Exploring Deep Learning for Multi-Topic Policy Text Classification

Chengman Jin (iD)
Department of Big Data Analytics, Kyung Hee University, Seoul, Korea
E-mail: jinchengman@khu.ac.kr

Ha Eun Park* (iD)
Department of Big Data Analytics, Kyung Hee University, Seoul, Korea
E-mail: gracepark129@khu.ac.kr

Baek Jeong (iD)
Department of Big Data Analytics, Kyung Hee University, Seoul, Korea
E-mail: ylbaek@khu.ac.kr

Kyoung Jun Lee (iD)
Department of Big Data Analytics, Kyung Hee University, Seoul, Korea
E-mail: klee@khu.ac.kr

ABSTRACT

As the demand for cross-national policy comparisons intensifies, automating the classification of vast volumes of policy texts through deep learning presents a viable solution to alleviate manual labor and enhance policy matching efficiency. This study proposes an automated multi-topic policy classification system grounded in the policy perceptron framework, augmented by the RoBERTa-whole word masking model pretrained on domain-specific corpora. The study curated a dataset of policy documents, applied the model for multi-label classification, and rigorously evaluated its performance via comparative experiments against baseline models. The results present performance for our model in multi-topic scenarios, with robustness to imbalanced classes and lengthy texts. Furthermore, the authors benchmarked the approach against general-purpose large language models, revealing that this domain-tailored model achieves higher classification efficacy. These findings highlight the advantages of specialized deep learning architectures for policy text analysis, examining the conceptual and methodological implications of automating policy classification and offering practical insights for scalable policy organization and future research in automated governance tools.

Keywords: deep learning, natural language processing, text classification, artificial intelligence, automation, public policy classification model

1. INTRODUCTION

In today's increasingly complex and data-driven government decision-making environment, the importance of methodologies that can systematically structure and classify vast amounts of policy information is becoming more prominent. Policy documents are not only voluminous and diverse in subject matter but are also updated at a rapid pace, creating a pressing need for both government agencies and researchers to develop a robust framework to efficiently organize and analyze such unstructured text data. In response to this challenge, one of the leading international initiatives is the comparative agendas project (CAP). CAP provides a standardized hierarchical coding system that enables consistent classification and cross-national comparison of public policy (Baumgartner et al., 2019). By coding policy topics into major topics and minor topics, CAP makes it possible to conduct longitudinal and cross-country comparative analyses to examine how policy agendas have evolved over time, and which issues each political system has prioritized.

However, the current CAP classification system relies heavily on manual coding performed by trained personnel, which consumes considerable amount of time and resources. When processing large-scale texts collected from diverse sources such as legislation, administration, and the media, manual methods reveal clear limitations in scalability, hinder the timely provision of data, and may also compromise coding consistency. Due to these constraints, there has been growing interest in developing automation-based classification techniques that can preserve the CAP coding framework while ensuring accuracy and interpretability (Cavari et al., 2022; Qin & Huang, 2024).

Building on this problem awareness, the present study aims to address the following research question: How can an automatic classification model be developed for non-English policy documents while preserving the hierarchical structure of the CAP coding scheme? To address these questions, this study focuses on policy documents and proposes a novel classification framework that extends the existing Policy Perceptron model with deep learning techniques and integrates it with the RoBERTa-whole word masking (wwm) language model (LM). The objectives of this study are to 1) explore a novel classification framework that extends the existing Policy Perceptron model with deep learning techniques and integrates it with the RoBERTa-wwm LM, focusing on multi-topic classification of Chinese policy documents, 2) conduct exploratory comparative evaluations of the model's performance through comparative experiments, including benchmarks against general-purpose models, to explore its effectiveness in handling hierarchical and multi-label classification tasks, and 3) offer theoretical and practical insights by exploring domain-adapted deep learning applications in policy analysis and offering a potential scalable tool for automated policy text organization.

Most prior studies on policy text classification using deep learning models have been confined to specific policy domains, resulting in models with limited generalizability to broader policy contexts (e.g., Liang & Yi, 2021; Wang et al., 2022; Yu et al., 2022). Furthermore, attempts to apply multi-topic classification frameworks such as CAP beyond the context of the United States remain very limited and there remains a scarcity of studies focused on developing automated policy text classification models. To address this research gap, the present study focuses on policy documents issued by the State Council of China, extending and applying the Policy Perceptron model to the Chinese policy context. Building on the original Policy Perceptron framework, this study develops a scalable deep learning-based classification system that integrates the RoBERTa-wwm pretrained LM. The proposed approach maintains alignment with the CAP framework while capturing the thematic complexity of China's national policies, thereby providing a robust reference model for automatic multi-topic classification of policy texts.

From a technical implementation standpoint, this study introduces several novel contributions. First, it proposes a multistage classification architecture that aligns with CAP's dual classification scheme (major-minor topic), thereby enabling the model to capture hierarchical relationships across policy topics. Second, to capture the semantic structures of Chinese-language policy documents, this study employs a domain-adapted pretrained LM (RoBERTa-wwm), which significantly enhances classification accuracy in non-English contexts. Third, selected experiments incorporate a summarization-enhanced preprocessing method that extracts key information from lengthy policy texts before classification, thereby improving both processing efficiency and interpretability.

Beyond these technical innovations, this study makes both theoretical and practical contributions by adapting the Policy Perceptron model to the Chinese policy context and integrating it with the CAP framework. Theoretically, it advances the application of deep learning in structured policy analysis and provides empirical evidence that domain-adapted LM such as RoBERTa-wwm can perform effectively in complex, non-English policy environments.

While deep learning enables large-scale automation and consistency, it also raises questions about how algorithmic structures align with the intentions of human classification frameworks such as the CAP system. Practically, it delivers a scalable and reproducible approach for the automatic classification of Chinese policy texts, enhancing the efficiency and consistency of policy analysis and equipping researchers and policymakers with a concrete tool for large-scale, multi-topic policy monitoring and comparison.

2. LITERATURE BACKGROUND

2.1. Overview of Research on Natural Language Processing

Natural language processing (NLP), a subfield of artificial intelligence (AI), studies technologies that enable computers to understand and interpret human language in both speech and text forms (Goodfellow et al., 2016). The task of automatically classifying large volumes of documents into predefined categories is one of the core challenges in NLP and machine learning, and is generally referred to as text classification. The goal of text classification is to automatically assign text data to specific categories (Kim et al., 2020; Liu & Guo, 2019). Traditional machine learning approaches include Bayesian classifiers (Gong & Yu, 2010), support vector machines (Wei et al., 2013), decision trees (Johnson et al., 2002), hidden Markov models, gradient boosting trees, and random forests. While these methods achieve a certain level of performance, their effectiveness is limited by issues such as high dimensionality, data sparsity, and constraints on generalization (Minaee et al., 2021; Wu et al., 2020).

In recent years, deep learning has attracted growing attention as a suitable approach for tasks such as policy document classification, owing to its superior feature extraction capability, low-dimensional text embeddings, and enhanced semantic representation power. The introduction of word to vector (Word2Vec) (Mikolov et al., 2013) popularized word embedding techniques, while embeddings from language models (ELMo) (Peters et al., 2018) leveraged long short-term memory (LSTM)-based contextual embeddings to enable deeper semantic understanding. Subsequently, a variety of models such as bidirectional encoder representations from transformers (BERT) (Devlin et al., 2019), document to vector (Doc2Vec), and convolutional neural networks for sentence classification (TextCNN) have been applied to the automatic classification of policy documents. In addition, attempts have been made to improve text vectorization and classification performance by combining algorithms such as TextRank (Mihalcea & Tarau, 2004), Latent Dirichlet Allocation (LDA) (Blei et al., 2003), and convolutional neural network-bidirectional long short-term memory (CNN-BiLSTM)-Attention. Furthermore, Liang and Yi (2021) proposed a two-stage, three-branch reinforcement model designed to reduce the likelihood of misclassification in paragraph-level policy text classification.

Meanwhile, large-scale open-source LMs are regarded as so-called out-of-the-box models and have shown outstanding performance across a wide range of NLP tasks (Kirk et al., 2021; Song et al., 2022). These models are pretrained on massive general-domain corpora, equipping them with broad linguistic representation capabilities; however, they remain relatively limited in domain-specific knowledge. Differences in word distributions between general and specialized domains can lead to performance degradation, and a common approach to addressing this issue is domain-adaptive pretraining. This technique involves further training a LM, initially pretrained on general-domain corpora, with domain-specific data to supplement contextual knowledge, and prior research has reported significant performance improvements over general pretrained models (Beltagy et al., 2019; Gururangan et al., 2020; Lee et al., 2019). For example, exBERT (Tai et al., 2020), an extension of BERT tailored to the biomedical domain, achieved high accuracy while reducing training costs, and Gururangan et al. (2020) empirically presented the effectiveness of domain-adaptive pretraining across four domains and eight classification tasks. Similarly, Suzuki et al. (2023) developed a domain-specific LM for finance, showing superior performance in financial NLP tasks compared to general models. In line with this trajectory, the present study seeks to develop a deep learning-based model capable of automatically classifying policy documents, while simultaneously exploring methods to effectively capture semantic patterns specific to the policy domain.

Previous studies have primarily been conducted in limited experimental settings focused on general text classification or specific topic categories, with insufficient attention given to the multilayered structure of policy agendas or the applicability of the CAP framework (Liang & Yi, 2021). In addition, research on domain-adaptive LMs has largely centered on specific industries such as biomedicine and finance, with very few applications to the complex and time-sensitive domain of public policy (Gururangan et al., 2020; Suzuki et al., 2023).

In contrast, this study offers several novel contributions. First, it implements a multistage classification architecture grounded in the hierarchical structure of the CAP framework, enabling the model to capture layered relationships among policy topics and thereby ensuring both interpretability and precision in policy document classification. Second, as the first large-scale empirical study of policy documents issued by the State Council of China, it addresses the complexity and contextual richness of non-English policy data that prior research has largely overlooked. Third, by applying a RoBERTa-wwm–based domain-adaptive LM, it overcomes the limitations of general pretrained models and shows the value of domain-specific representation learning in the policy domain. Fourth, in selected experiments, it introduces a summarization-based preprocessing procedure that enhances classification accuracy by extracting essential information from lengthy documents. Collectively, these contributions not only advance NLP and policy text classification research but also provide specialized technical and empirical insights into the practical challenge of automating CAP-based policy classification.

2.2. Policy Text Classification and Deep Learning Techniques

The automation of policy classification provides policymakers and researchers with a foundation to track policy trends more quickly and accurately, while also enabling comparative analysis. With growing interest in this field, various studies have explored the potential of deep learning techniques for policy text classification. For example, Liang and Yi (2021) proposed an enhanced two-stage, three-branch framework for classifying policy paragraphs from two cities. Sheng et al. (2024) employed text analysis and the policy modeling consistency-autoencoder index model to evaluate policy documents, finding that while the textual content aligns with practical needs, improvements are needed in policy diversity, tool integration, target coverage, and incentive mechanisms.

Domain-specific approaches have also been developed. Zhang et al. (2020) proposed a BERT and multi-scale convolutional neural network (CNN)-based model for classifying science and technology policy, while Yu et al. (2022) introduced a BERT-based model for classifying policy-issuing institutions. In the energy sector, Wang et al. (2022) applied a CNN model to classify policies into three renewable energy-related topics, and in the education sector, Rao and Yang (2022) employed an attention mechanism focusing on titles and body texts to classify policies into four educational themes.

These prior studies provide important insights demonstrating the potential and applicability of policy classification automation; however, most have focused on a single policy domain, limiting the generalizability of their models (see Table 1 for more details) (Beltagy et al., 2019; Blei et al., 2003; Devlin et al., 2019; Gong & Yu, 2010; Gururangan et al., 2020; Johnson et al., 2002; Liang & Yi, 2021; Mihalcea & Tarau, 2004; Mikolov et al., 2013; Peters et al., 2018; Suzuki et al., 2023; Tai et al., 2020; Wei et al., 2013). Moreover, large-scale, multi-topic policy classification based on the CAP framework has been attempted only very sparingly outside of Korea. To address this research gap, the present study extends and applies the Policy Perceptron model to classify the vast and diverse set of Chinese policy documents. The Chinese language presents unique linguistic characteristics, such as the absence of spacing between words and the rich semantic content embedded in individual characters, which pose challenges for policy document classification when relying solely on English-based or multilingual models. Thus, a domain-specific pretrained model tailored to Chinese is required. To this end, the study employs the RoBERTa-wwm model (Cui et al., 2021), a pretrained LM deeply trained on a large-scale Chinese corpus. RoBERTa-wwm offers the advantage of more effectively capturing the contextual meaning and linguistic complexity of Chinese policy documents. Accordingly, this study designs and applies a deep learning-based model built on RoBERTa-wwm to automatically classify Chinese policy documents, with the goal of enhancing both the structural organization and the accessibility of policy information.

Traditional classification systems have historically been grounded in human interpretation, emphasizing conceptual reasoning, hierarchical organization, and expert judgment in defining categories. In contrast, automated classification systems rely on data-driven similarity measures and algorithmic inference to detect latent patterns across large text corpora. This fundamental difference underscores a key conceptual tension between interpretive understanding and computational efficiency. By positioning the proposed model within this broader theoretical landscape, the present study aims to bridge human-centered classification logic and deep learning-based automation, highlighting both their complementarities and inherent limitations.

Table 1. Summary of major prior studies on NLP-based policy document classification

Category	Key researchers and year	Techniques/models used	Main contributions and limitations
Traditional machine learning-based classification	Gong & Yu (2010); Wei et al. (2013); Johnson et al. (2002)	Naive Bayes, SVM, decision trees	They established the early foundation of text classification through traditional statistical algorithms; however, their generalizability was limited due to high-dimensional features and data sparsity
Deep learning-based classification models	Mikolov et al. (2013); Peters et al. (2018); Devlin et al. (2019)	Word2Vec, ELMo, BERT, TextCNN	They enabled contextual embeddings and semantic representations, significantly improving classification performance, but still face limitations in handling lengthy policy documents and hierarchical classification structures
Applications of hybrid algorithms	Mihalcea & Tarau (2004); Blei et al. (2003)	TextRank, LDA, CNN-BiLSTM-Attention	They improved text vectorization and classification accuracy by combining multiple algorithms, yet fell short in reflecting policy domain-specific structures
Policy document-specific models	Liang & Yi (2021)	Two-stage branch-based reinforcement model	They proposed models that segment policy documents at the paragraph level to enhance classification performance, but their application to internationally standardized frameworks such as CAP was not considered
Domain-adaptive language models	Beltagy et al. (2019); Gururangan et al. (2020); Tai et al. (2020); Suzuki et al. (2023)	BioBERT, exBERT, finance-specialized BERT, etc.	They achieved substantial performance gains using pretrained language models specialized for specific industries, but research applying such models to complex domains like public policy remains highly limited

NLP, natural language processing; SVM, support vector machine; Word2Vec, word to vector; ELMo, embeddings from language models; BERT, bidirectional encoder representations from transformers; TextCNN, convolutional neural networks for sentence classification; TextRank, graph-based ranking model for text processing; LDA, Latent Dirichlet Allocation; CNN-BiLSTM-Attention, convolutional neural network-bidirectional long short-term memory attention; CAP, comparative agendas project; BioBERT, bidirectional encoder representations from transformers for biomedical text mining; exBERT, extended bidirectional encoder representations from transformers.

3. DESIGN OF THE POLICY PERCEPTRON MODEL

The design choice for the automated multi-topic policy classification system represents one empirical approach to operationalizing policy categories, through the integration of the policy perceptron framework with the RoBERTa-wwm model pretrained on domain-specific Chinese policy corpora. This operationalization relies on defining policy topics as discrete, multi-label categories derived from thematic analysis of the curated dataset, where labels are assigned based on contextual embeddings that capture semantic nuances in policy language.

3.1. Data Collection

This study utilizes policy documents collected from the official policy document repository of the State Council of China. The selection of Chinese policy documents is motivated by several key factors. First, China generates an immense volume of policy outputs through the State Council, providing a rich, diverse dataset that exemplifies the challenges of large-scale, multi-topic policy classification in a non-English context. Second, there is a significant research gap in applying automated classification frameworks like CAP to non-Western and non-English policy environments, where linguistic complexities and hierarchical structures demand tailored deep learning approaches. By focusing on China, this study extends the applicability of the Policy Perceptron model, enabling cross-national comparisons and offering practical value for global policy analysis involving emerging powers. The period of analysis spans from January 10, 2000, to December 31, 2022, covering more than two decades that represent a critical juncture in China's rapid economic growth and significant social transformation. A total of 15,018 policy documents were collected and systematically organized during this period. The categorical values assigned on the website were directly adopted as major and subtopics.

The collected dataset encompasses a wide range of policy domains, administrative levels, and policymaking bodies, thereby providing a comprehensive foundation for reflecting the dynamics of policy change and development

during this period. According to the classification system employed by the State Council, these policy documents are categorized into 22 major topics (see Table 2), each of which is further subdivided into 121 subtopics (see Table 3). This classification framework offers a valuable structure for systematically organizing, categorizing, and analyzing policy documents.

All policy documents published on the State Council's official website over a period of approximately twenty-two years were included in the analysis. After removing duplicate entries and handling missing values, subtopics with fewer than ten documents were excluded from the analysis to enhance the reliability of model training. The subtopic categories were directly adopted from the classifications provided on the website. As a result of preprocessing, a total of 22 major topics, 107 subtopics, and 14,652 policy

documents were used in the final analysis.

3.2. Summary Generation

Policy-related documents are widely used as important data sources across various research fields. However, such documents are structurally complex and linguistically sophisticated, often characterized by considerable length and broadly dispersed information. When these documents are used for model training without preprocessing, there is a risk of information loss or performance degradation. This is because, in processing long texts, the model may fail to capture the overall context or to extract the key information effectively. Accordingly, reducing the input length of policy documents while preserving information density as much as possible has emerged as a critical challenge.

Table 2. Major topics of Chinese policy

No.	Major topic
1	Organization of the State Council
2	Comprehensive government affairs
3	National economic management and state-owned assets supervision
4	Finance and audit
5	Land resources and energy
6	Agriculture, forestry, water conservancy
7	Industry and transportation
8	Trade, customs, tourism
9	Market supervision and safety production supervision
10	Urban and rural construction and environmental protection
11	Technology and education
12	Culture, radio and television, press and publication
13	Health and sports
14	Population and family planning, women and children's work
15	Labor, personnel and supervision
16	Public security, security and justice
17	Civil affairs, poverty alleviation and disaster relief
18	Nationality and religion
19	External affairs
20	Work for overseas Chinese from Hong Kong, Macao and Taiwan
21	National defense
22	Other

Table 3. Examples of subtopics in Chinese policy

No.	Major topic	Subtopic
1	Organization of the State Council	101. State Council
		102. General office of the State Council
		103. Constituent departments of the State Council
		104. Ad hoc agencies directly under the State Council
		105. Agencies directly under the State Council
		106. Offices of the State Council
		107. Institutions directly under the State Council
		108. State bureaus managed by ministries and commissions of the State Council
		109. Deliberative and coordinating bodies of the State Council
		110. Other
2	Comprehensive government affairs	201. Public affairs
		202. Government supervision
		203. Emergency management
		204. E-government
		205. Secretarial work
		206. Confidentiality
		207. Letters and visits
		208. Counsellor, literature and history
		209. Other

Summary generation has been proposed as a potential solution to this problem. Text summarization refers to the process of transforming the original text into a more concise form using NLP techniques, while preserving key information. Summarization methods can be broadly categorized into extractive summarization and abstractive summarization (Cajueiro et al., 2023). Extractive summarization constructs a summary by selecting important sentences directly from the source text. Traditional extractive approaches include LDA-based summarization models and extractive summarization models built on the TextRank algorithm (Blei et al., 2003; Mihalcea & Tarau, 2004), both of which have been widely used to identify key sentences and generate summaries.

In contrast, abstractive summarization reconstructs a summary by generating new sentences using LMs, rather than selecting existing ones as in extractive summarization. Most current techniques in this area rely on deep learning-based sequence-to-sequence text generation approaches, and more recently, a variety of natural language generation models built on the transformer architecture (Vaswani et al., 2017) have been introduced. Representative models include Facebook's BART (Lewis et al., 2020), Google's T5 (Raffel et al., 2020), and pre-training with extracted gap-sentences for abstractive summarization (PEGASUS) (Zhang et al., 2020). PEGASUS is a transformer-based language generation model developed by Google, which employs a novel pretraining objective of removing key sentences from large-scale text data and learning to reconstruct them, thereby acquiring a generalized understanding of language and context. Prior re-search has shown that PEGASUS performs well compared to many existing models across a variety of text summarization benchmark datasets (Zhang et al., 2020).

Accordingly, this study adopts the PEGASUS model, currently one of the best-performing models, for the summarization stage of the preprocessing module. To enhance adaptation to Chinese policy documents, the authors additionally employ the Randeng series of the Fengshenbang model (Zhang et al., 2023), which is a PEGASUS-based model trained for Chinese-language scenarios.

3.3. Policy Perceptron Model

The dataset constructed in this study consists of 22 major topics and 107 subtopics. If the entire classification task were performed using a single model, the model would need to learn all 107 categories, which could negatively impact both training efficiency and performance due to severe data imbalance issues. To effectively address this problem, the present study adopted the Policy Perceptron model at the design stage, a model that has shown performance in the field of policy multi-class classification.

However, the original Policy Perceptron model was built on Korean BERT, making it unsuitable for the classification of Chinese policy texts. To overcome this linguistic limitation, the present study adopts the RoBERTa-wwm model specialized for Chinese policy documents (Cui et al., 2021). This model is based on the WWM strategy of Chinese BERT, which enhances language understanding by masking entire words in accordance with traditional word segmentation methods. RoBERTa builds on the original BERT architecture but incorporates several
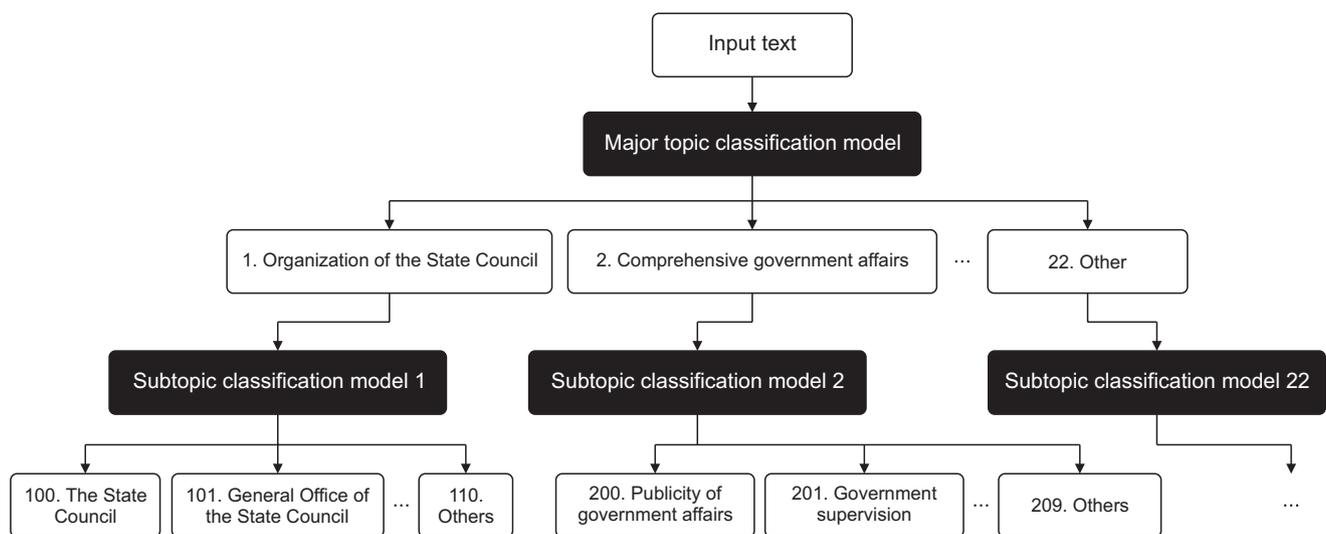

Fig. 1. Policy perceptron model.

refinements to improve performance: expanding batch size and sequence length to leverage larger datasets, removing the next-sentence prediction task, and introducing a dynamic masking strategy in the masked language modeling process. In this study, the authors also follow prior research in applying the same parameters (learning rate 5×10^{-5}, batch size 32, epoch 3).

Fig. 1 provides a visual representation of the overall architecture of the policy classification model proposed in this study. The structure consists of a major topic classification model and subtopic classification models. Both models are trained on RoBERTa-wwm. The policy summary text is first processed by the major topic classification model, which selects the most probable topic among the 22 major topics. Based on the selected major topic, the corresponding subtopic classification model is then activated, and this model ultimately predicts the subtopic with the highest probability from the input text.

The process of the policy classification model is divided into three main stages. In the first stage, policy documents are collected and subjected to a series of preprocessing tasks, such as text cleaning, stopword removal, and duplicate elimination, ensuring the data are ready for analysis. In the second stage, a summarization algorithm is applied to generate summaries that optimize input length while preserving key information from the long and complex source texts. Specifically, the PEGA-SUS model is employed to more accurately capture the semantic context of policy documents. Finally, in the third stage, the Policy Perceptron model is used to sequentially classify documents into major topics and then subtopics. Each stage is interconnected and has been carefully designed and optimized to maximize the model's overall performance and classification accuracy. This hierarchical and integrated approach provides a robust foundation for effectively processing and analyzing high-dimensional text data such as policy documents. Fig. 2 illustrates the step-by-step processing workflow of the overall model proposed in this study.

4. EXPERIMENTAL DESIGN AND RESULTS

4.1. Experimental Design

The dataset constructed in this study comprises a total of 14,652 policy documents. To ensure randomness and representativeness, data were randomly sampled across each classification label. For dividing the dataset into training and evaluation sets, the authors applied the commonly used 8:2 ratio: 80% of the data were used for model training, while the remaining 20% were set aside as an independent test set. This test set is intended for comparing and evaluating the performance of different models. To comprehensively and rigorously assess model performance, this study adopts four representative evaluation metrics: accuracy, precision, recall, and F1-score. By employing these metrics in combination, the proposed model's performance can be analyzed and compared from multiple perspectives.

4.2. Experimental Results

In the model design of this study, particular attention was given to the summary generation stage, with a detailed examination of both the theoretical background and practical applications of various summarization models. To more thoroughly analyze the impact of summarization quality on the final classification performance, three different summarization approaches were compared.

First, a baseline model without any summarization was established. Second, TextRank, a widely used extractive summarization model, was selected. This model represents text in a graph structure and generates summaries by extracting the most representative information from the original text. Finally, PEGASUS, a recently prominent abstractive summarization model, was included for comparison. Unlike extractive approaches, PEGASUS leverages semantic associations and transfer learning to generate new expressions rather than relying solely on words present in the source text, thereby conveying the deeper meaning of the original text in a more refined manner.

In addition, this study designed an additional experiment to compare the performance differences between

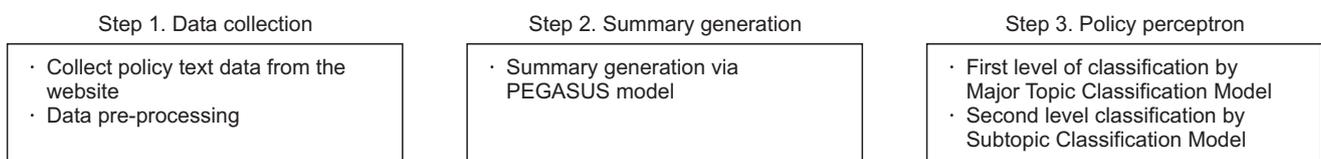| Step 1. Data collection	Step 2. Summary generation	Step 3. Policy perceptron
· Collect policy text data from the website · Data pre-processing	· Summary generation via PEGASUS model	· First level of classification by Major Topic Classification Model · Second level classification by Subtopic Classification Model

Fig. 2. Overall process of the policy perceptron model.

the Policy Perceptron model and a standalone classification model. In this comparative experiment, a RoBERTa-based standalone classification model was established as the baseline, and the performance gap between the two models was analyzed using various evaluation metrics (see Table 4). Through this comparative analysis, the performance of the Policy Perceptron model can be evaluated more precisely, while also providing clearer insights into the model's strengths and limitations.

Careful observation and analysis of the experimental results yielded several encouraging conclusions. Specifically, the policy classification system developed by combining the PEGASUS summarization model with the Policy Perceptron framework performed well compared

Table 4. Model comparison results

Overall performance		Macro average		
Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score
Non-summary+RoBERTa	60.80	45.12	42.48	41.62
TextRank+RoBERTa	59.40	46.84	45.18	44.10
PEGASUS+RoBERTa	67.42	51.34	50.96	49.37
Non-summary+Policy Perceptron	66.46	59.87	53.03	53.73
TextRank+Policy Perceptron	66.70	62.82	55.34	56.51
PEGASUS+Policy Perceptron	74.82	65.85	63.49	62.59

Table 5. Example of correct text classification by ChatGPT

Question by the user	"Central Bank: The policy stipulates that the upper limit on the interest rate for commercial housing loans and personal housing loans in Shanghai shall be adjusted upward by more than 20 basis points from the official market interest rate during the loan period, and the upper limit on the interest rate for commercial housing loans and personal housing loans in 2nd-tier housing shall be implemented in accordance with the current regulations." Among the following 22 policy categories, which number does this policy most likely belong to? Please select the one with the highest likelihood. 1. Labor, personnel, and inspection 2. Public safety, security, and affairs 3. Civil affairs, disaster relief, and disaster management 4. Nationality and religion 5. Diplomatic affairs 6. Affairs related to Hong Kong, Macau, and Taiwan 7. National defense 8. Other 9. Health and sports 10. Population and family planning, women and children 11. Science, technology, and education 12. Culture, broadcasting, media, and publishing 13. Urban, rural construction, and environmental protection 14. Commerce, taxation, and industry and commerce 15. Market regulation and work safety regulation 16. Industry and transportation 17. Agriculture, forestry, and water resources 18. Land, resources, and energy 19. Finance, banking, and audit 20. Comprehensive government affairs 21. National economic management and state-owned assets supervision 22. Organizational affairs of state organs Please answer in the following format: "Most likely:"
Response from ChatGPT	Most likely: 19. Finance, banking, and audit
Correct answer	Most likely: 19. Finance, banking, and audit

to other comparative models across multiple evaluation metrics. These findings provide empirical support for the study's hypothesis and serve as important evidence that the proposed PEGASUS+Policy Perceptron model is a highly efficient and precise approach for policy text classification tasks.

4.3. Exploratory Experiment with ChatGPT

In recent years, the advancement of LMs has led to the emergence of various chatbots. Representative examples include language model for dialogue applications, a Transformer-based chatbot (Thoppilan et al., 2022), and Sparrow, which was trained using reinforcement learning from human feedback (RLHF) (Glaese et al., 2022). Following these initiatives, ChatGPT was developed. ChatGPT is similar to InstructGPT (Ouyang et al., 2022) in its use of RLHF-based training methods, but differs in its data collection environment and other configurations, and has

shown impressive performance across diverse domains. The advent of ChatGPT marked a turning point in the performance of generative LMs, and this technological progress suggests potential utility in handling complex texts with embedded topical relevance, such as policy documents. Against this backdrop, the present study empirically examines how effectively ChatGPT-3.5 and GPT-4 can perform multi-topic classification inference tasks for policy documents under zero-shot and few-shot conditions.

In this study, pre-edited question templates were employed, into which specific policy summaries were inserted, thereby prompting ChatGPT to infer the most appropriate category among the 22 major topics. Concrete examples of implementation are illustrated in Table 5 (a correct text classification case) and Table 6 (an incorrect text classification case).

For GPT-3.5, performance was evaluated on the en-

Table 6. Example of incorrect text classification by ChatGPT

Question by the user	"In 2019, regarding the notice on the implementation of outpatient medical expense reimbursement under the National Health Insurance (including contacting medical insurance departments in other regions), the National Health Insurance Bureau stated that it would make every effort to promote the settlement of outpatient medical expenses in other places. With the insurance card, patients can receive treatment at designated hospitals nationwide and pay the expenses immediately, allowing elderly people and the general population to conveniently access medical services." Which of the following 22 policy categories does this policy most likely fall under? Please select the number with the highest likelihood. 1. Labor, personnel, and inspection 2. Public safety, security, and affairs 3. Civil affairs, disaster relief, and disaster management 4. Nationality and religion 5. Diplomatic affairs 6. Affairs related to Hong Kong, Macau, and Taiwan 7. National defense 8. Other 9. Health and sports 10. Population and family planning, women and children 11. Science, technology, and education 12. Culture, broadcasting, media, and publishing 13. Urban, rural construction, and environmental protection 14. Commerce, taxation, and tourism 15. Market regulation and work safety regulation 16. Industry and transportation 17. Agriculture, forestry, and water resources 18. Land, resources, and energy 19. Finance, banking, and audit 20. Comprehensive government affairs 21. National economic management and state-owned assets supervision 22. Organizational affairs of state organs Please answer in the following format: "Most likely:"
Response from ChatGPT	Most likely: 9. Health and sports
Correct answer	Most likely: 1. Labor, personnel, and inspection

tire test set, while for GPT-4, due to server-side performance constraints, experiments were conducted on 360 samples, corresponding to 12% of the test set. The accuracy achieved by the proposed model and by ChatGPT in major topic classification is presented in Tables 7 and 8, respectively.

As shown in the Tables 7 and 8, ChatGPT-3.5 achieved a classification accuracy of 62.81% under the zero-shot condition, while GPT-4 achieved 66.39% under the few-shot condition. In contrast, the model proposed in this study attained accuracies of 82.67% and 82.78%, respectively, demonstrating higher specialization and classification performance in domain-specific tasks such as policy topic classification. Nevertheless, given that ChatGPT shows potential utility as a supplementary tool in situations with limited sample data, its role as a complementary resource warrants consideration in future applications.

5. DISCUSSION AND IMPLICATIONS

This study provides empirical evidence on how domain-adapted models interact with structured classification systems, offering insight into the ways in which domain-oriented training can shape model performance in multi-label policy text analysis. This study offers the following theoretical contributions. First, it extends the Policy Perceptron framework to Chinese policy documents from the State Council, thereby enriching theoretical insights into the adaptability of automated classification models across diverse national and linguistic environments. This extension addresses the limitations of prior models that have primarily been tested in different language settings, demonstrating how the model on region-specific corpora can enhance cross-cultural applicability and reduce biases inherent in global AI systems. By validating the model's performance on large-scale, hierarchical policy texts, the study contributes to ongoing discussions on generalizing perceptron dynamics to real-world policy learning scenarios (Liang & Yi, 2021). This shows how domain-specific model can bridge gaps in cross-cultural policy research, paving the way for more inclusive automated tools in global governance studies (Wang et al., 2022; Yu et al., 2022).

Second, through the enhancement of RoBERTa-wwm (Cui et al., 2021) and comparative evaluation against general LLMs like ChatGPT, it contributes to theories of domain adaptation in NLP, emphasizing the superior efficacy of specialized pretrained models in capturing contextual semantic features unique to government policy texts, thus informing future developments in multilingual and sector-specific AI applications. This comparison reveals that models like RoBERTa-wwm perform well compared to general-purpose LLMs in text classification within domain-specific corpora, particularly in handling unstructured policy data. By exploring automated text classification with political corpora, it underscores the continued relevance of models like RoBERTa-wwm in the era of advanced LLMs, offering theoretical guidance for adapting NLP tools to high-stakes environments like governance analysis.

Third, unlike prior studies confined to single policy domains (Wang et al., 2022; Yu et al., 2022), this work

Table 7. Comparison between policy perceptron and GPT-3.5

Overall performance		Macro average		
Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score
Non-summary+RoBERTa	77.28	69.78	65.01	65.83
GPT-3.5	62.81	47.13	52.53	44.10
PEGASUS+Policy Perceptron	82.67	74.80	67.85	70.29

Table 8. Comparison between policy perceptron and GPT-4

Overall performance		Macro average		
Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score
Non-summary+RoBERTa	65.00	66.17	57.94	56.97
GPT-4	66.39	64.06	65.19	61.04
PEGASUS+Policy Perceptron	82.78	82.59	79.07	79.81

contributes to scalable classification by validating deep learning approaches for multi-topic analysis across vast, diverse policy texts. It extends the conceptual framework of automated agenda-setting research (Baumgartner et al., 2019) by showing how attention mechanisms in models like RoBERTa-wwm (Cui et al., 2021) can disentangle overlapping policy themes (e.g., economic vs. environmental policies) in real-world datasets. This advances theoretical discussions on model robustness, highlighting factors such as data imbalance and topic granularity that influence generalizability, and offers a blueprint for hybrid human-AI classification pipelines in policy studies.

From a practical perspective, this study introduces a scalable, deep learning-based model leveraging RoBERTa-wwm to automatically classify Chinese policy documents across diverse thematic areas, significantly reducing the time, labor, and inconsistency associated with manual classification. By enabling rapid and systematic organization of large-scale government datasets, the model empowers policymakers, analysts, and researchers to efficiently monitor and compare national policy agendas in near real-time. Practically, this enhances the accessibility and usability of policy data, which is often underutilized due to classification bottlenecks. For instance, government agencies can use this tool to track policy shifts, evaluate policy implementation, and identify emerging priorities, such as environmental or economic initiatives, with greater precision and speed. This capability supports evidence-based decision-making and facilitates timely responses to dynamic policy challenges, ultimately improving governance efficiency in data-intensive environments.

The proposed Policy Perceptron model offers a transferable framework for classifying policy documents in non-English and linguistically complex contexts (Baumgartner et al., 2019; Cui et al., 2021). This practical advancement allows researchers and policymakers to extend automated classification to other languages and national contexts, fostering cross-national comparative studies of policy agendas. For example, the framework can be adapted to analyze policy documents in other languages. By providing a reproducible and adaptable methodology, the model enables global research institutions and international organizations to standardize policy data analysis, uncover cross-country patterns in agenda-setting, and promote collaborative policy evaluation. This scalability enhances the potential for harmonized global policy research, supporting initiatives like sustainable development goal tracking or comparative governance studies.

In addition, analysis of misclassifications provides deeper insights into the ambiguities and overlaps inherent in policy topics. For instance, errors occurred at the boundaries between environmental and economic policies, where texts discussing sustainable development were inconsistently labeled due to shared terminologies like "green growth" that blur categorical distinctions. Similarly, overlaps in social and health policy domains led to misclassifications in documents addressing public welfare during pandemics, highlighting how automation can sometimes exacerbate interpretive challenges rather than resolve them. These patterns invite reflection on the behavior of classification boundaries under automation, suggesting that while specialized models enhance efficiency, they may reinforce rigid categorizations that fail to capture the fluid, multifaceted nature of real-world policy discourse, thereby underscoring the need for hybrid human-AI approaches in future implementations.

6. CONCLUSION AND FUTURE RESEARCH

This study proposes empirical evidence of automated policy classification tailored to the Chinese policy context. The Policy Perceptron framework was integrated with RoBERTa-wwm, a LM pretrained on a large-scale Chinese corpus. The proposed model supports multi-topic classification of national policy documents within the structure of the CAP framework, addressing the scalability and consistency limitations inherent in manual coding approaches. In addition, considering the lengthy and complex nature of policy documents, the PEGASUS summarization model was introduced to condense inputs while preserving essential content. This enhances the efficiency of the preprocessing stage and contributes to input optimization for the model. Building on this foundation, RoBERTa-wwm was finely tuned within the Policy Perceptron architecture to more accurately capture the linguistic and contextual characteristics of Chinese policy texts and improve classification accuracy. Experimental results presented that the proposed model performs effectively in multi-class classification tasks, providing empirical validation of the approach. Overall, this study contributes to the theoretical advancement of NLP applications in specialized domains, while also offering a practical contribution by developing a scalable policy analysis tool applicable to non-English policy environments.

However, this study also has several limitations. First, since the model was developed based on policy data from the State Council of China, there may be constraints in generalizing its applicability to other levels of government

or different countries. Second, in some subtopics, the low density of data may lead to topic imbalance, which could result in relatively lower classification accuracy for certain categories. Third, potential biases may be introduced by the dataset's focus on Chinese policies, which may not fully account for cross-cultural variations or evolving policy discourses in other regions. The findings of this study are context-bound and may not readily generalize beyond the specific dataset and models examined. To address these limitations, future research may consider the following directions. First, by collecting and integrating policy data from various countries and local governments, the generalizability of the model and its applicability across nations can be further validated. Second, to mitigate the issue of data imbalance, techniques such as minority class augmentation (e.g., data synthesis, oversampling) or cost-sensitive learning could be employed to achieve more refined improvements in classification performance. Third, beyond the CAP framework, applying or comparing multiple policy classification schemes in parallel could help develop the model into a more flexible policy analysis platform capable of accommodating diverse classification systems. In line with the current evidence, the conclusion highlights that domain-adapted models show promising potential; however, further comparative studies are needed to more fully establish their performance and applicability across diverse contexts.

While our domain-tailored model demonstrates improved performance in multi-topic policy classification—surpassing well-performing baseline and general-purpose large LMs in both accuracy and robustness—it also reveals several open issues that warrant further investigation in automated policy analysis. For instance, the algorithmic modeling of policy categories raises questions about their stability, as predefined labels may not consistently capture the fluid, context-dependent nature of policy discourse, potentially leading to brittle categorizations under varying linguistic or cultural shifts. Moreover, transitioning from human interpretive classification to automation involves trade-offs: Gains in scalability and efficiency are evident, yet losses may occur in nuanced understanding, such as overlooking implicit policy intents or subtle rhetorical elements that human experts intuitively discern, which could inadvertently standardize diverse policy narratives into overly rigid frameworks. To mitigate these challenges, hybrid approaches that integrate algorithmic inference with expert review could enhance reliability, fostering iterative refinements that balance computational speed with interpretive depth and ultimately advancing more resilient

tools for automated governance.

In conclusion, this study makes a substantive contribution to the advancement of automatic classification techniques for Chinese policy documents. The proposed approach enables faster and more accurate retrieval of policy information, thereby serving as an important support tool for policymakers and researchers in tracking policy changes, analyzing shifts in policy priorities, and understanding long-term policy trends.

CONFLICTS OF INTEREST

No potential conflict of interest relevant to this article was reported.

REFERENCES

Baumgartner, F. R., Breunig, C., & Grossman, E. (2019). The comparative agendas project: Intellectual roots and current developments. In F. R. Baumgartner, C. Breunig, E. Grossman, F. R. Baumgartner, C. Breunig, & E. Grossman (Eds.), *Comparative policy agendas: Theory, tools, data* (pp. 3-16). Oxford University Press.

Beltagy, I., Lo, K., & Cohan, A. (2019, November). SciBERT: A pretrained language model for scientific text. In K. Inui, J. Jiang, V. Ng, & X. Wan (Eds.), *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)* (pp. 3615-3620). Association for Computational Linguistics.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3, 993-1022.

Cajueiro, D. O., Nery, A. G., Tavares, I., Melo, M. K. D., dos Reis, S. A., Weigang, L., & Celestino, V. R. R. (2023). *A comprehensive review of automatic text summarization techniques: Method, data, evaluation and coding*. https://arxiv.org/abs/2301.03403

Cavari, A., Rosenthal, M., & Shpaizman, I. (2022). Introducing a new dataset: The Israeli policy agendas project. *Israel Studies Review*, 37(1), 1-30. https://doi.org/10.3167/isr.2022.370102

Cui, Y., Che, W., Liu, T., Qin, B., & Yang, Z. (2021). Pre-training with whole word masking for Chinese BERT. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 29, 3504-3514. https://doi.org/10.1109/TASLP.2021.3124365

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019, June). BERT: Pre-training of deep bidirectional transformers for

language understanding. In J. Burstein, C. Doran, & T. Solorio (Eds.), *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)* (pp. 4171-4186). Association for Computational Linguistics.

Glaese, A., McAleese, N., Trębacz, M., Aslanides, J., Firoiu, V., Ewalds, T., Rauh, M., Weidinger, L., Chadwick, M., Thacker, P., Campbell-Gillingham, L., Uesato, J., Huang, P. S., Comanescu, R., Yang, F., See, A., Dathathri, S., Greig, R., Chen, C., … Irving, G. (2022). *Improving alignment of dialogue agents via targeted human judgements*. https://arxiv.org/abs/2209.14375

Gong, Z., & Yu, T. (2010, November 7-9). Chinese web text classification system model based on Naive Bayes. In S. Walcheski (Ed.), *2010 International Conference on E-Product E-Service and E-Entertainment* (pp. 1-4). IEEE.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. The MIT Press.

Gururangan, S., Marasović, A., Swayamdipta, S., Lo, K., Beltagy, I., Downey, D., & Smith, N. A. (2020, July). Don't stop pretraining: Adapt language models to domains and tasks. In D. Jurafsky, J. Chai, N. Schluter, & J. Tetreault (Eds.), *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 8342–8360). Association for Computational Linguistics.

Johnson, D. E., Oles, F. J., Zhang, T., & Goetz, T. (2002). A decision-tree-based symbolic rule induction system for text categorization. *IBM Systems Journal*, 41(3), 428-437. https://doi.org/10.1147/sj.413.0428

Kim, J., Jang, S., Park, E., & Choi, S. (2020). Text classification using capsules. *Neurocomputing*, 376, 214-221. https://doi.org/10.1016/j.neucom.2019.10.033

Kirk, H. R., Jun, Y., Iqbal, H., Benussi, E., Volpin, F., Dreyer, F. A., Shtedritski, A., & Asano, Y. M. (2021). *Bias out-of-the-box: An empirical analysis of intersectional occupational biases in popular generative language models*. https://arxiv.org/abs/2102.04130

Lee, J., Yoon, W., Kim, S., Kim, D., Kim, S., So, C. H., & Kang, J. (2019). BioBERT: A pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics*, 36(4), 1234-1240. https://doi.org/10.1093/bioinformatics/btz682

Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., Stoyanov, V., & Zettlemoyer, L. (2020, July). BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In D. Jurafsky, J. Chai, N. Schluter, & J. Tetreault (Eds.), *Proceedings of the 58th Annual Meeting of the Association for*

Computational Linguistics (pp. 7871–7880). Association for Computational Linguistics.

Liang, D., & Yi, B. (2021). Two-stage three-way enhanced technique for ensemble learning in inclusive policy text classification. *Information Sciences*, 547, 271-288. https://doi.org/10.1016/j.ins.2020.08.051

Liu, G., & Guo, J. (2019). Bidirectional LSTM with attention mechanism and convolutional layer for text classification. *Neurocomputing*, 337, 325-338. https://doi.org/10.1016/j.neucom.2019.01.078

Mihalcea, R., & Tarau, P. (2004, July). TextRank: Bringing order into text. In D. Lin & D. Wu (Eds.), *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing* (pp. 404–411). Association for Computational Linguistics.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). *Efficient estimation of word representations in vector space*. https://arxiv.org/abs/1301.3781

Minaee, S., Kalchbrenner, N., Cambria, E., Nikzad, N., Chenaghlu, M., & Gao, J. (2021). Deep learning--based text classification: A comprehensive review. *ACM Computing Surveys*, 54(3), Article 62. https://doi.org/10.1145/3439726

Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C. L., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., Schulman, J., Hilton, J., Kelton, F., Miller, L., Simens, M., Askell, A., Welinder, P., Christiano, P., Leike, J., & Lowe, R. (2022). *Training language models to follow instructions with human feedback*. https://arxiv.org/abs/2203.02155

Peters, M. E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., & Zettlemoyer, L. (2018). *Deep contextualized word representations*. https://arxiv.org/abs/1802.05365

Qin, X., & Huang, J. (2024). Policy punctuations and agenda diversity in China: A national level analysis from 1980 to 2019. *Policy Studies*, 45(1), 21-41. https://doi.org/10.1080/01442872.2023.2183944

Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W., & Liu, P. J. (2020). Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of Machine Learning Research*, 21(1), Article 140.

Rao, Y., & Yang, F. (2022). A method for classifying information in education policy texts based on an improved attention mechanism model. *Wireless Communications and Mobile Computing*, 2022(1), 5467572. https://doi.org/10.1155/2022/5467572

Sheng, L., Chen, G., Gao, Y., Lin, Q., Lin, X., & Chen, Y. (2024). Quantitative evaluation of innovation policy based on text analysis – taking Wenzhou as an example. *Asian Journal of Technology Innovation*, 32(3), 457-480. https://doi.org/10.1080/19761597.2023.2252015

Song, D., Vold, A., Madan, K., & Schilder, F. (2022). Multi-

label legal document classification: A deep learning-based approach with label-attention and domain-specific pre-training. *Information Systems*, 106, Article 101718. https://doi.org/10.1016/j.is.2021.101718

Suzuki, M., Sakaji, H., Hirano, M., & Izumi, K. (2023). Constructing and analyzing domain-specific language model for financial text mining. *Information Processing & Management*, 60(2), 103194. https://doi.org/10.1016/j.ipm.2022.103194

Tai, W., Kung, H. T., Dong, X., Comiter, M., & Kuo, C. F. (2020, November). exBERT: Extending pre-trained models with domain-specific vocabulary under constrained training resources. In T. Cohn, Y. He, & Y. Liu (Eds.), *Findings of the Association for Computational Linguistics: EMNLP 2020* (pp. 1433–1439). Association for Computational Linguistics.

Thoppilan, R., Freitas, D. D., Hall, J., Shazeer, N., Kulshreshtha, A., Cheng, H. T., Jin, A., Bos, T., Baker, L., Du, Y., Li, Y., Lee, H., Zheng, H. S., Ghafouri, A., Menegali, M., Huang, Y., Krikun, M., Lepikhin, D., Qin, J., … Le, Q. (2022). *LaMDA: Language models for dialog applications*. https://arxiv.org/abs/2201.08239

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). *Attention is all you need*. Paper presented at 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA.

Wang, X., Zhou, Z., Sun, L., Xie, G., & Lou, Q. (2022, April 15-17). A new energy policy text classification and recognition model based on convolutional neural networks. In T. T. Lie, & Y. Liu (Eds.), *2022 7th Asia Conference on Power and Electrical Engineering (ACPEE)* (pp. 2083-2090). IEEE.

Wei, S., Guo, J., Yu, Z., Chen, P., & Xian, Y. (2013, May 25-27). The instructional design of Chinese text classification based on SVM. *2013 25th Chinese Control and Decision Conference (CCDC)* (pp. 5114-5117). IEEE.

Wu, H., Liu, Y., & Wang, J. (2020). Review of text classification methods on deep learning. *Computers, Materials & Continua*, 63(3), 1309-1321.

Yu, B., Deng, C., & Bu, L. (2022, February 4-6). Policy text classification algorithm based on BERT. *2022 11th International Conference of Information and Communication Technology (ICTech)* (pp. 488-491). IEEE.

Zhang, J., Gan, R., Wang, J., Zhang, Y., Zhang, L., Yang, P., Gao, X., Wu, Z., Dong, X., He, J., Zhuo, J., Yang, Q., Huang, Y., Li, X., Wu, Y., Lu, J., Zhu, X., Chen, W., Han, T., … Chen, C. (2023). *Fengshenbang 1.0: Being the foundation of Chinese cognitive intelligence*. https://arxiv.org/abs/2209.02970

Zhang, J., Zhao, Y., Saleh, M., & Liu, P. (2020). PEGASUS: Pre-training with extracted gap-sentences for abstractive summarization. In H. D. III, & A. Singh (Eds.), *Proceedings of the 37th International Conference on Machine Learning* (pp.11328-11339). Proceedings of Machine Learning Research.