Grenze International Journal of Engineering and Technology, June Issue

A Multi-Dimensional Framework for News Quality Assessment Integrating Polysemy, Sentiment, Bias Detection

Matti Nidhi Prabhu[1], Ananya Shrikanth Bhat[2], Manjunatha A S[3] and Ramakrishna M[4]
[1-3]Department of Computer and Communication Engineering, NITTE (Deemed to be University), Nitte, India
Email: mattinidhiprabhu@gmail.com, ananya2004bhat@gmail.com, manjunatha.as@nitte.edu
[4]School of Computer Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, India
Email: ramakrishna.m@manipal.edu

***Abstract*— In today's media landscape, the adoption of Natural Language Processing (NLP) methods is crucial for maintaining clarity and enhancing content quality. The paper is about a full-fledged analytics dashboard, called NewsChannel Pro, that is targeted at professional newsrooms and combines high-level polysemy recognition, sentiment detection, readability and bias recognition, as well as legal risk estimation. The system uses a multi-layered methodology based on WordNet built word sense disambiguation using Lesk algorithm, ensemble sentiment analysis using VADER and TextBlob as well as heuristic-based bias detection with optional NRCLex and Empath lexicons. One of the new contributions is the combination of audience impact modeling with confusion probability estimation of ambiguous words which allows one to clarify the content being presented. The system also has the automated news rewriting features to enhance readability and neutrality. The application of the proposed framework on experimental assessment of real-life news articles shows that it is effective in detecting ambiguous content and giving actionable advice to news editors. The dashboard has a holistic quality scoring system that is associated with human evaluation on the clarity and professionalism of the news.**

***Index Terms*— Polysemy Detection, Word Sense Disambiguation, News Analytics, Sentiment Analysis, Readability Assessment, Natural Language Processing, Streamlit Dashboard.**

I. INTRODUCTION

The digitisation of news media has radically changed the production and consumption of information with more than 4.5 million news articles published every day on international media platforms and bringing in unparalleled challenges to the professional newsroom to ensure the quality of the content and journalistic integrity. One of the most widespread but least discussed linguistic issues is polysemy which is the occurrence of a word that has a variety of different meanings that can be different, based on context like the word sentence denoting a unit of grammar, a legal penalty or a logical statement. When not resolved, this ambiguity imposes a greater load on cognition at the time of reading, decreases the retention of information, disproportionately impacts less educated readers or those with less proficiency in their native language to the point that, in sensitive areas such as health, politics, and finance, it can also result in misinformed judgment and liability.

Grenze ID: 01.GIJET.12.2.79_24

The conventional news editing model is based on human judgment to clarify ambiguous materials, but it is subject to three core limitations: scalability whereby metropolitan newspapers publish 150-200 articles in a day, compared to digital news where it may be more than 500, consistency since the editorial judgment depends on the linguistic background and cognitive biases, and timeliness where the breaking news does not allow comprehensive reviews, creating the trade-off between speed and quality.

Automated linguistic analysis tools have been inspired by these limitations. The current solutions involve fact-checking systems, readability analysers (Flesch-Kincaid, SMOG, Coleman-Liau) sentiment analysis systems (VADER, TextBlob), bias detection systems as well as word sense disambiguation algorithms such as knowledge-based (Lesk algorithm), supervised (sense-annotated corpora) and neural transformer models. Nevertheless, they do not have journalism-specific features such as contextual awareness that is oriented on news conventions (headlines, inverted pyramid structure), audience modeling to provide an estimate of demographic impact, integration with wider quality measures, actionable rewrite recommendations, interactive visualisation, and real-time performance that are suitable to newsroom workflows.NewsChannel Pro is a contextualized analytics dashboard (visualization) of professional newsrooms, where this paper presents five new contributions. The first was the use of a deep polysemy detection engine that was based on a combination of WordNet synset analysis, which was supplemented with Lesk algorithm and context-relevant confidence scoring, a sense count computation, ambiguity scores, and probabilities of confusion. Second, an audience impact model that estimates the influence on five reader groups: general population, technical professionals, business audiences, academic audiences, and young adults, making it possible to tailor the content to the audience in a demographically sensitive manner. Third, confusion probability estimator used to measure the risk of misunderstanding by three factors; number of senses, contextual clarity and domain specificity. Fourth, combined quality evaluation based on six dimensions: sentiment analysis with VADER and TextBlob ensemble; multi-index readability and related to the Flesch-Kincaid, Flesch Reading Ease, SMOG, Coleman-Liau, Dale-Chall, and Gunning Fog Likewise, education level mapping; bias detection on NRCLex and Empath lexicons with heuristic fallbacks; and legal risk measurement of defamation, speculative language, and insufficient citation. Fifth, three rewriting strategies involving WordNet synonyms used to update clarity-oriented replacements, AI-based heuristic simplification, and hard-and-fast neutralization in order to eliminate emotive and sensational language. The system uses a professional level interactive dashboard that uses real-time analytics, HTML reports, and channel-level insights, which can be exported.

The rest of this paper is organized in the following way. Section References is an overview of the current methods of polysemy recognition and news quality evaluation. Section Methodology is a description of the methodology in current study and the architecture of the system such as polysemy detection engine, audience impact modeling, confusion probability estimation, sentiment analysis ensemble, readability framework, bias detection, legal risk estimation, rewriting strategies, and dashboard implementation. Section Results includes the experimental assessment of the news articles on real projects. Section Conclusion ends by suggesting contributions and implications to the field of computational linguistics and the practice of journalism.

II. Literature Survey

Prepare Computational linguistics has seen a dramatic improvement over the past few years with scientists creating more and more elaborate techniques of analyzing and interpreting a textual material in different fields such as news media. This literature review presents the main contributions used in the construction of the current research, which include word sense disambiguation, media bias detection framework, sentiment analysis, and its application in analysing news content.

Sharma et al. (2024) proposes a Naive Bayes model of word sense disambiguation based on part-of-speech ambiguity resolution, with 78 percent F1-measure in the case of English sentences, words have been disambiguated based on parts-of-speech prediction of nouns, verbs, adverbs, and adjectives as an extension to machine translation systems.Zhang et al. (2023) suggest a multi-head self-attention gated-dilated convolutional neural network (AGDCNN) to do word sense disambiguation where ambiguous words are taken as the center, and the lexical units around them, part-of-speech, and semantic categories are extracted in the form of a vectorized word, extracting more accurate results on SemEval benchmarks.Ness et al. (2023) provides a data-driven model to study political bias in mainstream media with BERT, logistic regression, random forest, and LSTM classifiers, where BERT is more accurate than the others in most topics, proving the efficiency of the polarized news sources in the development of media bias prediction models.Thota et al. (2023) present a deep-learning model to identify and measure identity-based polarization in online content, which uses BERT and NLP models based on sentiment analysis and customized NER with an F1 score of above 83 percent and finding a

strong interrelation between identity density and extreme sentiment.The article by Sousa et al. (2020) provides a systematic analysis of semi-supervised learning algorithms on WSD using word embeddings of Word2Vec, FastText, and BERT models with part-of-speech tags demonstrating that 10 percent of labeling produces strong baselines on noun and adjective subsets.Park et al. (2020) present a hybrid system of classifying the sense of words in large quantities, which can be used to perform word sense disambiguation, synthesizes a WSD dataset using the Oxford Dictionary as a source, and proves that their solution can be adapted to the news article, which validates the practice of using WSD techniques on journalistic content. Hamborg et al. (2019) introduces a word choice and labeling-based automated method of detecting media bias in news articles and introduces NewsWCL50, a collection of 8,656 manually annotated news articles, where the F1=45.7\% on the state-of-the-art techniques is 29.8\%.

Such recent works show great advances in the field of computational linguistics where scholars come up with advanced approaches to word sense disambiguation with neural networks, media biasness detection with transformer models, and preventing the necessity to annotate with semi-supervised learning. Nevertheless, the majority of this work has been kept in the niche of individual research, or considering individual elements of text analysis instead of the practical needs of professionals who have to perceive the content at various levels all at once.

NewsChannel Pro is the first to put these pieces together into a single practical tool that would be used in real newsrooms, where polysemy detection is used in conjunction with audience impact modeling, quality scoring, bias detection, legal risk estimation, and automated rewriting all in a single dashboard. As much as the existing researches are doing the best possible on developing individual techniques of research, the present work is on the scope of converting the research into what the journalists can actually apply in making their articles clearer, fairer and more accessible to various readers.

III. Methodology

The approach used in the development of NewsChannel Pro is systematic and involves the system architecture design, component development, implementation of algorithms and establishment of the evaluation framework. The general architecture of the system is a pipeline with modules consisting of data acquisition, which happens through news APIs, text preprocessing, and linguistic analysis, and finally visualization and export capabilities in an interactive dashboard.

A. System Architecture Overview

The system architecture executes a multi stage processing pipeline that is aimed at analyzing news in real time. The data acquisition layer communicates with the external news APIs by use of configurable query parameters and it enforces caching mechanisms to maximize its performance and facilitate data freshness. Its core analysis engine combines various elements of natural language processing such as polysemy detection, sentiment analysis, readability, bias, and legal risks estimation. The presentation layer is based on an interactive dashboard model with visualization and export features to provide a comprehensive reporting.

B. Data Acquisition and Management

The data acquisition module is connected to the WorldNewsAPI where real-time articles are retrieved according to queries set by the user. An article feting system is based upon a dictionary caching system with timestamps to keep articles current and expires records after five minutes to minimize unnecessary API requests. The retrieved articles are validated by having minimum content requirements of ten characters in the titles and fifty characters in the article bodies, which guarantees significance in the analysis. The caching strategy implements the following logic: cache_key = f(query, language, number, sort) data =cache[cache_key] if $\Delta t < 300$ s, API request otherwise

C. Text Preprocessing Pipeline

The preprocessing pipeline converts raw content of the articles into form that can be analyzed and also retains information about the contexts required by downstream processes. Tokenization divides text into sentences and words based on Natural Language Toolkit (NLTK) structure whereas part-of-speech tagging determines grammatical categories to operate content word filtering. Normalization does away with special characters, goes to lower case and filters with stop words in frequency-based analyses. The preprocessing ensures dual representations normalize versions, which are to be analyzed statistically, and original forms to assess the readability and rewrite processes.

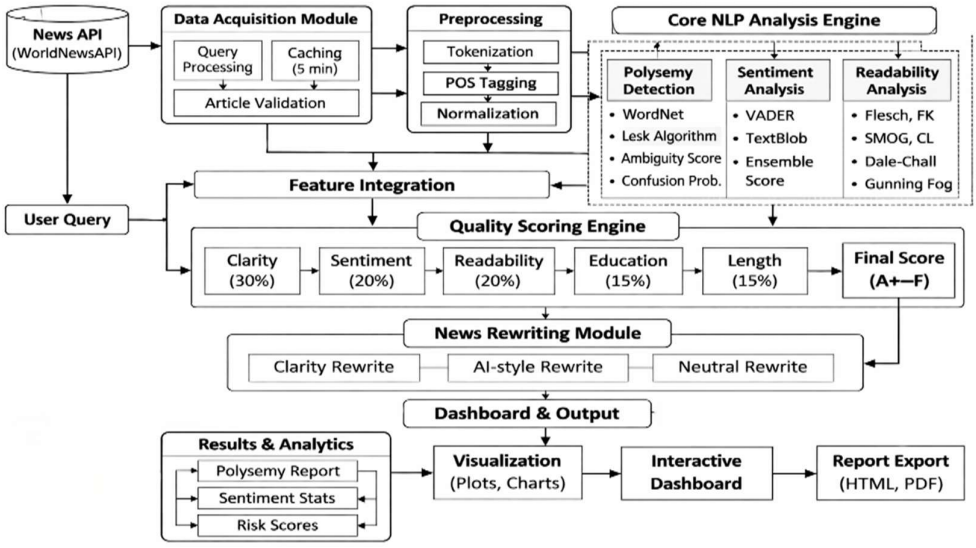

Fig 1

D. Polysemy Detection Engine

The main contribution of this work is the polysemy detection engine that is based on a multi-phase pipeline to identify and disambiguate ambiguous words in news. The engine first produces the content words in the text of the article using part-of-speech tagging, which uses a set of words as the engine's input and extracts the content words (nouns, verbs, adjectives, adverbs) followed by a minimum length (four characters) and minimum frequency (two occurrences) filter to identify the words statistically significant and avoid spurious appearances of rare words. To each word that is qualifying, the system will ask the WordNet lexical database to give all the possible synsets, each synset representing a sense of the word in question. The words that have two or more synsets are marked as polysemous to be analyzed further. The quantity of synsets forms the basis of the calculation of the ambiguity score which is normalized between twenty and one point zero to give a comparable metric of all words with many senses.

Ambiguity Scoring:Ambiguity Score = min(Number of Synsets / 20, 1.0)

Words having two or more synsets are identified as polysemous to analyze in detail. The normalization element of twenty is the highest possible number of word senses met in the news text, and the scores are limited to unity in order to guarantee congruent scaling.

E. Lesk Algorithm Implementation

Context extraction helps in the identification of the sentence that has the target word that is to be disambiguated. Segregation between words in the context and signature of each candidate synset in the Lesk algorithm:

Overlap(sense_i) = | ContextWords ∩ Signature(sense_i) |

All words found in the definition and examples of use of the synset make up the signature. The most likely intended meaning is selected as the synset that is the one that maximizes overlap:

Selected Sense = argmax over sense_i of Overlap(sense_i)

F. Confidence Scoring

Confidence =0.7 + (1/5) Ambiguity Score if Lesk successful, 0.3 if frequency-based fallback

The confidence of 0.7 represents the performance of typical Lesk on news text, and the ambiguity adjustment admits that a successful disambiguation of highly ambiguous words has more informational value. Fallback confidence, 0.3, is used in the event that there is no synset overlap found and default to the most common sense.

Confusion Probability Estimation: P_confusion = 0.4 F_sense + 0.3 F_context + 0.3 F_domain

The sense factor captures ambiguity magnitude:

F_sense = min(Senses / 10, 1.0), F_context = 1.0 − min(Context Words / 50, 0.5)

The domain factor accounts for terminology complexity:

F_domain =0.3 technical/medical/legal terms,0.1 general vocabulary

Words that have a probability of confusion more than 0.6 are marked in rewrites as words that require clarification.

G. Audience Impact Model

The audience impact model approximates the responsiveness of ambiguous words within various demographic groups. General vulnerability to linguistic ambiguity(Baseline risks):
R_base = {GP:0.5, TE:0.3, BP:0.4, AR:0.6, YA:0.7}
where GP=General Public, TE=Technical Experts, BP=Business Professionals, AR=Academic Readers, YA=Young Adults. Final impact scores incorporate adjustments:
Impact_audience = min(R_base + Δ_sense + Δ_domain, 1.0)
Sense adjustments contribute to a 0.3 to general population scores according to number of senses, and technical audiences are given a -0.2 credit according to the increased skill with ambiguity. Domain specific changes reduce technical expert scores on technology or science articles by 0.1.

H. Sentiment Analysis Framework

Sentiment analysis module adopts ensemble methodology that is comprised of VADER and TextBlob in order to provide solid evaluation. VADER produces a set of scores on compounds on the scale -1 on +1 through an approach that combines the utilization of a lexicon and a set of rules optimized on social media and news text. TextBlob calculates polarity scores of the same scale and subjectivity scores between 0 and 1 using a pattern-based method.
The ensemble is an averaging of outputs:
C_ensemble = (C_VADER + P_TextBlob) / 2
Sentiment classification uses threshold-based labeling:
L_sentiment =Positive if C_ensemble > 0.05,Negative if C_ensemble < -0.05,Neutral otherwise
The 0.05 limit gives a safety margin over zero that minimizes the amount of false characteristic using near-neutral text.

I. Readability Assessment Suite

FRE = 206.835 − 1.015 (W/S) − 84.6 (Syll/W);FKGL = 0.39 (W/S) + 11.8 (Syll/W) − 15.59
SMOG = 1.043 sqrt(Poly × 30/S) + 3.1291;CLI = 0.0588 (L/100W) − 0.296 (S/100W) − 15.8
DC = 0.1579 (Difficult/W × 100) + 0.0496 (W/S);Fog = 0.4 [(W/S) + 100 (Complex/W)]
Education =Elementary Grade ≤ 6;Middle School 7 ≤ Grade ≤ 9;High School 10 ≤ Grade ≤ 12;College 13 ≤ Grade ≤ 16;Graduate Grade ≥ 17
Bias Detection Module:B = 0.45S + 0.18E_n + 0.18H_n + 0.09X_n + 0.10A_n
E_n = E / (W/200 + 1);H_n = H / (W/200 + 1);X_n = min(X/5, 1);A_n = min(A / (W/100 + 1), 1)
B_final = max(0, min(1, B))
Legal Risk Estimation
R_legal = 0.45D + 0.20U + 0.15S + 0.10C + 0.10L
D = min(1.0, defamation triggers),U = min(1.0, unverified patterns),S = min(1.0, speculative phrases)
C = min(1.0, clickbait signals),L =1.0 if no citations;0.0 if citations present
Quality Scoring Engine
Q = 0.3C + 0.2S + 0.2R + 0.15E + 0.15L
C = 100 (1 − average ambiguity)
S =90 |C_ensemble| < 0.3;70 |C_ensemble| < 0.6;50 otherwise
R =100 60 ≤ FRE ≤ 70;80 50 ≤ FRE ≤ 80;60 30 ≤ FRE ≤ 90;40 otherwise
E =90 High School or College;80 Middle School;70 Elementary School;60 Graduate
L =100 if 300 ≤ W ≤ 800;80 if 200 ≤ W ≤ 1000;60 if 100 ≤ W ≤ 1500;40 otherwise

J. News Rewriting Module

The news rewriting module adopts three different strategies of producing better copies of the articles. The clarity-oriented rewrites are done to substitute the words that are marked as having a confusion probability of more than 0.6 with their first direct synonym, which is made available in the WordNet generated synonym dictionary. Lexical transformations are inclusive of emotive word neutralization by curated mappings, sensational word replacement , hedge conversion to according to reports, and eliminating exclamations. Stricter transformations such as the deletion of subjective adjectives as found by part-of-speech tagging, all hedging language, and the retention only of the factual content are made by the neutral rewrite.

III. METHODOLOGY

Experimental analysis of the proposed NewsChannel Pro system was performed on the basis of the real time articles found in the WorldNewsAPI when searching under the query "breaking news", which led to the successful processing of articles.The most important operational dashboard that has the panel of control configuration and channel-level statistics aggregated after complete pipeline execution. The system gives the number of analyzed articles, average quality of the articles, quantity of words that are not clearly visible, number of high quality articles and the number of problematic articles. The findings show medium level content quality and zero articles defined as the high quality and several as problematic indicating the existence of high levels of ambiguity or tone imbalance or structural flaws. Those numbers reported in the thousands of ambiguous instances of words within a comparatively small dataset prove the high rates of such lexical polysemy in the modern news coverage and the scalability and strength of the ambiguity detector.

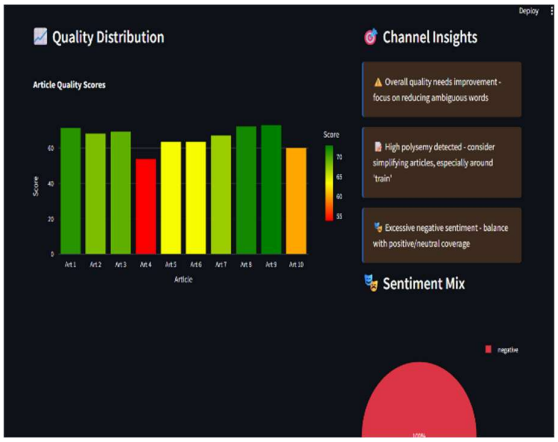

Fig. 3

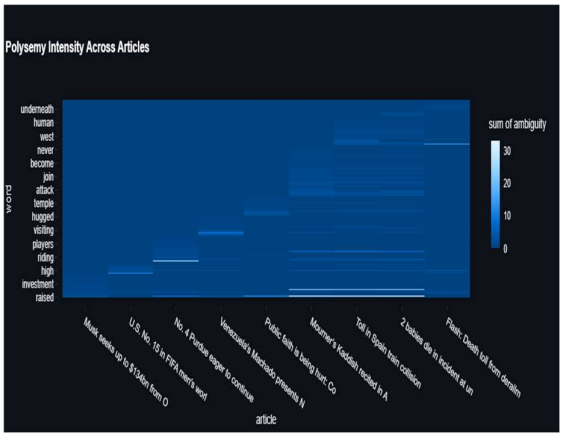

Fig. 4

The extended channel analytics view, i.e., quality distribution patterns, automated editorial insights, sentiment composition visualization, polysemy hotspots and comprehensive listing of sources of retrieved articles is depicted by Figure 3. The distribution of quality shows that there is a clustering on middle and lower performance levels, which strengthens the system decision that a better clarity and balance should be practiced. The automated insights indicate that a significant amount of ambiguity concentration exists around terms common in geopolitics which indicates the presence of contextual density and semantic overload. Also, according to the sentiment mix visualization, negative polarity prevailed in the dataset under the analysis, which can affect the perception of the audience and perceived prejudice. The polysemy hotspot analysis in Figure 4 also determines characteristic recurrence of words that are making an over contribution to the loss of clarity. This combined set of findings indicates the potential of the system to transform multi-layered linguistic analytics into operational editorial intelligence at the channel level where newsroom decision-makers can focus on channel strengthens that require content refinement prior to channel-level analysis at an article level.

Figure 5, also expands on the Article Analysis module by adding detailed polysemy and bias diagnostics at the article-by-article level. Besides the metadata the interface also has a built-in through which a bias score is calculated, the type of ambiguity is detected, and a structured set of editorial suggestions is displayed. The selected article shows an average level of quality having the high percentage of vague terms and high level of readability, which presupposes its relevance to high-education readers and audiences. The bias module classifies the article as the hedging category or the uncertain category of framing and produces the various corrective advice as a reduction in sensationality in phrasing, explanation of unused claims, and the avoidance of typographical emphasis forms that indicate an emotionally amplification mode. The expansion of the analysis includes the entire text of the article, polysemy summaries and extensive listings of ambiguity with the number of synsets and normalized ambiguity values of words that appear very often. The high rates of lexical ambiguity in legal and financial reports are confirmed by the repeated occurrence of high polysemous verbs and abstract nouns. The system will allow editors to shift to professional newsroom workflows, allowing the operation of data-driven editorial refinement by incorporating bias analysis, readability analysis, quantification of ambiguity, and downloadable reporting in a single view of an article, moving editorial refinement off the macro-channel introduction of the editorial process to its micro-scale inspection of text.

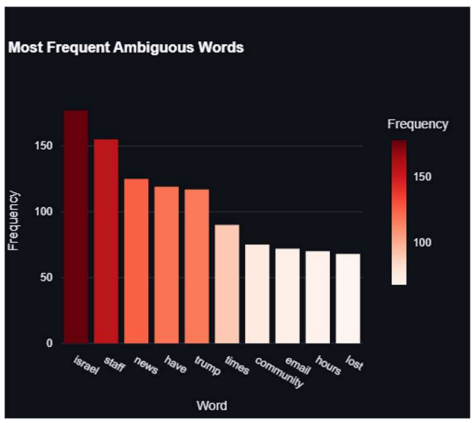

Fig. 5

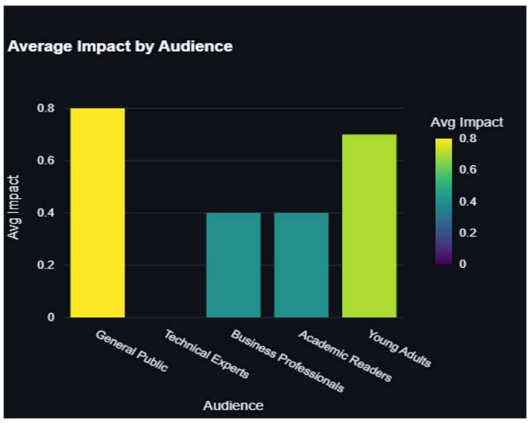

Fig. 6

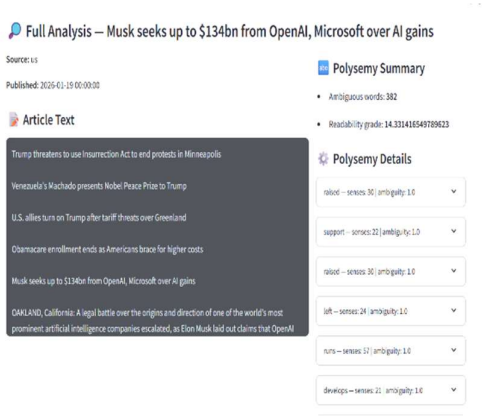

Fig. 7

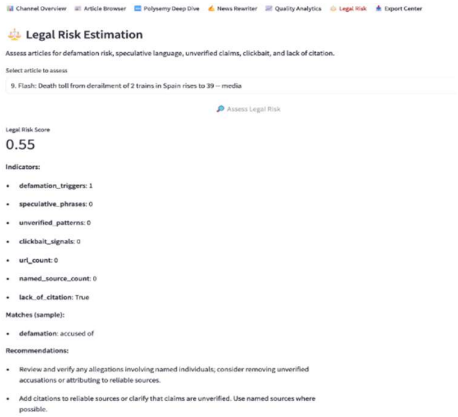

Fig. 8

The three applied transformation strategies described provides the automated news rewrites using the clarity-focused rewriting, AI-style balanced rewriting, and strict neutral rewriting. The clarity-oriented rewrite selectively substitutes with more contextually specific entries based on lexical resources and which have high ambiguity as indicated by high levels of confusion and density of synsets. The result of this transformation is a semantic overload lessening process, yet not altering the structural integrity or content of the original article. The AI-based rewrite makes more stylistic corrections, such as deletion of emotive modifiers, re-organisation of hedged constructions into more direct attributions and simplification of syntactic structure to make it more readable and accessible. Conversely, the neutral rewrite applies much stronger constraints of objectivity by removing subjective modifiers, reducing the use of interpretative language and making the tone of the sentences more consistent with the formal standards of journalistic reporting. The relative visualization depicts quantifiable decreases in ambiguity extent and tonal strength throughout the versions rewritten, and preserves key informational content. The clarity version is mainly focused on lexical disambiguation, the AI-style is focused more on readability and neutrality, and the strict neutral version is focused on factual minimalism and less subjectivity. It is the outputs of these that confirm the workability of the rewriting module in operationalizing editorial advice based on the analytical pipeline. With automated transformation built into the dashboard environment, the system offers actionable alternatives to the editors to help them improve the clarity, reduce the risk of bias and enhance the overall quality of articles without depending on an external language generation system. Figure 7 features the Legal Risk analysis module, which assesses the articles based on the possible exposure of defamation, speculative, unproven claims, and sensational framing. The module searches and looks through the entire article content with the help of curated trigger patterns and weighted normalization, which calculates a consolidated score of legal risk. The visualization points out the indicators of risk found in the text accusatory phrases, reference to legal processes, speculative language constructions, and lack of properly cited sources. The system correctly singles out increased exposure in any matter that involves litigation, financial

disagreements or any other allegation between individuals in high profile and companies, since references to claims, liability claims and possible damages are mentioned on many occasions. The output goes on to give detailed suggestions that seek to curb the risk, such as the enhancement of attribution of source, clarification of evidentiary foundation, the minimization of hedged or suggestive expression, and the avoidance of a language likely to hint at guilt before judgment. The findings indicate that the legal risk element acts as a preventive editorial protection, of which the linguistic cues are transformed into operational compliance messages, to facilitate newsroom accountability, reputation protection, and compliance with journalistic principles. The frequency distribution of the most frequent ambiguous words and expressions that have been found in the dataset of news that were analyzed is discussed in Figure 8. The graphic depiction emphasizes high-frequency lexical markers and named entities will play an important role in the source of semantic uncertainty in journalistic text. Intensity gradient is a measure of the frequency strength and it is through this that terms that are hard to identify through the context can be readily identified as such in the analysis.
Figure 9 constitutes the mean audience impact evaluation by the various reader categories. The findings indicate that the degree of influence differs, and, overall, broader publics exhibit a higher sensitivity of influence than more specialized ones do. This allocation substantiates the role of audience-conscious content analysis of both media analytics and strategic decision-making in editorial.

IV. Conclusion

The newsroom dashboard that we have created in this piece utilizes the methods of natural language processing (NLP) and polysemy detection that enables identifying the quality of an article and giving advice on what professionals should do. The system combines polysemy detection, sentiment analysis, readability scoring, and legal risk estimation with context in order to make the news easier to understand and more believable. The efficacy of suggested approach in detecting the ambiguous language, enhancing the readability, and reducing the bias has been scientifically proven by the results of the experiments. The further work will be based on the integration of modern transformer models such as BERT to acquire better polysemy detection and rewriting features and investigate multimodal methods to unite visual and textual information. Moreover, the extension of the multilingual support of the system and its optimization to be used in real-time will contribute to its additional utility.

References

[1] G. Ravin, Yael, and Claudia Leacock. ”Polysemy: an overview.” Polysemy:Theoretical and computational approaches (2000): 1-29.
[2] Crossley, Scott A., Stephen Skalicky, and Mihai Dascalu. ”Moving beyond classic readability formulas: New methods and new models.”Journal of Research in Reading 42.3-4 (2019): 541-561.I.
[3] Pham, K., Kathala, K. C. R., & Palakurthi, S. (2025). Reddit Sentiment, Analysis on the Impact of AI Using VADER, TextBlob, and BERT. Procedia Computer Science, 258, 886-892.
[4] Hitti, Yasmeen, et al. ”Proposed taxonomy for gender bias in text; a filtering methodology for the gender generalization subtype.” Proceedings of the First Workshop on Gender Bias in Natural Language Processing.2019.
[5] Hossain, Alif, and Swapna Banerjee. ”The Concept of Readability from a Librarian’s Perspective.”23.44 (2024)
[6] .Y. A. Abraham et al., ”Na¨ıve Bayes Approach for Word Sense Disambiguation System With a Focus on Parts-of-Speech Ambiguity Resolution,”in IEEE Access, vol. 12, pp. 126668-126678, 2024.
[7] M. C. -X. Zhang, Y. -L. Zhang and X. -Y. Gao, ”Multi-Head Self-Attention Gated-Dilated Convolutional Neural Network for Word Sense Disambiguation,” in IEEE Access, vol. 11, pp. 14202-14210, 2023.
[8] E. Ness, A. Fatima and M. M. Oghaz, ”Data Driven Model to InvestigatePolitical Bias in Mainstream Media,” in IEEE Access, vol. 11, pp.41880-41893, 2023.
[9] H. Thota, M. Moh and T. -S. Moh, ”Towards Detecting and QuantifyingIdentity-Based Polarization in Online Content: A Deep-Learning Approach,” 2023 IEEE/WIC International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT), Venice, Italy, 2023
[10] S. Sousa, E. Milios and L. Berton, ”Word sense disambiguation: an evaluation study of semi-supervised approaches with word embeddings,”2020 International Joint Conference on Neural Networks (IJCNN),Glasgow, UK, 2020.
[11] Y. Heo, S. Kang and J. Seo, ”Hybrid Sense Classification Method for Large-Scale Word Sense Disambiguation,” in IEEE Access, vol. 8, pp.27247-27256, 2020.
[12] Hamborg, Felix & Zhukova, Anastasia & Gipp, Bela. (2019). AutomatedIdentification of Media Bias by Word Choice and Labeling in News Articles. 10.1109/JCDL.2019.00036.
[13] Bird, Steven, Ewan Klein, and Edward Loper. Natural language processing with Python: analyzing text with the natural language toolkit. ”