apostolos Antonacopoulos, Annika Hinze, Benjamin Piwowarski, Mickaël Coustaty, Giorgio Maria Di Nunzio, Francesco Gelati, and Nicholas Vanderschantz, (Eds.) Series Title: Lecture Notes in Computer Science. Springer Nature Switzerland, Cham, 413–429. ISBN: 978-3-031-72436-7 978-3-031-72437-4. doi:10.1007/978-3-031-72437-4_24.

[18] Ruben Van Heusden, Maik Larooij, Jaap Kamps, and Maarten Marx. 2025. A collection of FAIR dutch freedom of information act documents. *Scientific Data*, 12, 1, (May 15, 2025), 795. doi:10.1038/s41597-025-05052-2.

[19] Gregor Wiedemann and Gerhard Heyer. 2021. Multi-modal page stream segmentation with convolutional neural networks. *Language Resources and Evaluation*, 55, 1, (Mar. 1, 2021), 127–150. doi:10.1007/s10579-019-09476-2.

[20] [n. d.] WooZM — de woo ZoekMachine. Retrieved Apr. 7, 2026 from https://woozm.nl/.

[21] Ministerie van Algemene Zaken. 2026. Wet open overheid (woo) - rijksoverheid.nl. Last Modified: 2025-05-08T10:58 Running Time: 617. (Apr. 28, 2026). Retrieved Apr. 6, 2026 from https://www.rijksoverheid.nl/onderwerpen/wet-open-overheid-woo.

A Appendix

The full code of the WOOLens system is available at:

- https://github.com/joriswex/WOO_organizer.