The Readability Paradox: Auditing the Twin Opacities of the Indian Union Budget

[1] Megha Agarwal, [2] Dr. Mamta Brahmbhatt

[1] Research Scholar, B.K. School of Professional and Management Studies, Gujarat University, India.
[2] Professor & Head (Department of Business Intelligence), B.K. School of Professional and Management Studies, Gujarat University, India.

[1] ORCID - https://orcid.org/0009-0000-5344-6354

Abstract:
The Indian Union Budget is a foundational document of national governance, yet its dense language has long been a barrier to public accountability. This study audits Union Budget documents from 2019 to 2025, interrogating both their linguistic properties and their structural integrity as digital artifacts. Our analysis uncovers a critical "readability paradox": summarized 'Highlights' documents yield statistically higher, and thus more complex, readability scores (avg. Flesch-Kincaid Grade: 19.24) than the unabridged 'Speeches' (12.75). We demonstrate that this anomaly is not a feature of the prose but a methodological artifact stemming from inconsistent digital formatting in official PDF releases, which corrupts automated text parsing. For the structurally sound documents, we find that syntactic complexity—specifically, average sentence length—is the dominant driver of difficulty, far outweighing the impact of specialized jargon. This paper posits that the challenge to fiscal transparency is twofold, introducing the "Twin Opacities" framework to distinguish between the linguistic barriers to human comprehension and the structural barriers to machine processability. We conclude that for transparency to be meaningful, a commitment to plain language must be accompanied by a rigorous standard for data quality and machine-readability, a principle we term "Transparency by Design."

Keywords
Fiscal Transparency, Computational Linguistics, Digital Governance, Readability, Structural Opacity, Machine-Readability, India

Highlights

- Uncovers a "readability paradox" where official summaries of the Indian Budget appear statistically more complex than their source documents.
- Demonstrates this paradox is a methodological artifact caused by Structural Opacity—inconsistent PDF formatting that breaks automated analysis.
- Proposes the "Twin Opacities" framework to distinguish barriers to human comprehension (Linguistic) from barriers to machine processing (Structural).
- Concludes that meaningful public accountability requires "Transparency by Design," mandating both plain language and machine-readable data formats.

1. Introduction: The Promise and Peril of Computational Accountability

The Union Budget of India is not merely a fiscal ledger. It is the nation's primary economic speech act, a performative instrument of governance that articulates state priorities and shapes the contours of public life for the year to come. In a democracy of over a billion people, the ability of citizens and civil society to engage with this document is a fundamental precondition for accountability. The budget's transparency, therefore, is not a procedural matter but an epistemic challenge central to the health of democratic deliberation.

For decades, the critique of the budget has centered on a single, familiar villain: linguistic opacity. The document's dense amalgam of specialized lexicon and convoluted syntax has been rightly condemned as a formidable barrier to comprehension. This inaccessibility stifles public discourse and fosters a democratic deficit, leaving a key instrument of governance opaque to the very public it serves. This critique, while true, is now dangerously incomplete.

This paper argues that a second, more insidious barrier has emerged: structural opacity. In an era of computational scrutiny, the very format of public documents—their digital encoding, layout, and adherence to machine-readable standards—can either facilitate or obstruct accountability at scale. When documents are released in inconsistent or poorly structured formats, they become opaque not just to the human eye, but to the automated tools that promise to enhance transparency. The challenge is no longer confined to the prose; it now extends to the digital architecture of the documents themselves.

To diagnose this dual challenge, we introduce and test the "Twin Opacities" framework. This model provides a necessary lens for transparency in the modern era, distinguishing between the traditional barrier to human comprehension (Linguistic Opacity) and the modern barrier to machine processability (Structural Opacity). It posits that a failure in either dimension creates a critical democratic deficit and, more importantly, that structural failures can create the illusion of linguistic phenomena, leading to dangerously flawed conclusions. Using the Indian Union Budget documents from 2019 to 2025 as a crucible, this paper conducts a computational audit that uncovers a "readability paradox"—an analytical anomaly that allows us to diagnose a fundamental failure in the digital transparency pipeline. Our contribution is not merely to quantify the budget's complexity, but to offer a critical methodological lesson for evaluating the integrity of public data in an age of automated accountability.

2. Review of Literature

This research is situated at the confluence of three distinct but converging streams of scholarship. To build the foundation for our central thesis, we first ground our inquiry in the public administration literature on fiscal transparency, then review the long-standing tradition of using readability metrics to quantify linguistic complexity, and finally, situate our work within the modern computational paradigm, exposing a critical gap where the promise of "text-as-data" collides with the messy reality of public information.

2.1 The Democratic Imperative of Fiscal Transparency

The principle that public budgets should be open to scrutiny is a cornerstone of modern democratic theory (Roberts, 2006). Scholars have long contended that fiscal transparency is not merely a matter of good governance but a prerequisite for meaningful citizen participation and accountability (Stiglitz, 2001). When governments provide clear and accessible information, it empowers civil society, reduces opportunities for corruption, and fosters public trust (Bellver & Kaufmann, 2005; Meijer, 2014). In developing nations, this transparency has been directly linked to improved fiscal discipline (Glennerster & Shin, 2008). Conversely, opacity creates profound information asymmetries that concentrate power within a technocratic elite, producing what Fung, Graham, and Weil (2007) have

termed a "democratic deficit." While the shift to e-governance promised to mitigate this, the mere availability of data is no guarantee of transparency (Heald, 2006); without a commitment to clarity, digital archives can become overwhelming "data dumps" rather than tools of empowerment (Yu & Robinson, 2012). Our study proceeds from this normative foundation, treating the accessibility of the Union Budget as a direct measure of democratic health.

2.2 The Tradition of Readability Analysis in Public Documents

Building on this normative foundation, a significant body of empirical research has sought to measure the linguistic challenges of translating transparency into practice. This tradition, rooted in foundational work to quantify the cognitive burden of a text (Flesch, 1948), was soon adapted for official and technical contexts (Kincaid et al., 1975). Subsequent scholarship has applied these methods to a vast range of official documents, from central bank communications to corporate financial disclosures, consistently finding them written at a level far exceeding the comprehension of an average citizen (Loughran & McDonald, 2014; Li, 2008). Research has revealed a profound disconnect between the language of the law and the public's ability to understand it—a problem famously known as "legalese" (Tiersma, 1999)—while studies in public health have shown that critical information is often rendered inaccessible by its graduate-level complexity (Rudd, 2007). This scholarly tradition provides a robust, validated toolkit for quantifying what our framework identifies as Linguistic Opacity. Our work builds directly on this precedent, applying these established metrics to the Indian fiscal context.

2.3 The Computational Turn and its Structural Perils

The proliferation of digital documents and Natural Language Processing (NLP) has revolutionized the study of political texts. The "text-as-data" paradigm, as outlined by Grimmer and Stewart (2013), allows researchers to analyze vast corpora with unprecedented scale and objectivity. This computational turn holds immense promise for accountability, enabling what Diakopoulos (2015) calls "algorithmic accountability reporting" by journalists and civic-tech organizations.

This methodological wave, however, is not without its perils. The very scholars who championed "big data" have also issued stark warnings about the traps inherent in using "found" or administrative data without rigorous validation (Lazer et al., 2014; boyd & Crawford, 2012). The digital structure and format of documents can introduce profound biases and errors into any analysis, creating what O'Neil (2016) terms "weapons of math destruction." A PDF, for instance, is not a standardized data container; its internal structure is optimized for visual layout, not semantic meaning, a fact that leads to critical failures in automated text extraction (Baumann, 2014; Kaur & Singh, 2017). This problem of "dirty data" requires extensive cleaning, a process that is frequently under-documented in research (Gil & Staveloz, 2017).

This emerging literature on data quality provides the crucial context for our paper's primary contribution. It highlights a glaring gap in the existing transparency discourse. Within the fields of public administration and policy analysis, the challenge of machine-readability has largely been treated as a mere technical nuisance—a pre-processing headache for the researcher, as noted in software engineering contexts (Jedlitschka & Pfahl, 2005). It has not been theorized as a distinct and potent barrier to public accountability in its own right. Our concept of Structural Opacity aims to fill this gap, providing the necessary theoretical lens to understand this critical new frontier of computational governance.

3. The Baseline: Quantifying the Linguistic Opacity of the Budget Speech

Before investigating the more complex issue of structural inconsistencies, we first establish an empirical baseline. This requires isolating the most consistently structured documents in our corpus—the verbatim 'Budget Speeches'—to conduct a focused audit of their linguistic

properties. This analysis serves two purposes: first, to move beyond anecdotal claims by quantifying the precise level of the speeches' linguistic complexity, and second, to deconstruct this complexity to identify its primary drivers.

3.1 Methodology for Baseline Analysis

Our baseline analysis was confined to the seven 'Budget Speech' documents from 2019 to 2025. These files were consistently available in well-structured DOCX format, a choice made to ensure clean and reliable text extraction, thereby minimizing the risk of the parsing errors that can plague other formats. We applied a suite of linguistic metrics to this clean corpus. Readability was measured using the Flesch-Kincaid Grade Level. To identify the sources of complexity, we engineered two key features: Jargon Density, calculated as the frequency of 31 curated fiscal terms per 1,000 words, and Average Sentence Length, a robust proxy for syntactic complexity derived using a standard sentence segmentation model. Finally, a Pearson correlation analysis was conducted to determine the statistical relationship between these features and the final readability score.

3.2 Findings: Confirmation of High Complexity

The analysis confirms that the primary budget narrative is characterized by a high degree of linguistic opacity. The average Flesch-Kincaid Grade Level for the Budget Speeches across the seven-year period was 12.75. This score indicates that a university-level education is required for comfortable comprehension, a standard that places the nation's principal fiscal statement beyond the direct reach of a vast majority of its citizens. This finding provides a quantitative anchor for the long-standing critique of the budget's inaccessibility.

TABLE 1 - Descriptive Statistics of Linguistic Features for Budget Speeches (2019-2025)

Metric	Mean	Standard Deviation	Minimum	Maximum
Flesch-Kincaid Grade	12.75	0.69	11.88	13.57
Avg. Sentence Length	20.33	2.21	16.99	22.73
Jargon Density	9.10	2.05	5.87	11.10

3.3 Deconstructing Complexity: Syntax over Semantics

While the high complexity is now quantified, a more insightful finding lies in its composition. The common narrative often attributes the difficulty of economic texts to specialized jargon. Our correlation analysis, however, challenges this simplistic assumption.

As illustrated in Figure 1, the primary driver of high readability scores is Average Sentence Length, which exhibits a strong and significant positive correlation with the Flesch-Kincaid Grade ($r = 0.73$). This provides evidence that syntactic complexity—the use of long, multi-clause sentences—is the principal contributor to the budget's poor readability. In contrast, the correlation with Jargon Density is exceptionally weak ($r = 0.16$). While technical terms are present, their frequency alone does not appear to drive the difficulty. The core challenge for the reader lies in parsing the document's syntax, not just its vocabulary.

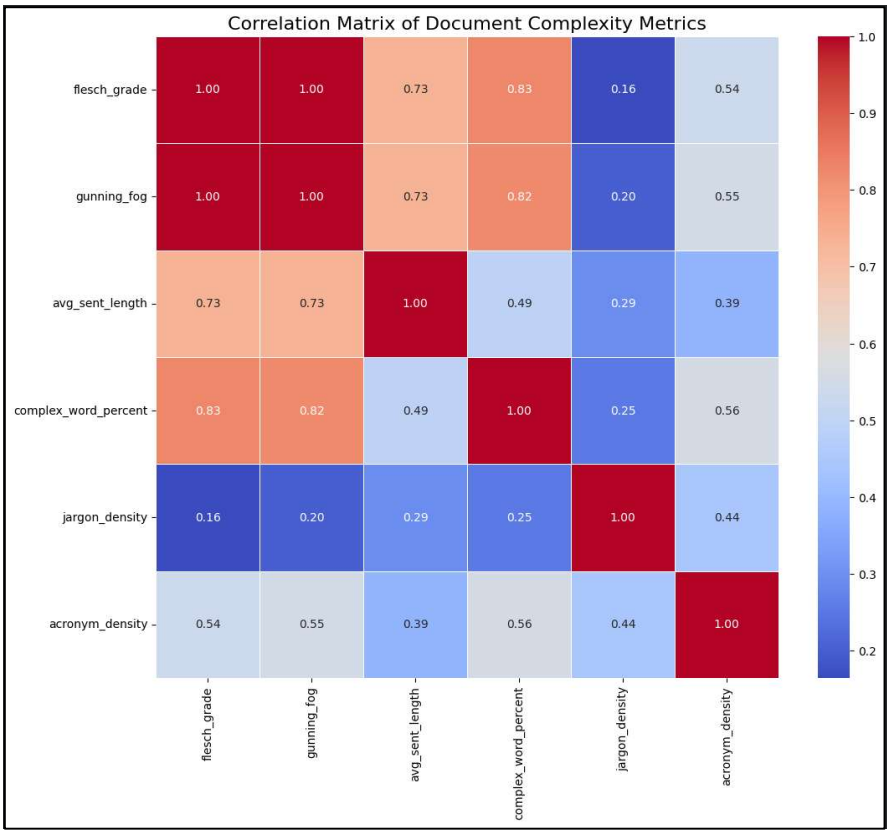

FIGURE 1 - Correlation Matrix of Complexity Drivers in Budget Speeches

3.4 Thematic Priorities and Rhetorical Stance

Beyond linguistic structure, our baseline analysis of the 'Budget Speeches' also examined their thematic content and rhetorical framing. A zero-shot classification model was used to identify the prominence of ten key policy areas, while a sentiment analysis model gauged the overall tone. The analysis reveals a consistent discursive focus on foundational sectors such as 'Social Welfare & Inclusion' and 'Infrastructure'. A discernible evolution in priorities is also evident over the seven-year period, with a notable increase in emphasis on 'Technology & Digital India' in more recent years.

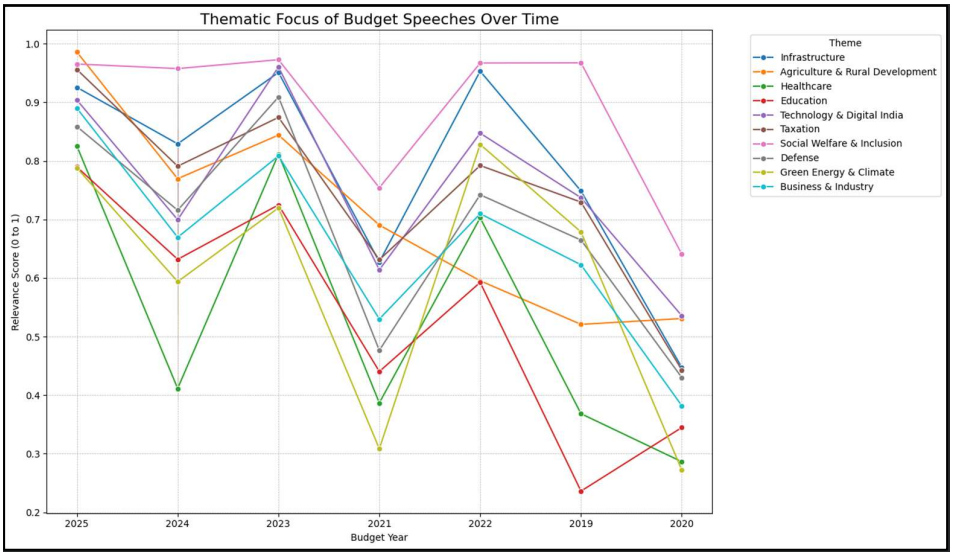

FIGURE 2 - Thematic Focus of Budget Speeches Over Time

Across all years, the speeches were consistently classified with a 'Positive' sentiment, an expected rhetorical feature of a formal government address designed to project confidence and outline a forward-looking economic vision.

TABLE 2 - Average Thematic Focus Across All Speeches (2019-2025)

Theme	**Average Relevance Score**
Social Welfare & Inclusion	0.897874
Infrastructure	0.788911
Taxation	0.751007
Technology & Digital India	0.749684
Agriculture & Rural Development	0.713318
Defense	0.689119
Business & Industry	0.660237
Green Energy & Climate	0.597877
Education	0.549026
Healthcare	0.525516

4. The Anomaly: Uncovering a Paradox in Readability Metrics

Having established the linguistic baseline of the unabridged Budget Speeches, the analysis logically proceeds to their official summaries, the 'Budget Highlights' documents. These documents are created for one clear purpose: to distill the main text into a more accessible format for wider dissemination. The logical expectation, therefore, is simple. They should be significantly easier to read.

Our empirical analysis reveals a stark and paradoxical contradiction to this expectation. The data refused to comply.

When the same readability metrics were applied to the 'Highlights' documents, the results were inverted. The average Flesch-Kincaid Grade Level for the 'Highlights' soared to 19.24—a figure that is not only substantially higher than the 12.75 average for the full speeches but is also nonsensical for any form of human-generated text. It suggests a level of complexity far beyond postgraduate academic writing, a finding that defies all reasonable explanation.

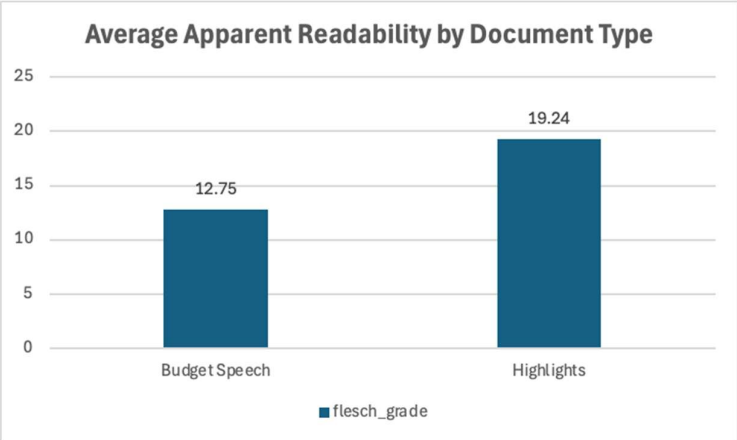

FIGURE 3 - Average Apparent Readability by Document Type

This counter-intuitive inversion, visualized in Figure 3, is not a statistical fluke. An independent sample t-test confirms the difference is statistically significant ($p = 0.0491$), falling below the conventional $p < 0.05$ threshold for significance. The data thus present a clear and unavoidable paradox: the very documents designed to simplify the budget appear, by standard automated metrics, to be orders of magnitude more complex than the source they are meant to summarize.

What can explain this analytical absurdity? It cannot be plausibly explained as a feature of the prose itself. It is inconceivable that government communicators, in an attempt to create a summary, would produce text that is dramatically more convoluted than the original. The finding therefore points away from a linguistic explanation and towards a methodological or structural one. The anomaly demands a forensic investigation, shifting the focus of our inquiry from the text's content to its underlying digital architecture.

5. The Diagnosis: Exposing the Roots of Structural Opacity

The paradox uncovered in the preceding section—where summary documents yield impossibly high complexity scores—necessitates a shift in analytical focus from the linguistic to the structural. These implausible scores are not a reflection of the text's content but are symptoms of a deeper issue. This section presents a forensic diagnosis of these symptoms to reveal the underlying cause: a failure of machine-readability rooted in inconsistent document formatting.

5.1. A Forensic Analysis of Outlier Documents

The investigation begins by isolating the specific documents responsible for skewing the aggregate results. A granular inspection of the data reveals that the anomalous scores are driven by extreme outliers, particularly the 'Highlights' documents for 2021, 2022, and both versions released for the year 2024, which produced Flesch-Kincaid scores of 33.5, 27.1, and 23.4 respectively.

To understand the source of these scores, we must examine their constituent parts. The Flesch-Kincaid formula is a function of two variables: average syllables per word and average words per sentence. While the former remained within a normal range across all documents, the latter revealed the true source of the anomaly. The 'Highlights' document for 2021, for instance, produced an average sentence length of 73.2 words. Such a figure is a clear statistical artifact; it does not represent human language but rather a catastrophic failure in the sentence segmentation process.

TABLE 3 - Diagnostic Metrics for Outlier 'Highlights' Documents

Year	Flesch-Kincaid Grade	Average Sentence Length (Words)
2021	33.51	73.27
2022	27.12	38.00
2024	23.42	71.70
2024	17.78	33.43

5.2. The Culprit: PDF Formatting and Parsing Failures

This diagnostic trail leads directly to the source files themselves. The outlier documents were all distributed as PDF files, a format notorious for its lack of a consistent, underlying text structure. A manual inspection of the raw text extracted from these files confirms the diagnosis: the automated PDF-to-text conversion process failed to correctly recognize line and paragraph breaks. Bullet points, short phrases, and even entire sections of text separated by visual whitespace were concatenated by the parser into single, massive character strings. The algorithm, unable to find sentence-terminating punctuation within these strings, correctly—by its own logic—identified them as single, monstrously long "sentences."

5.3. The Finding: From Linguistic Phenomenon to Methodological Artifact

The readability paradox is therefore resolved. The statistically significant difference in "complexity" between the 'Speeches' and 'Highlights' is not a linguistic phenomenon but a methodological artifact. The complexity being measured was not in the prose but in the inconsistent digital structure of the document.

This finding provides the first empirical validation of our framework. The nonsensical readability scores of the 'Highlights' are a direct measurement of Structural Opacity. They reveal a critical failure of machine-readability, where the digital format of an official document is so inconsistent that it obstructs—and in this case, actively misleads—the very computational tools designed to promote transparency. The problem was not the language; it was the digital container in which the language was delivered.

6. The Framework: Theorizing the Twin Opacities

The diagnosis of the readability paradox provides the empirical foundation for a broader theoretical contribution. The case of the Indian Union Budget is not merely an isolated incident of poor data formatting; it is a powerful illustration of a fundamental challenge in modern governance. It reveals that transparency in the digital age is not a monolithic concept but a dual challenge. We therefore propose the "Twin Opacities" framework as a formal model for analyzing and auditing state communication.

6.1. Formal Definition: Linguistic Opacity

Linguistic Opacity is the traditional barrier to public understanding, rooted in the intrinsic properties of the text itself. It is a function of semantic and syntactic complexity.

- **Domain:** The prose.
- **Barrier to:** Human comprehension.
- **Method of Detection:** Standard linguistic and readability analyses (e.g., Flesch-Kincaid) applied to structurally sound, validated text.
- **In this study:** Empirically demonstrated by the university-level reading difficulty (FK Grade 12.75) of the clean 'Budget Speech' documents, driven primarily by high syntactic complexity.

6.2. Formal Definition: Structural Opacity

Structural Opacity is a distinctly digital-era barrier, rooted in the properties of the document as a data artifact. It is a function of the document's format, encoding, and adherence to standards that permit clean, automated parsing.

- **Domain:** The digital container (e.g., PDF, DOCX, HTML).
- **Barrier to:** Machine processability and, by extension, scalable public scrutiny.
- **Method of Detection:** Forensic analysis of data artifacts, statistical outliers, and paradoxical results generated during automated processing.
- **In this study:** Empirically demonstrated by the "readability paradox," where parsing failures in poorly structured PDFs generated misleadingly high complexity scores, proving a failure of machine-readability.

6.3. The Interplay: How the Structural Masquerades as the Linguistic

The most critical insight of this framework is the interplay between the two opacities. They are not independent phenomena. As our findings demonstrate, Structural Opacity can create the illusion of extreme Linguistic Opacity. An uncritical, automated analysis of the 'Highlights' documents would have led to the dangerously flawed conclusion that the summaries were written in an incomprehensible, post-doctoral language. The structural failure would have been misinterpreted as a linguistic one.

This has profound implications. It suggests that without a preliminary audit for structural integrity, any large-scale computational analysis of government documents is at high risk of producing nonsensical results. The framework thus proposes a clear analytical hierarchy: an assessment of Structural Opacity must precede any measurement of Linguistic Opacity. A document must first be proven to be machine-readable before its human-readability can be meaningfully assessed.

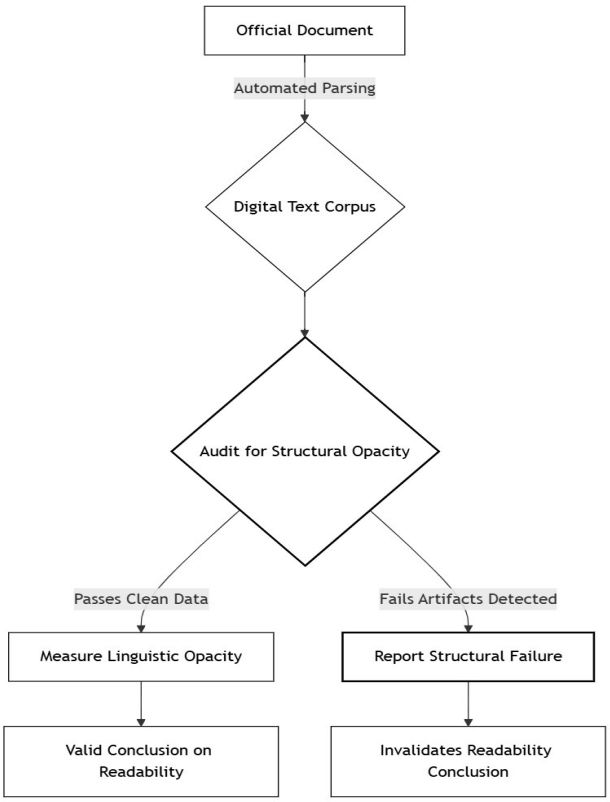

FIGURE 4 - The Twin Opacities Analytical Workflow

7. Conclusion: Towards Transparency by Design

This study began as a computational audit of the Indian Union Budget's complexity and, in the process, uncovered a more fundamental challenge to public accountability. Our investigation has empirically demonstrated that the barriers to fiscal transparency are twofold. We have proposed and validated the "Twin Opacities" framework, distinguishing between the traditional barrier of Linguistic Opacity—the convoluted prose that hinders human comprehension—and the modern, more insidious barrier of Structural Opacity—the inconsistent digital formatting that obstructs machine processability. The discovery of a "readability paradox," where summary documents produced nonsensical and misleadingly high complexity scores, served as a stark illustration of this structural failure. This finding is a critical warning: in an era reliant on automated scrutiny, inconsistent data is not a benign technical glitch but an active barrier to democratic accountability.

The implications of this framework are clear. For government bodies, the pursuit of transparency must evolve. It is no longer sufficient to simply publish documents online; they must be published with an ethic of "Transparency by Design." This requires a dual commitment: first, to linguistic clarity through the adoption of plain language principles, and second, to structural integrity through the enforcement of open, consistent, and truly machine-readable formats.

For researchers and practitioners in the computational social sciences, this study serves as a crucial methodological cautionary tale. It highlights the absolute necessity of a "defensive data analysis" paradigm—an approach that begins with a rigorous audit of a data source's structural integrity before any substantive conclusions are drawn. Trusting official digital documents at face value is a recipe for generating deeply flawed, even inverted, results.

Ultimately, the standards for public information must meet the capabilities of the age. A document that a citizen cannot understand is a failure of communication; a document that a machine cannot reliably parse is a failure of modern governance.

8. References

8.1 Academic Literature

Baumann, S. (2014). A Quantitative and Qualitative Analysis of the Characteristics of Public Sector PDFs. *Proceedings of the 14th European Conference on Digital Government*.

Bellver, A., & Kaufmann, D. (2005). *Transparenting transparency: Initial empirics and policy applications*. World Bank Policy Research Working Paper, (3765).

boyd, d., & Crawford, K. (2012). Critical Questions for Big Data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society, 15*(5), 662-679.

Diakopoulos, N. (2015). Algorithmic Accountability: Journalistic investigation of computational power structures. *Digital Journalism, 3*(3), 398-415.

Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology, 32*(3), 221–233.

Fung, A., Graham, M., & Weil, D. (2007). *Full disclosure: The perils and promise of transparency*. Cambridge University Press.

Gil, Y., & Staveloz, W. (2017). The Trouble with Data: A review of the challenges of data science. *Proceedings of the 2017 IEEE International Conference on Big Data*.

Glennerster, R., & Shin, Y. (2008). Does Transparency Pay? *IMF Staff Papers, 55*(1), 183-209.

Grimmer, J., & Stewart, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis, 21*(3), 267-297.

Heald, D. (2006). Transparency as an instrumental value. In *Transparency: The key to better governance?* (pp. 59-73). Oxford University Press.

Jedlitschka, A., & Pfahl, D. (2005). Reporting experiments in software engineering. In *Guide to advanced empirical software engineering* (pp. 201-228). Springer.

Kaur, G., & Singh, P. (2017). A review of text extraction techniques for PDF documents. *Journal of King Saud University-Computer and Information Sciences, 29*(3), 336-348.

Kincaid, J. P., Fishburne, R. P., Rogers, R. L., & Chissom, B. S. (1975). *Derivation of new readability formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy enlisted personnel*. Naval Technical Training Command Research Branch Report.

Lazer, D., Kennedy, R., King, G., & Vespignani, A. (2014). The parable of Google Flu: traps in big data analysis. *Science, 343*(6176), 1203-1205.

Li, F. (2008). Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and Economics, 45*(2-3), 221-247.

Loughran, T., & McDonald, B. (2014). Measuring readability in financial disclosures. *The Journal of Finance, 69*(4), 1643-1671.

Meijer, A. J. (2014). Transparency. In *The Oxford handbook of public accountability* (pp. 661-672). Oxford University Press.

O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown.

Roberts, A. (2006). *Blacked out: Government secrecy in the information age*. Cambridge University Press.

Rudd, R. E. (2007). Health literacy: An update of medical and public health literature. In *Review of adult learning and literacy* (Vol. 7, pp. 175-204).

Stiglitz, J. E. (2001). Transparency in Government. In *The Right to Tell: The Role of Mass Media in Economic Development*. World Bank Publications.

Tiersma, P. M. (1999). *Legal language*. University of Chicago Press.

Yu, H., & Robinson, D. G. (2012). The New Ambiguity of 'Open Government'. *UCLA Law Review Discourse, 59*, 178.

8.2 Technical & Software

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv preprint arXiv:1810.04805*.

Harris, C. R., Millman, K. J., van der Walt, S. J., et al. (2020). Array programming with NumPy. *Nature, 585*, 357–362.

Honnibal, M., & Montani, I. (2017). *spaCy 2: Natural language understanding with Bloom embeddings and convolutional neural networks*. (Software). https://spacy.io

Lewis, M., Liu, Y., Goyal, N., et al. (2019). BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension. *arXiv preprint arXiv:1910.13461*.

McKinney, W. (2010). Data Structures for Statistical Computing in Python. *Proceedings of the 9th Python in Science Conference*, 56-61. (for pandas)

Phaseit, Inc. (2022). *PyPDF2*. (Software). https://pypi.org/project/PyPDF2/

Python Software Foundation. (2023). *python-docx*. (Software). https://python-docx.readthedocs.io/

Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2019). DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. *arXiv preprint arXiv:1910.01108*.

Swatan, R. (2020). *textstat: a Python library to calculate readability*. (Software). https://github.com/textstat/textstat

Wolf, T., Debut, L., Sanh, V., et al. (2020). Transformers: State-of-the-Art Natural Language Processing. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, 38-45. (for Hugging Face Transformers)

Appendix A: Detailed Year-over-Year Metrics

The following charts visualize the year-over-year trends for the primary linguistic and structural metrics across all document types analyzed in this study. These figures are provided for comprehensive review and to illustrate the data underlying the aggregate statistics presented in the main text. Note the high volatility and extreme outlier values for the 'Highlights' document type in Figures A1 and A3, which correspond to the structural artifacts discussed in Section 5.

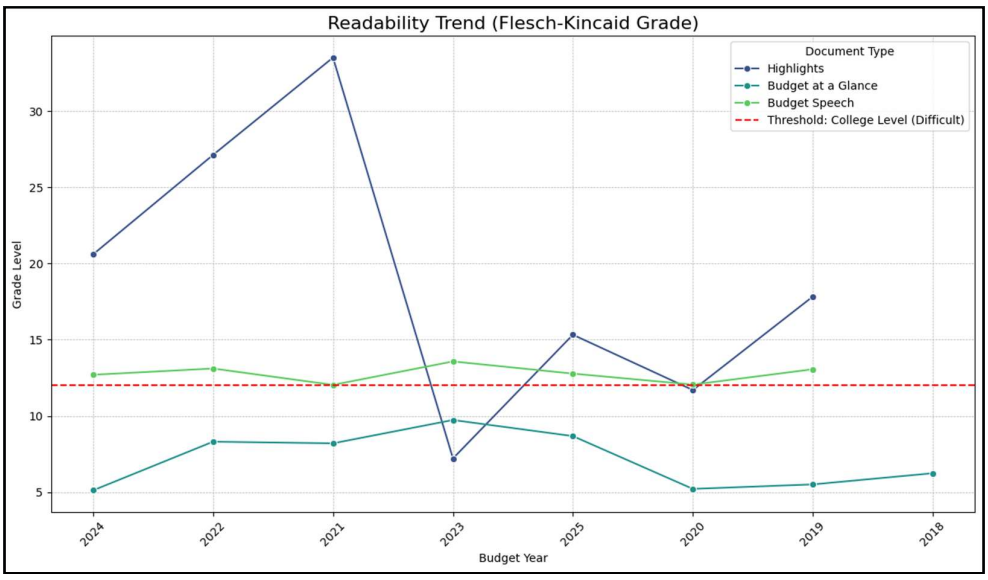

Figure A1: *Flesch-Kincaid Grade Level by Document Type (2019-2025)*

Figure A1 displays the readability score trend for the three main document types. Note the high volatility and extreme outlier values for the 'Highlights' series, which correspond to the structural artifacts discussed in Section 5.

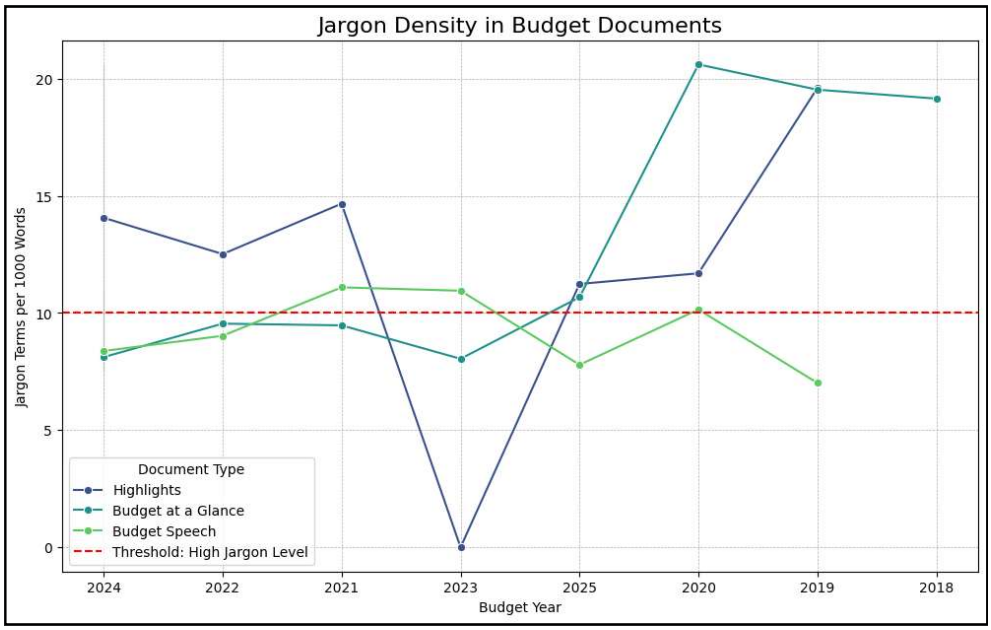

Figure A2: *Jargon Density by Document Type (2019-2025)*

Figure A2 displays the frequency of curated fiscal terms per 1,000 words for each document type. Note the relative stability of jargon across all document types, in contrast to the volatility of the structural metrics shown in Figure A3.

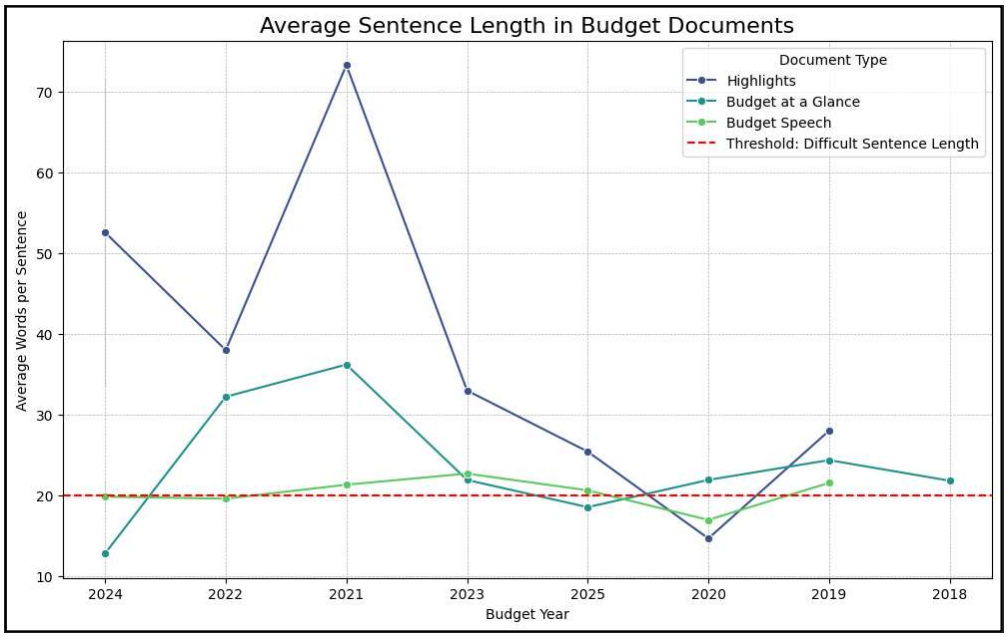

Figure A3: *Average Sentence Length by Document Type (2019-2025)*

Figure A3 displays the average number of words per sentence for each document type. The extreme outlier values for the 'Highlights' series directly correspond to the PDF parsing failures identified as the source of the readability paradox.

Appendix B: Complete Document-Level Data

For the purposes of transparency and replicability, the complete document-level data for all key metrics analyzed in this study are provided below.

Table B1 provides the complete, document-level readability data for both Flesch-Kincaid Grade Level and Gunning Fog index scores used in the analysis.

Table B1: *Complete Readability Scores by Document and Year*

Year	Budget Speech		Budget Highlights		Budget at a Glance	
	Flesch-Kincaid Grade	Gunning Fog	Flesch-Kincaid Grade	Gunning Fog	Flesch-Kincaid Grade	Gunning Fog
2019	13.06	15.90	17.83	21.03	5.51	9.06
2020	12.05	15.14	11.70	14.86	5.21	8.84
2021	12.04	14.99	33.51	36.90	8.20	11.86
2022	13.11	16.36	27.12	30.75	8.31	11.84
2023	13.57	16.81	7.21	8.40	9.73	13.39

2024	11.88, 13.51	14.97, 16.82	17.78, 23.42	21.29, 27.09	5.12	8.73
2025	12.77	15.75	15.33	18.32	8.68	12.31

Table B2 provides the complete, document-level data for Jargon Density (terms per 1,000 words) and Average Sentence Length (words), the two primary features used to deconstruct complexity.

Table B2: *Complete Jargon and Structural Metrics by Document and Year*

Year	**Budget Speech**		**Highlights**		**Budget at a Glance**	
	Jargon Density	**Avg. Sentence Length**	**Jargon Density**	**Avg. Sentence Length**	**Jargon Density**	**Avg. Sentence Length**
2019	7.01	21.57	19.61	28.02	19.53	24.39
2020	10.15	16.99	11.70	14.70	20.61	21.94
2021	11.10	21.35	14.67	73.27	9.47	36.21
2022	9.03	19.63	12.52	38.00	9.55	32.23
2023	10.95	22.73	0.00	33.00	8.05	21.94
2024	10.89, 5.87	17.96, 21.78	20.61, 7.52	71.70, 33.43	8.12	12.84
2025	7.79	20.65	11.24	25.45	10.67	18.57