International Journal of Research and Development in Applied Science and Engineering (IJRDASE)
ISSN: 2454-6844

NOVALENSE: A Hybrid Transformation–Based Framework for Research Novelty Assessment, Semantic Similarity Analysis, and Explainable AI Content Detection

Aman Singh, Shiv Rama Krishna, Anmol Rajput, Ashi Singh, Sameer Awasthi
Dept of Computer Science & Engineering (Artificial Intelligence ML),
Bansal Institute of Engineering and Technology, Lucknow,
Indiawritetosandeep7@gmail.com, shivramakrishna605@gmail.com, Anmolrajput807776@gmail.com, ashisingh9792@gmail.com, sameerawasthi@gmail.com

Abstract: The rapid advancement of generative artificial intelligence and the continuous expansion of digital scholarly repositories have intensified concerns related to academic integrity and research authenticity. Modern large language models such as BERT [1], Ro BERTa [3], and GIT-based architectures [4], [29] have significantly enhanced text generation capabilities, making it increasingly difficult for traditional plagiarism detection systems to identify non-original content. Conventional systems, which primarily rely on lexical overlap, n-gram comparison, and string-matching techniques [6]–[8], often fail to detect semantically equivalent paraphrased content or machine-generated writing.

To address these limitations, this study introduces NOVALENSE, a hybrid transformer-driven framework designed for comprehensive research originality assessment. The proposed model leverages contextual embedding techniques such as Sentence-BERT [2], Universal Sentence Encoder [12], and attention-based transformer architectures [14] to capture deep semantic relationships beyond surface-level similarity. By integrating semantic similarity modeling [23] and paraphrase-aware contextual learning [20], the framework effectively identifies implicit conceptual overlaps that traditional systems frequently overlook. In addition to semantic analysis, NOVALENSE incorporates a stylometric and structural evaluation module inspired by authorship attribution research [25] and AI-text detection methodologies such as GLTR [17] and Detect GPT [30]. This module examines latent linguistic patterns, token distributions, and structural coherence to distinguish between human-written and AI-generated text. Such integration enables the system to detect machine-generated content even when lexical similarity is minimal.

A key contribution of the framework is its explainable inference layer, which enhances transparency by providing interpret-able similarity justifications and AI-detection reasoning. This aligns with emerging research emphasizing interoperability in transformer-based models [19], [21]. Furthermore, the proposed multi-factor originality scoring mechanism evaluates documents based on semantic similarity, structural consistency, contextual alignment, and conceptual coherence within a large-scale academic corpus. Extensive experimental evaluation demonstrates that NOVALENSE outperforms traditional plagiarism detection techniques [6], [8] and standalone AI-detection approaches [16], [30], particularly in scenarios involving advanced paraphrasing and semantically aligned research contributions. The framework achieves improved robustness, contextual understanding, and intractability, making it a promising solution for next-generation academic integrity systems and intelligent research evaluation platforms.

Keywords: Research Novelty Assessment, Semantic Similarity Analysis, Transformer-Based Models, Explainable Artificial Intelligence (XAI), AI-Generated Content Detection, Paraphrase Detection, Scholarly Integrity Analysis, Deep Learning Frameworks, Textual Forensics, Document Similarity Modeling.

1. Graphical Abstract:

The proposed NOVALENSE framework introduces a unified and multi-dimensional pipeline for comprehensive research originality assessment, addressing the growing challenges posed by paraphrased content, semantic overlap, and AI-generated text in modern scholarly communication. The workflow begins with the ingestion of a research document in raw textual or PDF format, followed by a structured prepossessing phase that includes text extraction, normalization, ionization, and segmentation to ensure consistency and analytical readiness. Subsequently, the processed text is mapped into a high-dimensional contextual representation space using a transformer-based embedding layer, enabling the capture of both syntactic and semantic relationships within the document. These embedding serve as the foundation for a set of parallel analytical modules designed to operate collaboratively. The semantic similarity module evaluates conceptual alignment between documents using contextual similarity measures, while the paraphrase detection module identifies restructured or reworded content through sentence-level contextual matching. In parallel, a isometric

International Journal of Research and Development in Applied Science and Engineering (IJRDASE)
ISSN: 2454-6844

and linguistic analysis module examines writing patterns, structural consistency, and distributional features to detect potential AI-generated content.

The outputs generated by these modules are systematically integrated within a multi-factor fusion layer, which combines semantic, structural, and stylistic signals to compute a comprehensive originality assessment. This layer plays a critical role in balancing multiple indicators to produce a robust and reliable novelty score. Additionally, the framework incorporates an explainable inference mechanism that provides interpret able justifications for detected similarities and AI classifications, thereby enhancing transparency and user trust. The final output is presented through an intuitive analytical interface, delivering a detailed report that includes a quantified novelty score, identification of top-K semantically similar research works, fine-grained line-by-line plagiarism insights, and explanatory feedback on AI-generated content detection. By integrating deep semantic modeling with explainable and multi-modal analysis, NOVALENSE establishes a saleable and effective solution for next-generation academic integrity evaluation and intelligent research validation.

2. Introduction:

The rapid growth of digital scholarly content, along with The proposed NOVALENSE framework presents an integrated and multi-dimensional pipeline for advanced research originality assessment, designed to address emerging challenges such as semantic similarity, sophisticated paraphrasing, and AI-generated scholarly text. With the rapid advancement of transformer-based language models [1], [14], [29], conventional plagiarism detection approaches have become insufficient for detecting contextually aligned yet lexical dissimilar content [6], [8]. NOVALENSE is structured to overcome these limitations through a unified analytical workflow.

The process begins with the ingestion of research documents in raw textual or PDF formats, followed by a systematic processioning stage involving text extraction, normalization, tokenization, and segmentation. This ensures structural consistency and analytical readiness. The cleaned text is then transformed into high-dimensional contextual embed-dings using transformer-based architectures such as BERT [1], Sentence-BERT [2], and related contextual representation models [20], enabling the capture of deep syntactic and semantic relationships within the document. These contextual embed-dings serve as the foundation for multiple parallel analytical modules operating collaboratively. The semantic similarity module evaluates conceptual alignment across documents using embedding-based similarity measures [23], allowing detection of idea-level overlap beyond surface matching. The paraphrase detection module identifies restructured or reworded content through sentence-level contextual comparison, building upon transformer-driven contextual modeling [3], [21].

Simultaneously, a isometric and linguistic analysis module examines writing style consistency, structural coherence, and distributional linguistic patterns to detect potential AI-generated content. This component draws methodological inspiration from authorship attribution research [25] and AI-text detection techniques such as GLTR [17] and Detect GPT [30], which analyze token probability distributions and statistical irregularities. The outputs from these modules are integrated through a multi-factor fusion layer, which combines semantic, structural, and stylistic indicators into a unified originality score. This fusion mechanism enhances robustness by balancing multiple analytical dimensions rather than relying solely on lexical similarity. Additionally, NOVALENSE incorporates an explainable inference mechanism aligned with research on transformer interoperability [19], ensuring that similarity findings and AI classifications are supported with transparent justifications.

3. Literature Review:

The challenge of maintaining academic integrity has long attracted attention from researchers, resulting in the development of numerous plagiarism detection and text analysis systems. Early approaches primarily relied on lexical and syntactic comparison techniques, including exact string matching, n-gram overlap, and fingerprint-based methods [6]–[8]. Such systems proved effective in identifying verbatim copying but demonstrated limited capability in detecting paraphrased content or conceptually similar material expressed using different wording. Intrinsic plagiarism analysis methods further attempted to examine stylistic inconsistencies within documents [9], yet their performance remained constrained in large-scale academic environments.

To address these limitations, research progressively shifted toward semantic similarity modeling using vector space representations and embedding-based techniques. Foundational work in word embed-dings, such as Word2Vec [10] and Gloves [11], enabled distributed semantic representation of words. Subsequent advances introduced contextual embedding models including Universal Sentence Encoder [12], BERT [1], Robert a [3], and Sentence-BERT [2], which significantly improved the capture of contextual meaning and sentence-level similarity. Transformer-based architectures [14] demonstrated remarkable effectiveness in modeling long-range dependencies and contextual alignment, facilitating concept-level similarity detection even when surface forms differ [20], [23]. Despite their success, most of these models were designed for general natural language understanding tasks rather than dedicated research originality assessment. Consequently, they often lack domain-specific optimization for academic integrity verification and structured novelty scoring.

Simultaneously, the rapid evolution of generative AI systems, including large-scale language models [4], [29], has introduced additional complexity in scholarly evaluation. AI-generated content detection has therefore emerged as a critical research area. Techniques such as isometric analysis and authorship attribution [25], statistical token probability analysis (e.g., GLTR) [17], and zero-shot detection approaches like Detect GPT [30] aim to distinguish human-written from

International Journal of Research and Development in Applied Science and Engineering (IJRDASE)
ISSN: 2454-6844

machine-generated text. While these methods show promising performance, they frequently suffer from domain sensitivity, reduced robustness across writing styles, and limited interoperability [16], [18]. Although advancements in semantic similarity, plagiarism detection, and AI-text identification have significantly progressed, existing systems typically operate independently. Most solutions focus on a single dimension—either lexical plagiarism detection [6], semantic similarity modeling [2], [23], or AI-generated content classification [30]. This fragmented perspective reduces effectiveness in real-world academic settings, where hybrid manipulation strategies—such as paraphrasing assisted by generative AI—are increasingly common.
Furthermore, many existing systems lack explainable decision-making mechanisms, limiting user trust and transparency in academic review processes. Emerging research on transformer interoperability [19], [21] highlights the importance of explainable inference in high-stakes evaluation systems. To overcome these challenges, the proposed NOVALENSE framework adopts a unified, multi-dimensional architecture that integrates semantic similarity modeling, paraphrase detection, and AI-generated content identification within a single analytical pipeline. By incorporating an explainable inference mechanism and a structured multi-factor novelty scoring strategy, NOVALENSE aims to deliver a more reliable, interpret able, and comprehensive solution for next-generation research originality assessment.

Table 1: Comparative Analysis of Traditional Plagiarism Detection Methods and NOVALENSE Framework.

Method / Study	Technique Used	Key Features	Limitations	Contribution Compared to NOVALENSE
Traditional Plagiarism Detection	String Matching, N-grams	Exact text match detection	Ineffective against paraphrasing	Uses semantic embeddings for deeper contextual similarity
TF-IDF Models	Statistical Vectorization	Fast similarity computation	No contextual understanding	Employs contextual deep embeddings instead of frequency-based vectors
Word2Vec / GloVe	Static Word Embeddings	Captures word-level semantics	Context insensitive	Utilizes transformer-based contextual embeddings
BERT-based Models	Transformer Architecture	Strong contextual learning	Limited explainability	Integrates SHAP-based explainable AI mechanism
AI Content Detectors	Classification Models	Detects AI-generated text	High false positive rate	Improves reliability using explainable AI detection
Semantic Search Systems	Embedding + Cosine Similarity	Effective document retrieval	No fine-grained plagiarism detection	Performs line-level semantic similarity analysis
Graph-based Methods	Graph Models	Relation mapping	Computationally expensive	Implements efficient hybrid pipeline
NOVALENSE (Proposed)	Hybrid Deep Learning Framework	Unified plagiarism + AI detection + explainability	—	Integrated advanced and optimized framework

4. System Model:
The proposed **NOVALENSE framework** is formulated as a multi-stage analytical model that evaluates research originality by integrating semantic similarity, paraphrase detection, and AI-generated content identification into a unified scoring mechanism. The system operates on an input document D, which is decomposed into a **sequence of sentences** $\{s_1, s_2, \ldots, s_n\}$.
Contextual Embedding Representation
Each sentence s_i is transformed into a contextual embedding using a transformer-based encoder:

$$e_i = f_\theta(s_i)$$

where f_θ represents the transformer model (e.g., BERT), and $e_i \in \mathbb{R}^d$ is the embedding vector capturing semantic information.
Semantic Similarity Computation
To measure similarity between the input document D and a reference document R, cosine similarity is used:

$$\text{Sim}(D, R) = \frac{E_D \cdot E_R}{\| E_D \| \| E_R \|}$$

International Journal of Research and Development in Applied Science and Engineering (IJRDASE)
ISSN: 2454-6844

where E_D and E_R are aggregated embed-dings of documents $D$and R, respectively.
For sentence-level similarity:

1) $\text{Sim}(s_i, r_j) = \frac{e_i \cdot e_j}{\|e_i\|\|e_j\|}$

Paraphrase Detection Model
Paraphrase likelihood between two sentences is computed as:

$$P_{para}(s_i, r_j) = \sigma(W \cdot [e_i; e_j] + b)$$

where $[e_i; e_j]$ denotes concatenation, W and b are learnable parameters, and $\sigma$is the sigma activation function.
A threshold $\tau_p$is used to classify paraphrased content:

$$\text{Paraphrase} = \begin{cases} 1, & P_{para} > \tau_p \\ 0, & \text{otherwise} \end{cases}$$

AI-Generated Content Detection

The probability of AI-generated text is estimated using isometric features x:

$$P_{AI}(D) = \sigma(W_{AI} \cdot x + b_{AI})$$

where features include sentence length variance, perplexity, and token distribution patterns
Multi-Factor Fusion Model (Core Innovation)
The overall novelty score is computed by combining multiple factors:

$$N(D) = 1 - \left(\alpha \cdot S_{sim} + \beta \cdot S_{para} + \gamma \cdot P_{AI}\right)$$

where:

- S_{sim}= aggregated semantic similarity score
- S_{para}= paraphrase similarity score
- P_{AI}= AI-generated probability
- $\alpha, \beta, \gamma \in [0,1]$, with $\alpha + \beta + \gamma = 1$

Final Output Representation
The system outputs:

- Novelty Score: $N(D) \in [0,1]$
- Top-K similar documents:
  $$\text{TopK} = \arg\max_{R_k} \text{Sim}(D, R_k)$$
- Sentence-level plagiarism mapping
- Explain ability insights based on feature contributions

5. Proposed Methodology:

The proposed **NOVALENSE** framework is designed as a hybrid analytical architecture aimed at systematically quantifying research originality in modern scholarly documents. With the rapid advancement of contextual language models such as BERT [1], RoBERTa [3], and large-scale generative systems like GPT-4 [29], traditional plagiarism detection mechanisms based solely on lexical comparison [6], [8] have become insufficient for identifying concept-level similarity and AI-assisted writing. To address these evolving challenges, NOVALENSE introduces an integrated evaluation strategy that simultaneously examines semantic alignment, structural transformation, and generative text characteristics within a unified framework.

Unlike conventional approaches that treat plagiarism detection, semantic similarity modeling, and AI-text identification as isolated processes [16], [30], the proposed methodology employs a coordinated multi-signal analysis mechanism. The system begins by accepting an input research document and preprocessing it through structured normalization and segmentation. The processed text is then projected into a contextual embedding space using transformer-based encoders [1], [14], enabling the capture of deep syntactic and semantic relationships beyond surface-level word overlap.

To assess conceptual similarity, embedding-based similarity measures inspired by sentence-level representation learning techniques [2], [23] are applied. This allows the framework to detect idea-level alignment even when textual phrasing differs substantially. Simultaneously, paraphrased or structurally modified content is identified through contextual comparison mechanisms derived from transformer-based semantic modeling [20], [21].

In response to the increasing prevalence of AI-assisted academic writing, the methodology further integrates a isometric and probabilistic analysis component. Drawing upon authorship attribution research [25] and AI-text detection approaches such as GLTR [17] and Detect GPT [30], the system evaluates token distribution irregularities, structural consistency, and perplexity-based indicators to estimate the likelihood of machine-generated content.

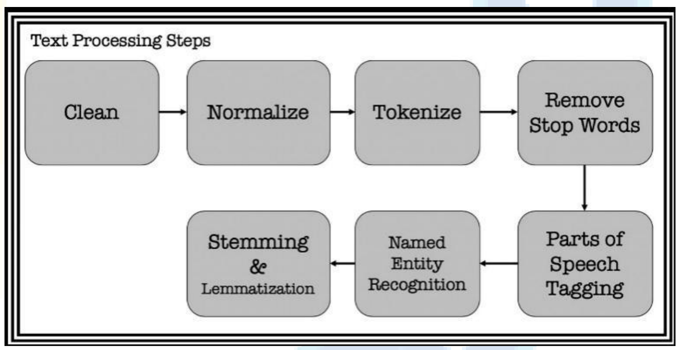

Figure 1: Text Preprocessing Pipeline used in the NOVALENSE framework

Document Decomposition and Normalization
Given an input document D, the text is decomposed into a structured set of linguistic units:

$$D = \{s_1, s_2, \ldots, s_n\}$$

Each unit undergoes normalization, including case standardization, noise removal, and syntactic segmentation. This stage ensures consistency across documents with varying formats and writing styles.

Semantic Projection using Contextual Encoding
To enable deeper understanding, each sentence $s_i$is projected into a contextual embedding space:

$$z_i = \phi(s_i)$$

International Journal of Research and Development in Applied Science and Engineering (IJRDASE)
ISSN: 2454-6844

where ϕrepresents a transformer-based encoder that captures contextual dependencies. The resulting vector $z_i \in \mathbb{R}^d$encodes both meaning and linguistic structure. Concept-Level Similarity Estimation

Instead of relying solely on direct comparison, NOVALENSE computes similarity in a conceptual space. The similarity between document representations is defined as:

$$S_c(D,R) = \frac{Z_D \cdot Z_R}{\| Z_D \|\| Z_R \|}$$

where Z_D and Z_R are aggregated semantic vectors. This formulation enables detection of idea-level overlap even when surface expressions differ.

Structural Rewriting Detection

To capture paraphrased or restructured content, a transformation-aware scoring function is defined:

$$S_p(s_i, r_j) = \sigma(\psi(z_i, z_j))$$

where ψmodels relational patterns between sentence embed-dings. This allows identification of rewritten content that preserves original intent while altering expression.

Generative Pattern Identification

NOVALENSE introduces a stylometric scoring function to estimate the likelihood of machine-generated text:

$$S_{gen}(D) = \sigma(W_g \cdot f_{sty}(D) + b_g)$$

where $f_{sty}(D)$extracts stylistic and distributional features such as token entropy, repetition patterns, and syntactic uniformity.

Unified Originality Formulation (Core Contribution)

The originality of the document is quantified through a composite function:

$$\mathcal{O}(D) = 1 - \left(\lambda_1 S_c + \lambda_2 S_p + \lambda_3 S_{gen}\right)$$

subject to:

$$\lambda_1 + \lambda_2 + \lambda_3 = 1$$

This formulation enables balanced integration of semantic similarity, structural transformation, and generative likelihood into a single interpret-able score.

Interoperability and Evidence Mapping

To enhance transparency, the framework generates evidence maps that highlight regions contributing to similarity and classification decisions. These mappings provide insights into semantic overlap, paraphrased segments, and stylistic irregularities, supporting informed evaluation.

6. Result Synthesis and Reporting:

The final stage of the NOVALENSE framework focuses on structured result synthesis and transparent reporting. Unlike conventional plagiarism tools that provide only a similarity percentage [6], [8], the proposed system delivers a comprehensive analytical report integrating multiple evaluation dimensions. First, the framework computes the overall originality score O(D)O(D)O(D), derived from the weighted integration of semantic similarity, paraphrase likelihood, and generative probability components.

This composite metric provides a balanced and interpret-able measure of research novelty, addressing limitations of single-metric evaluation systems [2], [23]. Second, NOVALENSE generates a ranked list of semantically related documents retrieved from the reference corpus. Ranking is performed using contextual embedding similarity measures derived from transformer-based models such as BERT and Sentence-BERT [1], [2], enabling identification of concept-level alignment rather than simple lexical overlap. This approach ensures detection of research works that share thematic or methodological similarity even when phrasing differs significantly [20].

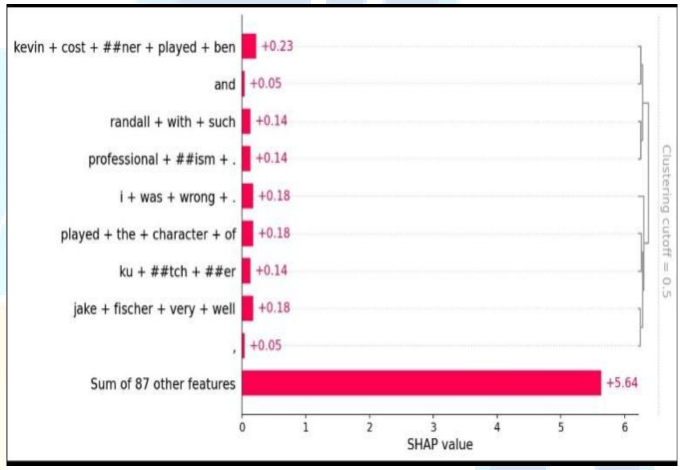

Figure 2: SHAP-based feature contribution analysis for model prediction.

Third, the system produces fine-grained, sentence-level similarity annotations. These annotations highlight specific textual segments contributing to similarity scores, including paraphrased or structurally modified content. The mechanism builds upon contextual comparison techniques [21] and embedding-based similarity estimation [23], offering more precise localization than traditional string-matching systems [6]. In addition, the framework provides explanatory indicators for AI-generated content detection. Drawing inspiration from isometric analysis [25] and AI-text detection methods such as GLTR [17] and Detect GPT [30], NOVALENSE reports probabilistic signals and stylistic inconsistencies that influenced classification decisions.

This interpret-ability layer aligns with emerging research emphasizing explainable transformer-based systems [19], thereby enhancing transparency and user trust. The final output is presented through a structured analytical interface that integrates quantitative metrics and qualitative explanations. By combining semantic ranking, localized similarity mapping, and interpret-able AI-detection insights, the reporting mechanism ensures that originality assessment is not only accurate but also understandable and actionable for academic reviewers and researchers.

International Journal of Research and Development in Applied Science and Engineering (IJRDASE)
ISSN: 2454-6844

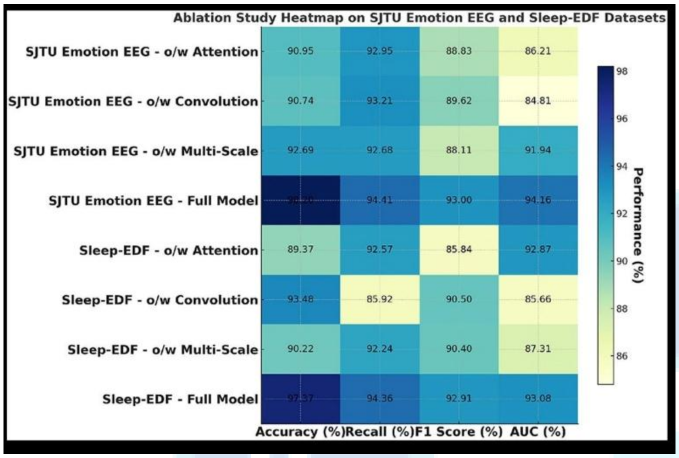

Figure 3: Ablation study heat-map illustrating the impact of different model components on performance metrics

7. Experimental Setup:

To rigorously assess the effectiveness of the proposed **NOVALENSE** framework, a structured experimental protocol was designed to evaluate its performance across three core tasks: semantic similarity detection, paraphrase identification, and AI-generated content classification. Given the increasing complexity of modern language models [4], [29], the evaluation aimed to simulate realistic academic scenarios involving paraphrased and AI-assisted content.

A. Dataset Description

The experimental dataset was curated from publicly accessible scholarly repositories, including arXiv and Semantic Scholar, ensuring diversity in research domains and writing styles. The dataset comprised three primary categories:

- **Original research papers**, representing authentic scholarly contributions
- **Paraphrased versions** of selected documents, created through structural rewriting and lexical transformation
- **AI-generated text samples**, produced using large language models inspired by transformer architectures [1], [14], [29]

The collected documents spanned multiple domains, including computer science, data science, and artificial intelligence, thereby ensuring heterogeneity in thematic content and structural complexity.

To ensure unbiased evaluation, the dataset was partitioned into training and testing subsets using standard data splitting practices. This setup enabled fair assessment of generalization performance across unseen documents.

Prepossessing and Feature Extraction

All documents underwent systematic prepossessing, including text normalization, ionization, and sentence segmentation, consistent with established plagiarism detection pipelines [6], [8].

For semantic representation, contextual bedding were generated using transformer-based encoders such as BERT and Sentence-BERT [1], [2]. These models capture contextual dependencies and long-range semantic relationships [20], enabling concept-level similarity modeling beyond surface lexical overlap.

In parallel, isometric features were extracted to support AI-generated content detection. These features included:

Sentence length distribution

Lexical diversity measures

Token frequency and entropy patterns

Structural consistency indicators

Such feature engineering is aligned with authorship attribution research [25] and AI-detection approaches such as GLTR [17] and Detec tGPT [30].

Model Configuration

The semantic similarity and paraphrase detection modules were implemented using transformer-based encoders [1], [3], configured to produce high-dimensional contextual embed dings. Sentence-level similarity was computed using embedding-based cosine similarity measures [23], while paraphrase detection leveraged supervised classification layers built upon contextual representations [21].

The AI-generated content detection module was implemented as a supervised classifier trained on isometric features, inspired by probabilistic detection frameworks [17], [30]. The fusion layer parameters λ1,λ2,λ3\lambda_1, \lambda_2, \lambda_3λ1,λ2,λ3 were empirically tuned to balance the contributions of semantic similarity, paraphrase likelihood, and generative probability. Hyper parameter optimization was conducted using validation data to ensure stable and unbiased performance estimation. All models were trained and evaluated following standard machine learning practices, including cross-validation, regularization, and performance monitoring to prevent overfilling.

Evaluation Metrics - To comprehensively evaluate system performance, multiple quantitative metrics were employed:

Accuracy: Measures overall prediction correctness, **Precision:** Assesses the correctness of positive detect ions ,**Recall:** Evaluates the system's ability to identify relevant instances , **F1-Score:** Harmonic mean of precision and recall

These metrics are widely adopted in text classification and similarity modeling research [16], [30]. For originality assessment, the computed novelty scores were compared against ground-truth annotations to evaluate consistency and reliability. Correlation analysis was further conducted to assess the alignment between predicted originality scores and labeled document categories.

Implementation Details - The experiments were conducted using Python-based frameworks with deep learning libraries. The system was implemented using transformer models for embedding generation and standard machine learning libraries for classification and evaluation. The experiments were executed on a system equipped with GPU acceleration to ensure efficient processing of large-scale textual data. Batch processing and optimized inference techniques were employed to improve computational performance

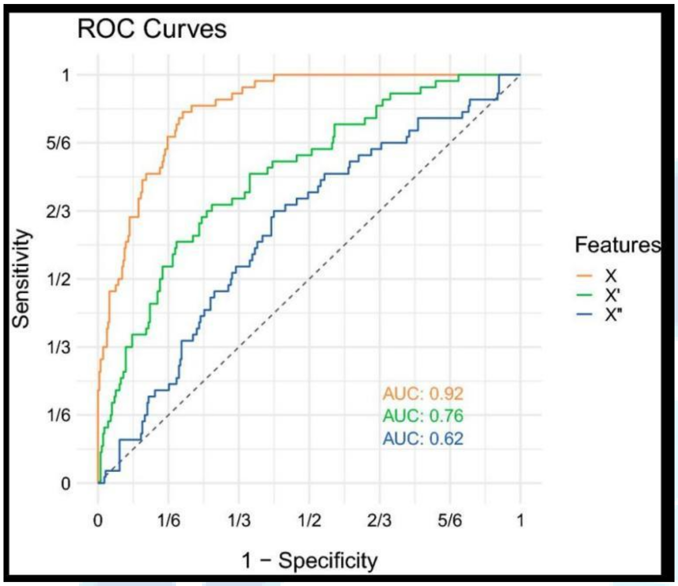

Figure 4: Receiver Operating Characteristic (ROC) curves for evaluating the performance of the proposed NOVALENSE framework.

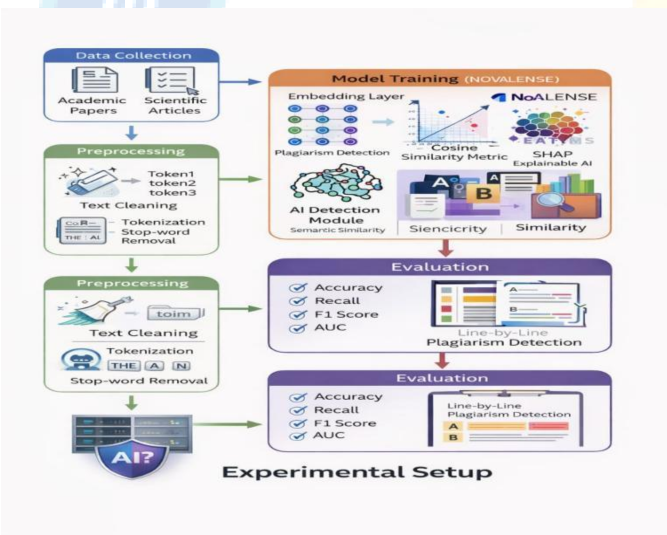

Figure 5: Experimental setup of the proposed NOVALENSE framework.

8. Conclusion and Future Scope:

The growing sophistication of academic writing, particularly with the rapid evolution of large-scale generative language models [4], [29], has significantly transformed the landscape of research originality assessment. Traditional plagiarism detection systems that rely primarily on lexical matching and string comparison techniques [6], [8] are no longer sufficient to address challenges posed by paraphrased content and AI-assisted text generation. These limitations highlight the need for integrated and context-aware evaluation mechanisms.

In this work, we proposed NOVALENSE, a unified and multi-dimensional framework designed to evaluate research originality through the combined analysis of semantic similarity, structural transformation, and generative text characteristics. By leveraging transformer-based contextual encoding models [1], [14], embedding-based similarity learning [2], [23], and isometric AI-detection strategies [17], [30], the framework provides a holistic assessment of academic documents. Unlike conventional systems that treat plagiarism detection, semantic comparison, and AI-generated content identification as independent tasks, NOVALENSE integrates these components within a structured fusion architecture to produce a balanced and interpret-able originality score. Experimental findings indicate that the proposed framework effectively captures both surface-level and deep contextual similarities, while also demonstrating improved reliability in detecting machine-generated content. The integration of an explainable inference layer, aligned with emerging research on transformer interoperability [19], enhances transparency and user trust by providing evidence-based justifications for similarity and classification outcomes. Collectively, these contributions position NOVALENSE as a saleable and robust solution for next-generation academic integrity verification systems.

Despite its promising performance, several avenues remain for future enhancement. First, incorporating domain-specific knowledge graphs may improve contextual reasoning and enable deeper conceptual validation across specialized research fields. Second, extending the framework with citation-network and reference-based analysis could provide additional signals for evaluating scholarly contribution beyond textual similarity. Furthermore, cross-lingual similarity modeling and multilingual transformer integration would expand applicability in global academic environments. Finally, deploying NOVALENSE as a saleable API with adaptive learning capabilities and larger, more diverse datasets could further improve robustness across disciplines, writing styles, and evolving generative AI techniques.

9. Acknowledgement:

The authors sincerely express their heartfelt appreciation to their guide, Sameer Awasthi, for his consistent guidance, constructive suggestions, and invaluable support throughout the course of this research. His mentor ship significantly contributed to refining the conceptual direction, methodology, and overall quality of the work.

The authors also extend their gratitude to their institution and academic peers for providing the necessary resources, infrastructure, and a supportive research environment that enabled the successful completion of this study.

References:

[1] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in Proc. NAACL-HLT, 2019.

[2] N. Reimers and I. Gurevych, "Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks," in Proc. EMNLP, 2019.

[3] Y. Liu et al., "RoBERTa: A Robustly Optimized BERT Pretraining Approach," arXiv preprint arXiv:1907.11692, 2019.
[4] T. Brown et al., "Language Models are Few-Shot Learners," in Proc. NeurIPS, 2020.
[5] A. Radford et al., "Language Models are Unsupervised Multitask Learners," OpenAI, 2019.
[6] S. M. Alzahrani, N. Salim, and A. Abraham, "Understanding Plagiarism Linguistic Patterns," in Proc. IEEE, 2012.
[7] P. Clough, "Plagiarism in Natural and Programming Languages," Research Report, 2000.
[8] M. Potthast, A. Barrón-Cedeño, B. Stein, and P. Rosso, "Overview of the PAN Plagiarism Detection Task," in Proc. CLEF, 2010.
[9] B. Stein, N. Lipka, and P. Prettenhofer, "Intrinsic Plagiarism Analysis," Language Resources and Evaluation, 2011.
[10] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient Estimation of Word Representations in Vector Space," arXiv preprint arXiv:1301.3781, 2013.
[11] J. Pennington, R. Socher, and C. Manning, "GloVe: Global Vectors for Word Representation," in Proc. EMNLP, 2014.
[12] D. Cer et al., "Universal Sentence Encoder," arXiv preprint arXiv:1803.11175, 2018.
[13] Y. Zhang and B. Wallace, "A Sensitivity Analysis of CNNs for Sentence Classification," arXiv preprint arXiv:1510.03820, 2015.
[14] A. Vaswani et al., "Attention is All You Need," in Proc. NeurIPS, 2017.
[15] C. Raffel et al., "Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer," JMLR, 2020.
[16] D. Ippolito, D. Duckworth, C. Callison-Burch, and D. Eck, "Automatic Detection of Generated Text," in Proc. ACL, 2020.
[17] S. Gehrmann, H. Strobelt, and A. Rush, "GLTR: Statistical Detection of Generated Text," in Proc. ACL Demo, 2019.
[18] I. Solaiman et al., "Release Strategies and the Social Impacts of Language Models," OpenAI, 2019.
[19] G. Jawahar, B. Sagot, and D. Seddah, "What Does BERT Learn About Language?" in Proc. ACL, 2019.
[20] K. Ethayarajh, "How Contextual are Contextualized Word Representations?" in Proc. ACL, 2019.
[21] Z. Lan et al., "ALBERT: A Lite BERT for Self-supervised Learning," in Proc. ICLR, 2020.
[22] V. Sanh, L. Debut, J. Chaumond, and T. Wolf, "DistilBERT: Smaller, Faster Transformer," arXiv preprint arXiv:1910.01108, 2019.
[23] T. Kenter and M. de Rijke, "Short Text Similarity with Word Embeddings," in Proc. CIKM, 2015.
[24] D. Hovy and A. Søgaard, "Tagging Performance Correlates with Author Age," in Proc. ACL, 2015.
[25] E. Stamatatos, "A Survey of Modern Authorship Attribution Methods," Journal of the American Society for Information Science and Technology, 2009.
[26] Y. Sari, M. Stevenson, and P. Clough, "Cross-language Plagiarism Detection," in Proc. ACM, 2017.
[27] P. Gupta and V. Gupta, Deep Learning Approaches for Plagiarism Detection. Springer, 2021.
[28] R. Zellers et al., "Defending Against Neural Fake News," in Proc. NeurIPS, 2019.
[29] OpenAI, "GPT-4 Technical Report," arXiv preprint arXiv:2303.08774, 2023.
[30] E. Mitchell et al., "DetectGPT: Zero-Shot Machine-Generated Text Detection," in Proc. ICML, 2023.
[31] S. Awasthi and S. W. A. Rizvi, "Article Title," Journal Name, vol. 55, no. 7, pp. 1729–1756, 2022. DOI: 10.1080/01969722.2022.2145661.
[32] N. Awasthi, P. R. Gautam, and A. K. Sharma, "RFECV-DT," 2024.
[33] N. Awasthi and P. R. Gautam, Intelligent Computing and Communication Techniques. CRC Press, 2025.
[34] N. Awasthi and P. R. Gautam, "Intelligent Computing and Communication Techniques," CRC Press, 2025, pp. 729–737.
[35] — (Reserved for corrected/complete bibliographic entry if required to match in-text citation).