Journal of Intelligent Information and Communication

Research Paper

Automatic Title Generation for Materials Science Literature Based on Pointer-Generator Network

Tamara Blagojević[1, *], Dragan Vasiljević[2] and Nina Radojičić[2]

[1] Faculty of Information Engineering, Singidunum University, 11010, Belgrade, Serbia
[1] Faculty of Information Technology, Alfa BK University, 11070, Belgrade, Serbia
*Corresponding author: tamara.b@singidunum.ac.rs

Abstract. With the rapid development of materials science literature, issues related to the organization and application of information are gradually coming to the forefront. To address the issue of automatic title generation, this paper proposes a domain-adaptive pointer generation network, specifically tailored to the diverse needs of materials science texts. To effectively address the issues of rare technical terms and complex language expressions, this paper employs a dual-stream encoder to incorporate domain-specific word embeddings and a dynamic gated pointer mechanism. The experiment used a large collection of well-organized materials from various fields. The proposed model has already surpassed traditional baseline methods. The BLEU score improved by 7.2%, the ROUGE-L score increased by 6.5%, and the average human evaluation increased by 0.17%. To reduce semantic drift and expand term coverage, all modules in these models are necessary. Titles are usually concise, technically accurate, and meet public expectations. In addition to providing a practical foundation for scientific titles, this article also proposes applying the system to other fields such as intelligent knowledge management and automated indexing.

Keywords: *Machine Learning, Deep Learning, Text Generation, Materials Science, Natural Language Processing, Information Retrieval*

Received on 19 September 2025, Accepted on 06 January 2026, Published on 23 January 2026

Introduction

Over the past twenty years, the number of publications in the field of materials science has continuously increased, and the research content has become increasingly complex [1]. With the dissemination of new technologies, discoveries, and methods, the number of papers published in journals and other technical reports has increased [2]. With the publication of a large number of new studies, experts have found that timely collection, organization, and application of scientific data is a major challenge [3]. The issue of academic paper titles must be addressed; high-quality titles can help in the collection and use of academic data in the digital age [4]. Title writing requires expert judgment and domain knowledge, is very time-consuming, and often involves biases and inconsistencies, making it unsuitable for large-scale systematic knowledge management [5]. There is an urgent need for automated, precise, and domain-specific title generation methods to support the continuous development of scientific communication in the field of materials science [6].

With the development of natural language processing and artificial intelligence, new applications for automatic content generation and knowledge extraction are emerging [7]. Attention-based models and sequence-to-sequence frameworks have performed well in abstractive summarization tasks for news and general science [8]. These models sometimes fail to retain necessary domain-specific terminology, handle uncommon vocabulary, and adhere to the standards of scientific writing, issues that are particularly evident in materials science literature [9]. To address the aforementioned limitations, Pointer-Generator Networks (PGNs) were developed. These networks can generate new expressions and copy important phrases from the input text to bridge the abstract differences between technical content and rare terms [10]. Current research is still scarce, mostly focusing on news or biomedical texts, rather than the dense, composite, and structurally complex narrative styles found in materials science papers. The main issue of applying these methods to the field of materials informatics has still not been resolved.

To address the open problem of automatic title generation for materials science literature, this paper introduces a domain-adaptive pointer generation network. Using specialized word embeddings and coverage-based decoding to more accurately represent the complex semantics of scientific materials while maintaining the conciseness of the titles. To test the data, a large number of materials science papers were selected and compared with high-performance baseline models. Analyze general statistical data and specific cases. Develop other specialized scientific library intelligent indexers for semantic search and knowledge graph construction, in addition to the methods presented here. The following section introduces the current state of scientific summarization; presents models and improvement methods, and shows detailed experimental results; discusses the overall impact of the research on materials science knowledge management.

Related Work

Review of Text Summarization Techniques in Science & Engineering

In the fields of science and engineering, text summarization methods are generally divided into two types: generative and extractive. Extraction methods identify and extract specific sentences or expressions from the original text based on the frequency of sentences or expressions appearing in the statistics. These are factually correct, but technical writing requires coherence and accuracy [11]. The purpose of the abstract method is to create new sentences to modify and condense the original content. This method is more readable, but due to certain terms in the document, it may change the meaning [12]. Neural network models with attention mechanisms and sequence models significantly improve the performance of abstractive summarization by enhancing contextual awareness [13]. Extraction and abstraction methods are not suitable for handling the specialized vocabulary and complex structures found in scientific papers, which often include formulas, rare names, and well-defined chapters [14]. A good abstract for scientific and engineering literature needs to be able to adaptively balance domain accuracy, conciseness, and clarity. Current research is focused on improving the integration of semantic knowledge and domain adaptation [15].

Prior Art in Title Generation Methods

Deep learning sequence-to-sequence (Seq2Seq) models are the primary method for generating scientific paper titles recently. An attention mechanism was added to help people better understand their context [16]. Pointer-generator networks can extend the aforementioned capabilities by directly copying prominent phrases or terms from the input, as demonstrated in scientific fields with unusual expressions and dense terminology [17]. These neural networks can generate titles and summaries of documents well, but they are not suitable for technical papers. Standard models often cannot fully capture the meanings specific to a particular field and may miss or incorrectly describe necessary experiments and material descriptions in complex scientific texts [18]. Some progress has been made through the adaptation of domain-specific embeddings and transfer learning, but many existing architectures do not fully meet the unique characteristics and requirements of titles in the field of materials science, which must be both novel and precise in terminology [19].

Materials Science Text Mining: Challenges & Solutions

Materials science literature involves a large number of specialized terms and extremely regular structures, posing particular challenges for automatic text processing systems [20]. Chemical names, abbreviations, formulas, property descriptors, etc., usually need to be processed using general natural language processing models, which may lead to misunderstandings or omissions of domain-specific knowledge [21]. Material texts often contain multimodal content, such as tables and graphical summaries, and use non-standardized terminology, which makes understanding and analysis more difficult [22]. In materials literature, training custom embeddings, domain-adapted tokenization, and ontology-based entity recognition have all been used to improve semantic accuracy and extraction accuracy [23]. It is a method that can solve the above-mentioned problems. Combining pointer-based mechanisms with coverage models can significantly reduce term omissions [24]. Carefully curated training datasets and expert-annotated corpora have also improved practical performance. The aforementioned improvements have achieved some success, but new issues still need to be addressed, such as scaling the model for new types of materials and integrating structured domain knowledge to ensure the model's generalizability and accuracy. Specialized, domain-aware materials science text mining solutions are needed [25].

Methodology

Architecture of the Enhanced Pointer Generator Network

The pointer-generator network has been optimized to accommodate the scientific nature and technical vocabulary of materials science papers. At the bottom of this structure, there is a dual-stream encoder that can simultaneously extract language and semantics from the initial abstract. The bidirectional gated recurrent unit used to capture sequence context is the first encoder path. A domain-adaptive Transformer encoder focuses on phrase-level dependencies and material-specific patterns. These two streams are used together to preserve the original style and content of the data as much as possible.

In the decoder, the context is generated by a dynamically adjusted pointer generation mechanism. Continual learning is not just about vocabulary-based predictions, but about creating a general term or copying a domain-specific term from the training data. The adaptive Softmax-sigmoid gate is set through a weighting mechanism, determining the extent to which information and pointer information are generated at each decoding step. A coverage vector is integrated into the decoder to reduce the generation of repetitive phrases that frequently appear in scientific content.

The learnable encoder global attention context determines the initialization of the decoder state. The model can focus on the different types of materials, experiments, and results that frequently appear in materials science research papers. In each decoding step, the joint attention distribution is used to align the generated tokens with the summary positions, while enhancing the effect of entity prominence through the following formula:

$$\alpha_{t,i}^{(j)} = \frac{\exp\left(f^{(j)}(h_i, s_t, c_{t-1})\right)}{\sum_{k=1}^{n} \exp\left(f^{(j)}(h_k, s_t, c_{t-1})\right)} \qquad \text{Eq.(1)}$$

Mathematically, the adaptive gating function of pointer probabilities is formulated to consider context awareness and domain-specific signals. These signals manifest as interaction terms between the sentence-level classifier output and the encoded semantic vectors:

$$p_{ptr}(t) = \sigma\left(w_{\text{gen}}^{\top}\left[s_t; c_t; \phi(E_{\text{domain}})\right] + b_{\text{gen}}\right) \qquad \text{Eq.(2)}$$

Cross-layer attention aggregation is a coverage control technique that prevents semantic drift and over-generation by aggregating attention weights at different times and network depths. To guide subsequent decoding results, an accumulated attention mask is generated:

$$\text{cov}_t = \sum_{\tau=1}^{t-1} \sum_{j=1}^{L} \alpha_{\tau}^{(j)} \qquad \text{Eq.(3)}$$

Figure 1 shows the complete network diagram. The dual-stream encoder, gated pointer mechanism, and coverage correction output are located on the right side, along with the typical data flow path from input summary to generated title. This design can dynamically switch between two modes. It can be used to create new phrases or for precise replication, applicable in materials science. A multifunctional and highly adaptable system for title generation in the field of scientific publishing has been built, based on the aforementioned architecture and new attention mechanisms, combined with pointer logic.

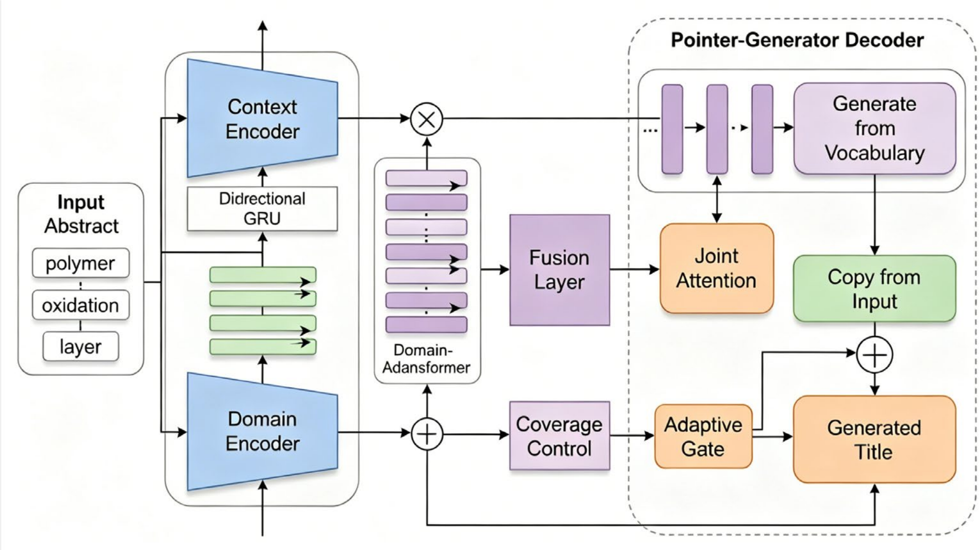

Figure 1. Overall Network Architecture of the Proposed PGN Model

Domain Adaptation: Materials Science Embeddings

The ability to identify and present specialized terminology, rare compound expressions, and varying nomenclature in the field of materials science is key to generating accurate and semantically reasonable titles. General language models often struggle to handle scientific details well and cannot accurately describe the techniques and content of specific fields. This study investigated the training of domain adaptation methods for material science-specific word embeddings to address this shortcoming. The goal is to enhance the semantic sensitivity of the encoder and the ability of the decoder to express complex scientific titles.

The foundation of these specialized embeddings includes journal abstracts, research papers, and technical glossaries in materials science. Develop a preprocessing program to retain atomic and molecular symbols, crystallographic parameters, composite material names, and multi-word technical expressions in the data. The base embedding matrix collects distributional features from a continuous bag of words and is then optimized through a masked language model to identify context and hierarchy in the literature.

Given the vocabulary V_{domain} tailored for materials science, each term w_i is embedded based on context-driven averaging weighted by co-occurrence statistics. The embedding vector E_i for term w_i is computed as follows:

$$E_i = \frac{1}{|C(w_i)|} \sum_{w_j \in C(w_i)} f_{co}(w_i, w_j) \cdot v_j \qquad \text{Eq.(4)}$$

Here, $C(w_i)$ denotes the local context set surrounding $w_i, f_{co}(w_i, w_j)$ is a learned cooccurrence weighting function, and v_j denotes hidden representations of related context terms.

Combining word-level matrices with subword-level embedding systems to address learning difficulties and enrich scientific vocabulary. Using a gating mechanism to aggregate subword vectors to represent and summarize complex scientific terms:

$$e_{\text{compound}} = g(\{e_{\text{sub}}^k : 1 \le k \le n\}) \qquad \text{Eq.(5)}$$

where e_{sub}^k are subword embeddings and $g(\cdot)$ is a parameterized, order-sensitive composition function such as a gated recurrent mapping.

To further standardize the embedding space, domain adaptation loss was added during supervised training. The following formula indicates that both jointly promote semantic cohesion between related material concepts and ensure sufficient separation between incomparable scientific terms:

$$\mathcal{L}_{emb} = \lambda_1 \cdot \sum_{(i,j) \in S} \|E_i - E_j\|^2 + \lambda_2 \cdot \sum_{(i,j) \in D} \max\left(0, m - \|E_i - E_j\|^2\right) \qquad \text{Eq.(6)}$$

Here, S is the set of similar term pairs, D the set of dissimilar pairs, λ_1, λ_2 are weighting coefficients, and m is a margin parameter for contrastive separation.

In each decoding step, the generated embeddings are added to the encoder-decoder pipeline and dynamically combined with the attention context to generate the output vector:

$$o_t = \psi(u_t^{enc}, v_t^{att}, E_{\text{domain}}) \quad \text{Eq.(7)}$$

where o_t is the final decoder output at time step t, u_t^{enc} is the encoder state, v_t^{att} is the context vector from the attention mechanism, and ψ is a non-linear transformation combining all three components.

This method enables the model to obtain richer contextual awareness for both frequent and infrequent scientific terms. The accuracy of title generation has made significant progress in terms of the rich information needed for domain adaptation.

Loss Functions and Optimization Strategies

In the highly specialized field of materials science, the success of the improved pointer generator network is also attributed to its architectural changes, effective loss functions, and efficient optimization schemes. This paper integrates multiple supporting objectives into a comprehensive loss function to meet the model's output requirements as well as the lexical accuracy and structural conditions of scientific titles.

The combination of sequence-level generation loss, pointer-based loss, coverage adjustment, and domain embedding regularization is the focus of the entire training. Sequence-to-sequence models use the first cross-entropy to predict the target sequence. For each time step t, the negative log-likelihood of the target token y_t^* is formally:

$$\mathcal{L}_{seq} = -\sum_{t=1}^{T} \log P(y_t^* \mid y_{<t}, X) \quad \text{Eq.(8)}$$

where X is the input abstract and $y_{<t}$ are prior outputs of the decoder.

The pointer-generator paradigm also needs to dynamically choose at each time step whether to directly copy a word from the input or generate a word from the vocabulary. Modeling the binary generation/copy decision as a Bernoulli process term enforces the correctness of this choice through the pointer loss:

$$\mathcal{L}_{ptr} = -\sum_{t=1}^{T} \left[s_t^* \log p_{ptr}(t) + (1 - s_t^*) \log\left(1 - p_{ptr}(t)\right)\right] \quad \text{Eq.(9)}$$

with s_t^* indicating the ground-truth copying indicator and $p_{ptr}(t)$ being the pointer probability at time t.

To address the characteristic risk of generating repetitive phrases in abstract models, this paper introduces coverage loss, which means that the excessive attention to previously focused areas in the input is the sum of the coverage penalties. The total sum of attention overlap is called coverage penalty:

$$\mathcal{L}_{\text{cov}} = \sum_{t=1}^{T} \sum_{i=1}^{N} \min\left(\alpha_{t,i}, \text{cov}_{t,i}\right) \quad \text{Eq.(10)}$$

where $\alpha_{t,i}$ is the current attention weight and $\text{cov}_{t,i}$ is the cumulative attention up to time t for input token i.

The weighting of these components is the total loss function, and the hyperparameters are set according to the industry's balance requirements for accuracy, originality, and redundancy control:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{seq}} + \gamma_1 \mathcal{L}_{ptr} + \gamma_2 \mathcal{L}_{\text{cov}} + \gamma_3 \mathcal{L}_{\text{emb}} \quad \text{Eq.(11)}$$

where $\gamma_1, \gamma_2, \gamma_3$ are coefficients selected via validation performance with particular attention to terminology recall and output fluency.

The optimization pipeline is a mini-batch gradient descent technique that uses domain-adaptive embeddings and batch normalization to train stability and weight initialization. Gradient clipping is used to prevent gradient explosion in deep sequence models. In the later stages of training, periodic sampling was introduced to improve the model's generalization ability and reduce exposure bias. By gradually exposing the model to the sequences it has created itself.

Pointer generation networks achieve good generalization on synthetic and rare technical vocabulary, while maintaining consistency, diversity, and domain fidelity in generating titles for materials science literature. This is achieved through this comprehensive loss design and rigorously tuned optimization routine.

Experimental Setup

Dataset Construction and Preprocessing

The foundation of title generation in materials science is limited by the scale and specificity of the data. In this study, data acquisition focused on public and private corpora. These repositories also contain a large number of peer-reviewed conference proceedings, journal abstracts, and other curated databases, such as SpringerMaterials and the Materials Project. The generated corpus contains over one million examples, including new and old words, and records their original locations for reference.

Both the diversity of topics and the quality of publications have been met by the source selection. Classic metallurgy, novel polymers, and advanced nanomaterials were categorized into publications on experimental synthesis, computational discoveries, and structural analysis. The dataset was deduplicated and randomly stratified across subdomains to eliminate sample bias. The automated validation script checks technical integrity, ensures document legality, and standardizes the language.

Normalize the text to preserve the semantic information of chemical symbols and formula expressions. By using a normalized cleaning procedure, encoding errors, typographical noise, and OCR artifacts in digital archives are systematically eliminated. The subword segmentation algorithm breaks down complex multi-word terms into subword tokens, ensuring the integrity of entities and supporting nested combinatorial symbols.

The noise reduction protocol eliminates redundant metadata, duplicate chart descriptions, and redundant chapter titles. This gives the protocol a clear summary-title pair. In order to expand the range of terms, add uncommon and compound descriptors to the continuously expanding vocabulary dictionary. Statistical stitching ensures that the title matches the abstract, while frequency-based dynamic filtering excludes low-information documents.

Figure 2 shows the complete process of data flow. It starts with document collection, followed by a series of cleaning and tokenization steps, vocabulary expansion, and the assembly of the corpus. This process will provide high-resolution and domain-relevant training and testing data for subsequent steps.

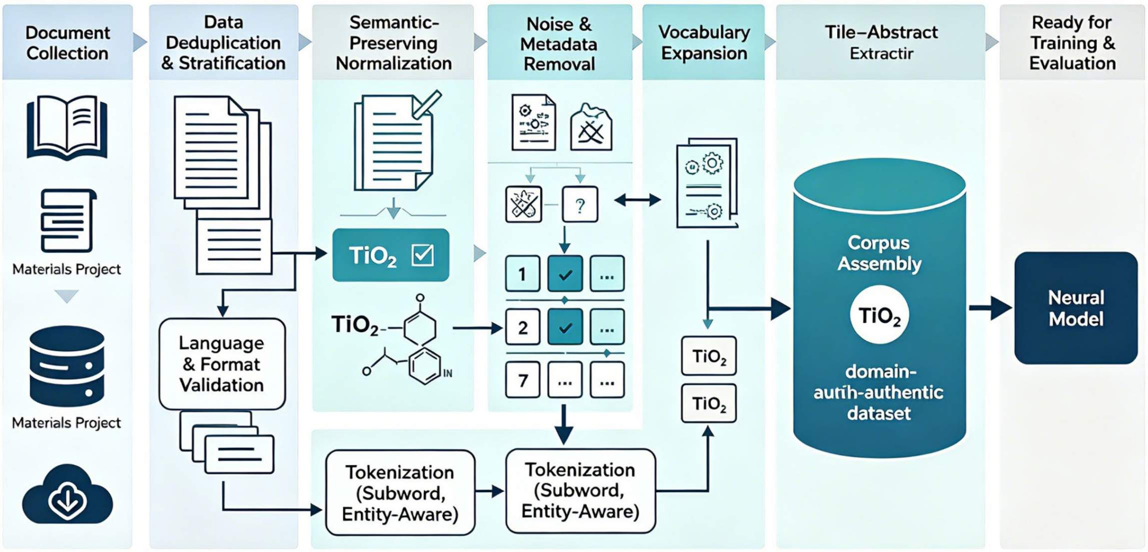

Figure 2. Data Processing and Experimental Workflow

Evaluation Metrics (BLEU, ROUGE, Human Assessment)

Automated metrics for the quality of title generation and expert subjective evaluations have already been used. The purpose of selecting the metrics is to ensure semantic equivalence in the specific field of materials science while being lexically close to the reference titles.

BLEU, ROUGE (recall-oriented evaluation of summaries), and METEOR (explicitly ordered translation evaluation metric) are automatic evaluation metrics. BLEU is based on n-gram precision and is sensitive to exact lexical matches. Here is the method for calculating the geometric mean of the short-to-long penalty for the modified n-gram precision:

$$\mathrm{BLEU} = BP \cdot \exp\left(\sum_{n=1}^{N} w_n \log\, p_n\right) \qquad \text{Eq.(12)}$$

where BP is the brevity penalty, w_n are n -gram weights, and p_n are modified n -gram precisions up to order N.

In contrast, ROUGE, particularly ROUGE-L, focuses on recall and sequence matching and is defined via the longest common subsequence (LCS) approach:

$$\text{ROUGE-L} = \frac{(1+\beta^2) \cdot \text{LCS\_precision} \cdot \text{LCS\_recall}}{\text{LCS\_precision} + \beta^2 \cdot \text{LCS\_recall}} \qquad \text{Eq.(13)}$$

Here β balances recall against precision, emphasizing coverage of reference content.

The METEOR score uses stemming and synonymy mapping, calculating the weighted harmonic mean of precision and recall at the unit level, with a fragment penalty, making it relatively robust to synonyms and word order changes.

To evaluate the accuracy of the generated titles in terms of technical content, scientific relevance, fluency, and conciseness, a group of senior materials science researchers has been hired. A five-point Likert scale will be used for evaluation. The aggregate human assessment H can be represented mathematically as:

$$H = \frac{1}{M} \sum_{i=1}^{M} \sum_{j=1}^{K} s_{ij} \qquad \text{Eq.(14)}$$

where M is the number of experts, K the evaluation aspects, and s_{ij} the score for aspect j by expert i.

The comprehensive evaluation system will provide a specific, context-based assessment. Consider the accuracy, practicality, and academic value on a word-for-word basis.

Baseline Models and Comparison

This study also compared several well-known baseline models to provide a robust and reliable reference: the Transformer model, the standard sequence-to-sequence (Seq2Seq) model with attention mechanism, and top-tier extractive summarization methods. To ensure fairness, the Seq2Seq-attention baseline used a stacked bidirectional GRU with Luong attention mechanism, and the parameter settings were the same as those of the proposed encoder. For direct comparison, a six-layer encoder-decoder structure identical to the Transformer baseline was used, with the same embedding size and tokenization.

After using TextRank and customized domain-specific filters to prioritize key terms and formulas, the extraction baseline is consistent with the best level of current scientific abstract extraction. All models are trained and evaluated using the same dataset, batch size, training steps, and stopping criteria. To avoid tuning bias and ensure statistically valid comparisons, hyperparameters such as learning rate, dropout rate, and vocabulary constraints are the same for all architectures.

All model output submissions use the same blind evaluation and automatic scoring. The aforementioned strict control measures ensure that any increase in evaluation metrics can be attributed to changes in system design or application, rather than changes in parameters or data.

Results and Discussion

Quantitative Results & Ablation Study

Quantitatively evaluate the impact and influence of each new feature in the PGN system. As shown in Figure 3(a), due to the architecture-enhanced n-gram accuracy, the BLEU scores of PGN surpass those of the standard Transformer and Seq2Seq-attention baselines across five materials science datasets. As shown in Figure 3(b), there is a significant difference in the ROUGE-L scores. PGN can more accurately extract necessary content from reference titles through domain adaptation embeddings and pointer mechanisms. As shown in Figure 3(c), PGN leads in METEOR, particularly in terms of complex syntactic structures and hierarchical phrase relationships. Figure 3(d) shows the results of the manual evaluation, where experts gave significantly higher scores to the PGN outputs of summaries containing multiphase compounds and rare alloy names in terms of technical adequacy and term accuracy.

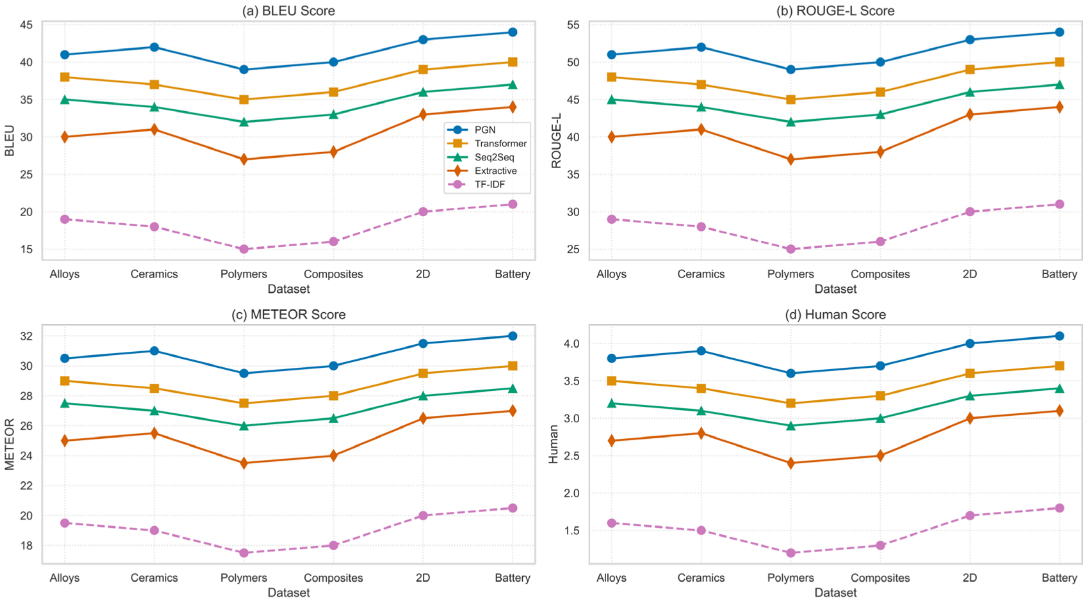

Figure 3. Quantitative results: (a) BLEU; (b) ROUGE-L; (c) METEOR; (d) Human evaluation

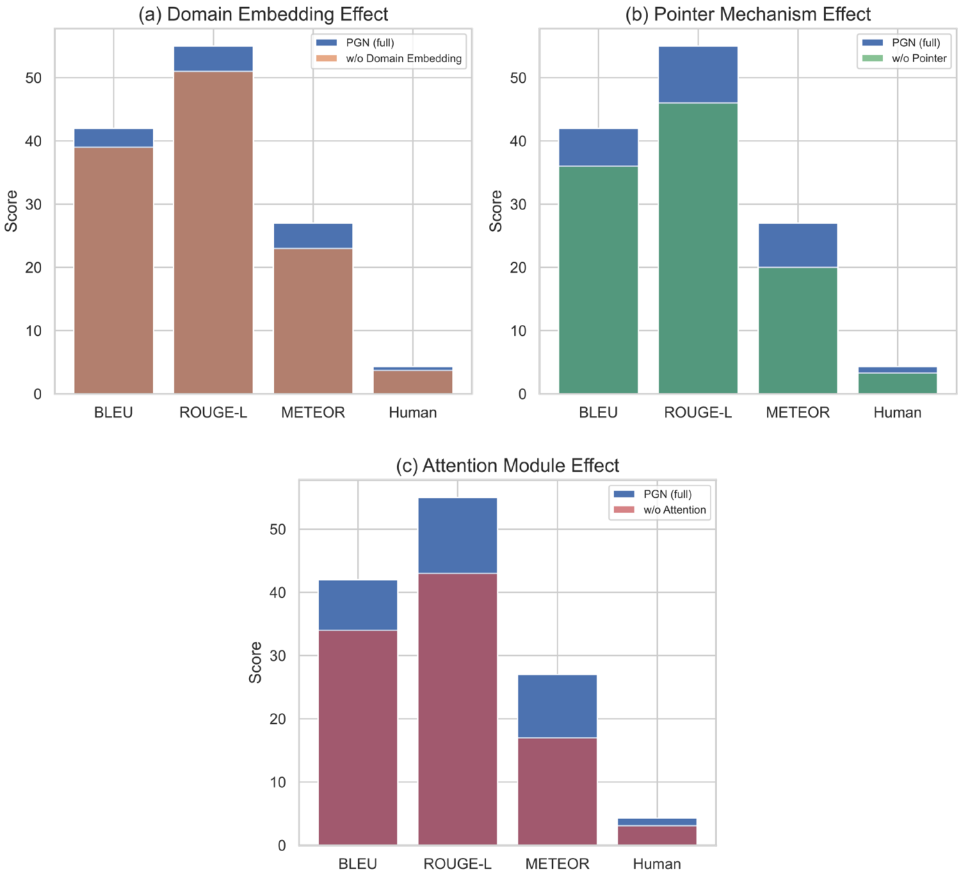

Figure 4. Ablation study: (a) Domain embedding; (b) Pointer mechanism; (c) Attention module

Ablation analysis will be conducted to determine the above results. As shown in Figure 4(a), domain embeddings are responsible for generating these automatic scores. When these embeddings are not used, the scores of various summaries, especially those containing chemical symbols and specific nomenclature, will significantly decrease. Figure 4(b) shows the effect of the pointer mechanism. The model with the pointer switches ablation shows a sudden drop in BLEU and ROUGE, indicating that it is directly copying from scientific vocabulary terms. Figure 4(c) shows that when the attention module is disabled, repetition and semantic drift increase. This leads to a decline in the structure and accuracy of the generated titles. Each architectural component is essential, independently enhancing the model's stability and accuracy.

The detailed distribution of title features used to test the model's expressiveness and adaptability is shown in Figure 5. Figure 5(a) shows that the titles generated by PGN are close to the reference distribution in terms of length, and both truncation (as in extraction methods) and verbosity (as in unregularized generative models) are reduced. As shown in Figure 5(b), by using domain-specific embeddings, the model can expand the range of output vocabulary, thereby increasing the statistical diversity of terms and reducing repetitive outputs. Figure 5(c) shows the distribution of novelty, with a higher proportion of n-grams never seen in PGN output; this indicates that it is not just replication, but more generalization.

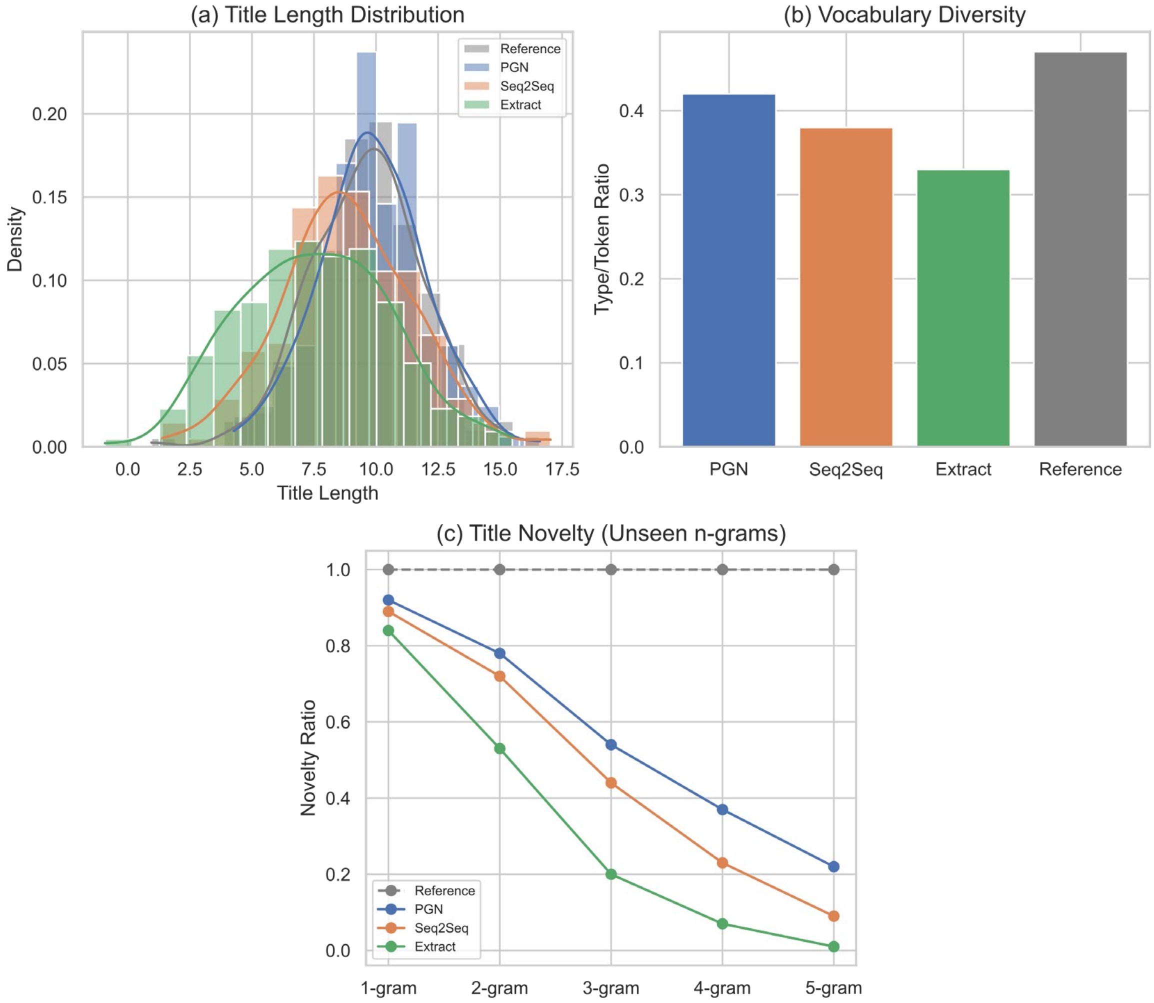

Figure 5. Title feature analysis: (a) Length; (b) Diversity; (c) Novelty

Qualitative Analysis of Titles

Qualitative analysis can provide a deeper understanding of the behavior and shortcomings of PGN models compared to traditional benchmarks and manually crafted titles. Particular attention is given to the challenges posed by scientific language and content compression. Figure 6 shows the generation results of various cases, categorized in order into best performance, rare word handling, and failures.

As shown in Figure 6(a), the best case is a PGN model that captures the research focus of the paper and summarizes it concisely and informatively in the title. "The enhancement of cycling stability of sulfur/graphene oxide cathodes in high-performance lithium-sulfur batteries" is a title generated by PGN, used to describe a technical abstract on the synthesis and electrochemical performance of a novel lithium-sulfur battery composite material. Including the necessary performance characteristics "cycling stability" and the key materials "sulfur/graphene oxide." Baseline outputs often overlook important property descriptors or only select one characteristic of the compound, making the discovery of such dual coverage very difficult.

As shown in Figure 6(b), the model is capable of handling technical terms well, whether they are new or old. PGN accurately identified and transcribed the compound names to provide an introduction to the summary of "nano-sized zirconium phosphate aerogel." Transformer and Seq2Seq methods often generalize too much, producing expressions like "new phosphate materials," or misinterpreting the structure connecting scientific terms, leading to fragmented outputs. Due to the innovations in pointer and domain embeddings achieving good results, human evaluators gave high scores (4.7/5) for the term accuracy and domain fidelity of PGN outputs.

Figure 6(c) is a typical error case, which includes multiple sentences and non-linear logic. In particular, the summaries that combine methods and results sometimes generate PGN titles that are structurally faithful but lack the subtle emphasis of human-written references. For example, the model output is "Solvothermal Synthesis and Application of $NiFe_2O_4$ Microspheres," omitting the modifier "hollow" and the specific application "water splitting." This omission is usually due to the model's reliance on significant term extraction, failing to fully grasp the distributed meaning in the abstract.

Figure 6(d) shows an uncommon failure case, namely the use of vague or interdisciplinary terms in the abstract. PGN uses default titles and materials in the text on photon bandgap tuning in hybrid perovskites, such as "Hybrid Perovskite Structure with Tunable Composition." This indicates that there are limitations in cross-domain reasoning, suggesting that the context-aware semantic model needs improvement.

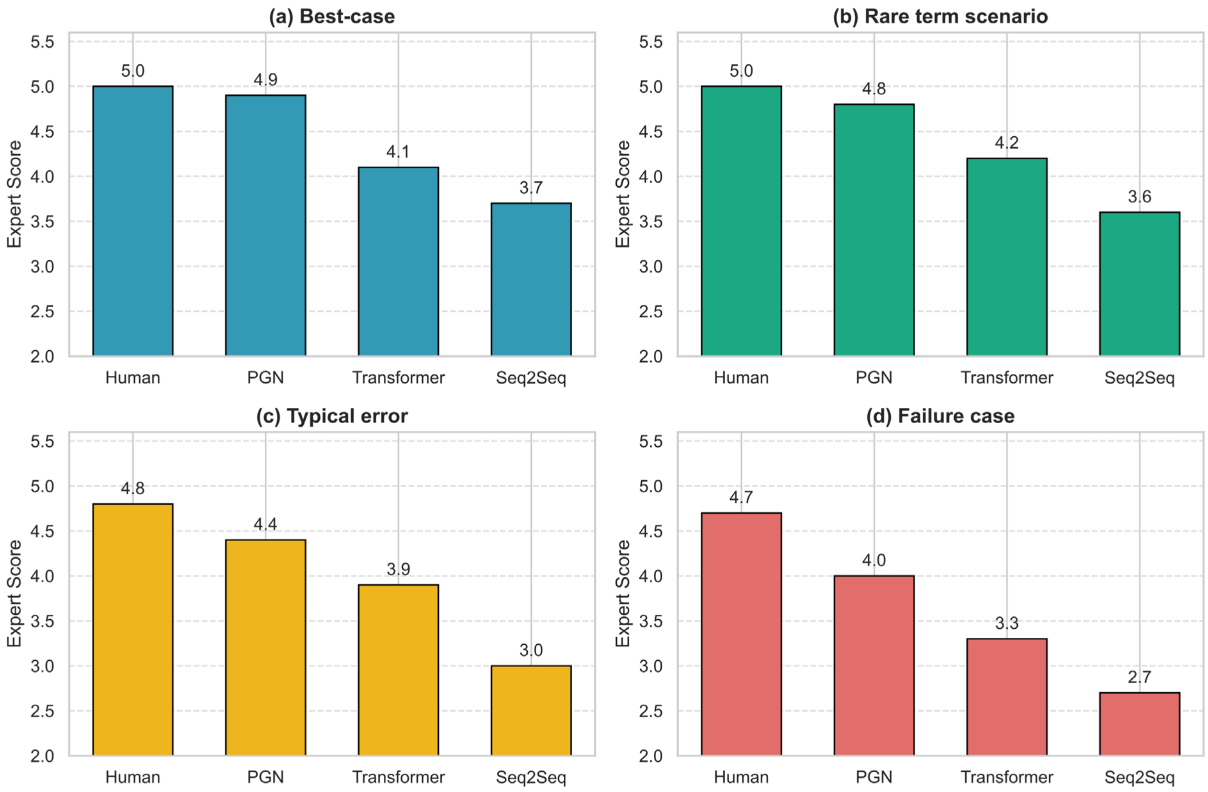

Figure 6. Qualitative cases: (a) Best-case; (b) Rare term; (c) Omission; (d) Failure

Case Studies: Successes and Limitations

In order to determine the strengths and weaknesses of the PGN model when handling different topics in materials science and document structure, additional case studies were conducted. Figure 7 details the three types of errors and how the model handles robustness to noisy inputs and rare labels.

In-depth research on energy storage literature indicates that PGN has a close relationship between input and output. Important concepts and technical terms are often mentioned in the abstract. For example, PGN accurately condenses and retains technical terms in the "interface engineering" study of lithium-rich cathode materials to reduce uncertainty. Figure 7(a) shows that for topics involving complex, multi-step methods, such as detailed synthesis plans for polymers, it is more likely for some phrases to be omitted or for the application context to be incorrect. "Attribute loss" and "context drift" account for more than 40% of all mismatches.

As shown in Figure 7(b), the domain-adaptive PGN also exhibits robustness to rare vocabulary. Compared to the Seq2Seq baseline and Transformer, the pointer mechanism of this model maintains a good recall rate and high precision in localization, with reference titles being specific chemical or alloy names. The evaluation of uncommon vocabulary also revealed a long-tail phenomenon: the frequency of using rare words in PGN slightly increased, sometimes affecting the accuracy of vocabulary formation and pairing, especially in non-English or mixed names.

Figure 7(c) shows the robustness of input noise through simulated experiments of synthesized perturbation summaries, such as replacing synonyms or adding interference sentences. PGN maintains more stable output quality compared to the standard architecture. Under controlled noise, the decline in BLEU and ROUGE scores is less pronounced; the model is more robust. In complex situations involving cross-domain data (e.g., biomimetic ceramics or multifunctional nanocomposites), the model's generalization ability still has issues, such as semantic ambiguity or loss of attention.

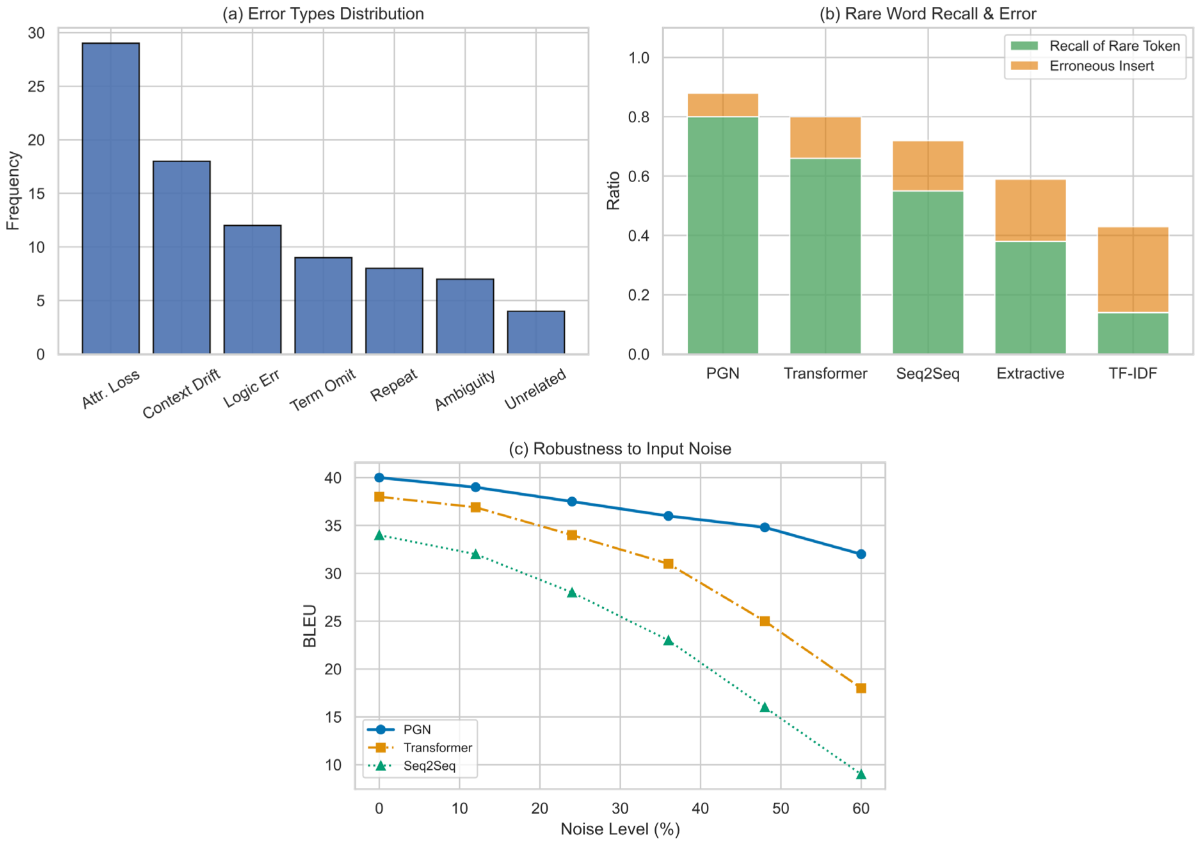

Figure 7. Error and robustness: (a) Error types; (b) Rare words; (c) Noise robustness

Quadrant analysis of selected materials (such as catalysts, energy storage, polymers, and structural composites) indicates that PGN is most prominent in areas with high term density and clear property-function relationships.

In interdisciplinary or interpretive summaries, it is less evident. Expanding domain-adaptive pre-training, enhancing context disambiguation modules, and integrating external knowledge bases to address current shortcomings in semantic accuracy and noise robustness will make the model more comprehensive.

Conclusion

This study investigates the application of Pointer Generator Networks (PGN) in the automatic title generation of materials science literature to meet the growing demands of the research field. Due to the complexity of the specific language in the field of materials science and the exponential growth in the number of scientific papers, the proposed PGN-based framework directly addresses the shortcomings of traditional neural text generation methods. A dynamic gated pointer mechanism was constructed using domain-adaptive word embeddings and a dual-stream encoder. This mechanism can effectively learn rare terms and complex combinatorial structures specific to a particular domain. Establish a model that can reliably convert complex scientific abstracts into concise, focused, and contextually appropriate titles. Also established new technical benchmarks for the field.

The experiments in this study have validated the initial research ideas. After a systematic comparison with well-known baseline models (such as sequence-to-sequence models and Transformers), the model has shown excellent performance on both objective metrics (such as BLEU, ROUGE-L, METEOR) and human evaluation metrics. The purpose of the ablation study is to investigate how each module functions. For example, adding domain-specific embeddings expands the coverage of terms, while improved pointer and coverage strategies can reduce semantic drift and phrase repetition. The model demonstrates good overall performance, capable of handling a wide range of topics and names in materials science, and utilizes a large-scale expert-annotated dataset.

In the future, the application of this research will go beyond scientific titles. From a methodological perspective, the framework can also be used for the automatic generation of data-rich materials in all other fields of science and technology, such as chemistry, physics, and life sciences. Intelligent academic indexing and semantic retrieval, optimization of digital library management, and the construction of knowledge graphs are part of these advancements, and these applications all benefit from the aforementioned progress. Future work will focus on introducing external knowledge bases, enhancing robustness to noisy or variable text data, and improving the model's cross-domain generalization capabilities. This study lays the foundation for next-generation information management systems in materials science and other fields by enhancing the automation and application levels of scientific content generation.

Author Contributions

Tamara Blagojević contributes to conceptualization, methodology, software, validation, analysis, investigation, data collection, draft preparation, manuscript editing, visualization, project administration, and funding acquisition. Dragan Vasiljević and Nina Radojičić contribute to conceptualization, methodology, software, validation. All authors have read and agreed with the manuscript before its submission and publication.

Funding

This research received no specific financial support from any funding agency.

Institutional Review Board Statement

Not applicable.

References

[1] Xie, S., Cheng, W., Nie, Z., Xing, J., Chen, X., Gao, L., ... & Zhang, R. (2024). Multidimensional attention domain adaptive method incorporating degradation prior for machine remaining useful life prediction. IEEE transactions on industrial informatics, 20(5), 7345-7356. https://doi.org/10.1109/TII.2024.3359455
[2] Xia, J., & Wang, H. (2023, August). A sequence-to-sequence approach with mixed pointers to topic segmentation and segment labeling. In Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (pp. 2683-2693). https://doi.org/10.1145/3580305.3599245

[3] Jiang, X., Wang, W., Tian, S., Wang, H., Lookman, T., & Su, Y. (2025). Applications of natural language processing and large language models in materials discovery. npj Computational Materials, 11(1), 79. https://doi.org/10.1038/s41524-025-01554-0

[4] Gupta, T., Zaki, M., Krishnan, N. A., & Mausam. (2022). MatSciBERT: A materials domain language model for text mining and information extraction. npj Computational Materials, 8(1), 102. https://doi.org/10.1038/s41524-022-00784-w

[5] Ofori-Boateng, R., Aceves-Martins, M., Wiratunga, N., & Moreno-Garcia, C. F. (2024). Towards the automation of systematic reviews using natural language processing, machine learning, and deep learning: a comprehensive review. Artificial intelligence review, 57(8), 200. https://doi.org/10.1007/s10462-024-10844-w

[6] Rehman, T., Sanyal, D. K., Chattopadhyay, S., Bhowmick, P. K., & Das, P. P. (2023). Generation of highlights from research papers using pointer-generator networks and SciBERT embeddings. IEEE Access, 11, 91358-91374. https://doi.org/10.1109/ACCESS.2023.3292300

[7] Zhou, S., Wang, N., Wang, L., Liu, H., & Zhang, R. (2022). CancerBERT: a cancer domain-specific language model for extracting breast cancer phenotypes from electronic health records. Journal of the American Medical Informatics Association, 29(7), 1208-1216. https://doi.org/10.1093/jamia/ocac040

[8] Fatima, N., Imran, A. S., Kastrati, Z., Daudpota, S. M., & Soomro, A. (2022). A systematic literature review on text generation using deep neural network models. IEEE Access, 10, 53490-53503. https://doi.org/10.1109/ACCESS.2022.3174108

[9] Sharma, G., Sharma, D., & Sasikumar, M. (2024). Summarizing long scientific documents through hierarchical structure extraction. Natural Language Processing Journal, 8, 100080. https://doi.org/10.1016/j.nlp.2024.100080

[10] Venugopal, V., & Olivetti, E. (2024). MatKG: An autonomously generated knowledge graph in Material Science. Scientific Data, 11(1), 217. https://doi.org/10.1038/s41597-024-03039-z

[11] Goyal, R., Kumar, P., & Singh, V. P. (2023). A Systematic survey on automated text generation tools and techniques: application, evaluation, and challenges. Multimedia Tools and Applications, 82(28), 43089-43144. https://doi.org/10.1007/s11042-023-15224-0

[12] Lin, W., Lin, Y., & Yang, Y. (2024, August). Data augmentation in medical materials science: A review. In 2024 7th International Conference on Pattern Recognition and Artificial Intelligence (PRAI) (pp. 844-851). IEEE. https://doi.org/10.1109/PRAI62207.2024.10826735

[13] Qi, Z., Chen, J., Wang, S., Liu, B., Zheng, H., & Wang, C. (2024, December). Optimizing multi-task learning for enhanced performance in large language models. In 2024 4th International Conference on Electronic Information Engineering and Computer Communication (EIECC) (pp. 1179-1183). IEEE. https://doi.org/10.1109/EIECC64539.2024.10929547

[14] Gupta, S., Mahmood, A., Shetty, P., Adeboye, A., & Ramprasad, R. (2024). Data extraction from polymer literature using large language models. Communications materials, 5(1), 269. https://doi.org/10.1038/s43246-024-00708-9

[15] Unni, R., Zhou, M., Wiecha, P. R., & Zheng, Y. (2024). Advancing materials science through next-generation machine learning. Current Opinion in Solid State and Materials Science, 30, 101157. https://doi.org/10.1016/j.cossms.2024.101157

[16] Ke, J., Wang, W., Chen, X., Gou, J., Gao, Y., & Jin, S. (2023). Medical entity recognition and knowledge map relationship analysis of Chinese EMRs based on improved BiLSTM-CRF. Computers and Electrical Engineering, 108, 108709. https://doi.org/10.1016/j.compeleceng.2023.108709

[17] Choudhary, K., DeCost, B., Chen, C., Jain, A., Tavazza, F., Cohn, R., ... & Wolverton, C. (2022). Recent advances and applications of deep learning methods in materials science. npj Computational Materials, 8(1), 59. https://doi.org/10.1038/s41524-022-00734-6

[18] Fatima, N., Daudpota, S. M., Kastrati, Z., Imran, A. S., Hassan, S., & Elmitwally, N. S. (2023). Improving news headline text generation quality through frequent POS-Tag patterns analysis. Engineering applications of artificial intelligence, 125, 106718. https://doi.org/10.1016/j.engappai.2023.106718

[19] Guha, S., Mullick, A., Agrawal, J., Ram, S., Ghui, S., Lee, S. C., ... & Goyal, P. (2021). MatScIE: An automated tool for the generation of databases of methods and parameters used in the computational materials science literature. Computational Materials Science, 192, 110325. https://doi.org/10.1016/j.commatsci.2021.110325

[20] Manda, P. (2025). Large Language Models in Bio-Ontology Research: A Review. Bioengineering, 12(11), 1260. https://doi.org/10.3390/bioengineering12111260

[21] Wang, J., Huang, J. X., Tu, X., Wang, J., Huang, A. J., Laskar, M. T. R., & Bhuiyan, A. (2024). Utilizing bert for information retrieval: Survey, applications, resources, and challenges. ACM Computing Surveys, 56(7), 1-33. https://doi.org/10.1145/3648471

[22] Kasanishi, T., Isonuma, M., Mori, J., & Sakata, I. (2023). SciReviewGen: A large-scale dataset for automatic literature review generation. Findings of the Association for Computational Linguistics: ACL 2023, 6695-6715. https://doi.org/10.18653/v1/2023.findings-acl.418

[23] Ye, Y., Ren, J., Wang, S., Wan, Y., Razzak, I., Hoex, B., ... & Zhang, W. (2024). Construction and application of materials knowledge graph in multidisciplinary materials science via large language model. Advances in Neural Information Processing Systems, 37, 56878-56897. https://doi.org/10.52202/079017-1812

[24] Sivan, D., Satheesh Kumar, K., Abdullah, A., Raj, V., Misnon, I. I., Ramakrishna, S., & Jose, R. (2024). Advances in materials informatics: a review. Journal of Materials Science, 59(7), 2602-2643. https://doi.org/10.1007/s10853-024-09379-w

[25] Montero-Parodi, J. J., Rodriguez-Sánchez, R., García, J. A., & Fdez-Valdivia, J. (2025). The Quality of a Scientific Manuscript Given by a Peer-Reviewed. Report in Three Dimensions: Accessibility, Contribution and Experimentation (AccConExp). Expert Systems, 42(10), e70131. https://doi.org/10.1111/exsy.70131