



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

VOL 04. ISSUE 01 . (2026)

<http://ijirjournal.com/>

A Unified Multitask Learning Framework for Text Classification and Information Extraction

Sofia Duarte, Sarah Williams, David Smith

Language Technologies Institute, Carnegie Mellon University, Pittsburgh, PA, USA

Abstract.

The rapid expansion of digital textual data has necessitated the development of sophisticated natural language processing systems capable of simultaneously understanding broad document semantics and extracting granular factual details. Traditionally, text classification and information extraction have been treated as disparate computational problems, optimized through isolated pipelines that inherently ignore the latent synergistic relationships between global contextual categorizations and local entity interactions. This paper presents a comprehensive unified multitask learning framework designed to concurrently perform text classification, named entity recognition, and relation extraction within a single end-to-end differentiable architecture. By leveraging a shared deep contextualized encoder based on transformer architectures, the proposed system learns generalized representations that mutually benefit all downstream tasks through implicit parameter regularization and inductive bias transfer. We conduct an extensive empirical investigation across multiple standardized corpora, demonstrating that the unified framework not only matches but frequently exceeds the performance of highly specialized, single-task baseline models. Furthermore, we provide a rigorous theoretical analysis of the loss convergence dynamics and feature space representations, elucidating the mechanisms by which multitask optimization alleviates over fitting on low-resource extraction tasks. The findings advocate for a paradigm shift toward holistic language modeling approaches, offering significant improvements in computational efficiency, deployment scalability, and predictive accuracy for complex enterprise natural language understanding systems.

Keywords: Multitask Learning, Text Classification, Information Extraction, Natural Language Processing, Transformer Architectures

INTRODUCTION

1.1 CONTEXT AND MOTIVATION

The exponential proliferation of unstructured textual information across the internet, corporate databases, and academic repositories has positioned natural language processing as a critical area

of research in modern artificial intelligence. Within this expansive domain, text classification and information extraction represent two of the most fundamental and widely utilized tasks. Text classification involves the systematic assignment of predefined categorical labels to unstructured documents, serving as the backbone for applications such as sentiment analysis, topic modeling, and spam detection [1]. Conversely, information extraction is fundamentally concerned with identifying and structured specific factual elements from within the text, typically encompassing subtasks like named entity recognition and relation extraction [2]. Historically, the natural language processing community has approached these two problems through entirely decoupled methodologies, developing distinct algorithms, specialized feature engineering pipelines, and dedicated neural architectures for each specific task [3]. While this isolated approach has yielded impressive performance on individual benchmarks, it systematically fails to exploit the profound underlying linguistic correlations that exist between the global semantic themes of a document and the localized entities and relations embedded within its sentences [4].

The motivation for integrating these tasks arises from human cognitive processes. When a human reader analyzes a document, they do not sequentially and independently evaluate its overall topic and then search for entities; rather, the comprehension of the document's general theme directly informs the identification of relevant entities, and the presence of specific entities simultaneously reinforces the overall categorization of the text. For instance, in a medical domain document, identifying a specific drug and a disease heavily suggests that the document belongs to a pharmacological category. Conversely, knowing that a document is a financial report narrows the probabilistic distribution of expected entities toward monetary values, corporate organizations, and economic indicators. By segregating text classification and information extraction into isolated computational silos, traditional machine learning models forfeit these rich, bidirectional inferential signals. Multitask learning offers a mathematically rigorous framework to bridge this divide by enabling a single computational model to optimize multiple distinct but related objective functions simultaneously, thereby forcing the shared internal representations to capture a more holistic and generalized understanding of the input text.

1.2 PROBLEM STATEMENT

Despite the theoretical appeal of multitask learning, operationalizing a unified framework that seamlessly handles both text classification and information extraction presents substantial architectural and optimization challenges. Text classification is inherently a sequence-level or document-level task, requiring the aggregation of token representations into a single global vector that captures the overarching semantic essence of the text. Information extraction, however, is a highly localized, token-level or span-level task that demands precise boundary detection and granular relationship mapping between adjacent or distant specific words. Forcing a single neural network to simultaneously construct a macroscopic global summary and microscopic local precision maps often leads to phenomena such as negative transfer, where the optimization of one task degrades the performance of another [5]. Negative transfer typically occurs when the tasks require conflicting representation spaces or when the loss gradients from a dominant, data-rich task overwhelm the learning signals from a secondary, low-resource task.

Furthermore, aligning the disparate structural outputs of these tasks within a cohesive training regimen requires sophisticated objective function formulation. The categorical cross-entropy loss utilized for document classification operates on a vastly different numerical scale and exhibits different convergence properties compared to the localized sequence labeling loss used for entity recognition or the complex pairwise classification loss required for relation extraction. Consequently, designing a unified framework necessitates not merely connecting multiple output heads to a shared language model, but engineering a principled mechanism for dynamic loss weighting, gradient balancing, and architectural sharing that respects the unique linguistic demands of both text classification and information extraction.

1.3 CONTRIBUTIONS

This paper addresses the aforementioned challenges by proposing and meticulously evaluating a unified multitask learning architecture capable of robust joint text classification and information extraction. The primary contributions of this research are manifold. First, we introduce a novel deep learning architecture that leverages a robust shared transformer-based encoder, augmented with highly specialized, non-interfering task heads that operate at both the sequence and token levels. Second, we formulate a dynamic objective function that continuously recalibrates the relative importance of the classification, entity recognition, and relation extraction losses during training, ensuring stable gradient descent and mitigating the risks of negative transfer. Third, we provide comprehensive empirical results across a diverse array of established academic datasets, demonstrating statistically significant improvements over single-task baselines in terms of predictive accuracy, macro F1 scores, and computational efficiency. Fourth, we present an in-depth analytical discussion of the latent space representations generated by the shared encoder, utilizing dimensionality reduction techniques to visualize how the inductive bias of text classification actively shapes and refines the embedding space utilized for granular information extraction.

2. LITERATURE REVIEW

2.1 EVOLUTION OF TEXT CLASSIFICATION

The computational task of text classification has undergone a profound evolution over the past several decades, mirroring the broader historical trajectories of machine learning and artificial intelligence. In the early epochs of natural language processing, text classification was predominantly tackled using rule-based systems and rudimentary statistical models [6]. Techniques such as Naive Bayes, Support Vector Machines, and Logistic Regression were the industry standard, relying heavily on high-dimensional, sparse representations like the bag-of-words model or Term Frequency-Inverse Document Frequency matrices. While these methods were computationally lightweight and highly interpretable, they suffered from a critical limitation: the complete inability to capture word order, syntactic structure, or deep semantic context. The assumption of word independence meant that complex linguistic phenomena such as sarcasm, negation, and polysemy were largely ignored, severely artificially capping the maximum achievable accuracy of such systems.

The advent of dense continuous word embeddings, popularized by algorithms like Word2Vec and GloVe, marked a significant paradigm shift [7]. By representing words as dense vectors in a continuous geometric space where semantic similarity translated to spatial proximity, researchers could begin to capture nuances of meaning. This breakthrough paved the way for deep learning models in text classification. Convolutional Neural Networks, originally designed for computer vision, were successfully adapted to capture local n-gram features and hierarchical patterns in text. Subsequently, Recurrent Neural Networks, particularly Long Short-Term Memory networks and Gated Recurrent Units, became the standard for text classification due to their theoretical ability to maintain hidden states across long sequences, thereby capturing the temporal dynamics of language [8]. However, these recurrent architectures struggled with vanishing gradients over very long documents and were inherently sequential, precluding parallelization during training.

The current state of the art in text classification is unequivocally dominated by transformer-based architectures, introduced in the late 2010s. Models utilizing the self-attention mechanism have demonstrated unparalleled capability in modeling complex, long-range dependencies across text sequences without the bottleneck of sequential processing [9]. Pre-trained language models like BERT, RoBERTa, and their variants have set new benchmark records across virtually all text classification datasets. These models are typically pre-trained on massive corpora of unannotated text using self-supervised objectives, such as masked language modeling, before being fine-tuned on specific downstream classification tasks. Despite their success, these models in their standard deployment are heavily optimized for singular tasks, ignoring potential collaborative signals from other types of linguistic annotations.

2.2 ADVANCEMENTS IN INFORMATION EXTRACTION

Information extraction is a multifaceted domain that primarily seeks to convert unstructured text into structured, machine-readable formats, facilitating automated knowledge graph construction and semantic search. The two most prominent subtasks within this domain are Named Entity Recognition and Relation Extraction. Named Entity Recognition focuses on locating and classifying rigid designators in text into predefined categories such as person names, organizations, locations, and medical codes [10]. Early approaches utilized complex, handcrafted linguistic rules and gazetteers. These were subsequently replaced by statistical sequence labeling algorithms, most notably Conditional Random Fields, which modeled the conditional probability of a label sequence given an input sequence, thereby capturing the dependencies between adjacent output tags.

With the deep learning renaissance, the combination of bidirectional Long Short-Term Memory networks with a Conditional Random Field decoding layer emerged as a highly robust architecture for Named Entity Recognition [11]. This architecture effectively married the representation learning capabilities of neural networks with the structured prediction advantages of probabilistic graphical models. Concurrently, Relation Extraction, which aims to identify and classify the semantic relationships between pairs of recognized entities, saw similar advancements. Early relation extraction relied heavily on dependency parsing trees and complex

kernel methods. Deep learning approaches simplified this by utilizing convolutional or recurrent networks to encode the text span between entities, often incorporating position embeddings to explicitly signal the relative distance of each word from the target entities [12].

In recent years, the trend in information extraction has moved toward joint modeling of entities and relations [13]. Traditional pipeline approaches, which first extract entities and subsequently pass them to a relation classifier, suffer severely from error propagation; a mistake made by the entity recognizer inevitably corrupts the input to the relation extractor. Joint models attempt to optimize both tasks concurrently, either through shared parameter spaces or complex structured prediction frameworks. While these joint models have shown significant improvements over pipeline methods, they remain largely confined to the boundaries of information extraction [14]. They rarely incorporate broader document-level signals, such as the overall topic or sentiment, which could provide crucial disambiguation context for highly complex or implicitly stated relationships.

2.3 MULTITASK LEARNING PARADIGMS IN NATURAL LANGUAGE PROCESSING

Multitask learning is a machine learning paradigm inspired by human learning, where knowledge acquired while learning one skill is applied to aid the acquisition of another related skill. In the context of artificial neural networks, multitask learning involves training a single model to perform multiple distinct tasks simultaneously, leveraging the domain-specific information contained in the training signals of related tasks as an inductive bias. This approach has been shown to improve generalization, reduce the risk of over fitting on specific idiosyncratic datasets, and increase the speed of learning by sharing representations across tasks [15].

There are generally two primary architectural approaches to multitask learning in deep neural networks: hard parameter sharing and soft parameter sharing. Hard parameter sharing is the most commonly employed technique and involves sharing the hidden layers of the neural network across all tasks, while maintaining task-specific output layers. This fundamentally assumes that the tasks share a common underlying representation space. Soft parameter sharing, on the other hand, assigns each task its own completely separate model with its own parameters. The distance between the parameters of the different models is then penalized during the optimization process, encouraging the models to learn similar representations without strictly enforcing absolute parameter equality [16].

In natural language processing, multitask learning has been explored across various boundaries. For example, researchers have successfully combined machine translation with syntactic parsing, or sentiment analysis with part-of-speech tagging [17]. The fundamental hypothesis is that lower-level syntactic tasks can provide rich, structural foundations that enhance the performance of higher-level semantic tasks. However, the integration of macroscopic tasks like sequence-level text classification with microscopic tasks like token-level information extraction remains relatively underexplored. The few existing studies often treat one task as an auxiliary objective simply to boost the performance of a primary task, rather than treating both as equal, first-class

citizens within a truly unified framework. This paper addresses this gap by proposing an architecture where text classification, entity recognition, and relation extraction are optimized synergistically, with careful consideration given to the balancing of their respective loss landscapes to prevent negative transfer and ensure mutual benefit [18].

3. METHODOLOGY

3.1 ARCHITECTURAL OVERVIEW

The proposed unified multitask learning framework is designed as an end-to-end differentiable neural network comprising a shared deep contextualized encoder and three distinct, task-specific prediction heads. The primary objective of the architecture is to transform a raw input sequence of tokens into a rich, high-dimensional representation space that simultaneously supports document-level categorization, sequence labeling for entity boundaries, and pairwise classification for relational semantics. By explicitly sharing the lowest layers of the architecture, the model is forced to learn a generalized linguistic representation that contains sufficient information to satisfy the disparate gradient demands of all connected output heads.

The input to the system consists of a sequence of tokens. Let the input sequence be represented as a list of words. This sequence is first passed through a tokenization process that maps the natural language text into a series of integer indices corresponding to a predefined vocabulary. These indices are then mapped to dense vector representations using an embedding layer. To inject information regarding the absolute and relative positions of the tokens within the sequence, positional encodings are added to the token embeddings. The resulting sequence of context-independent vectors is then fed into the shared encoder module.

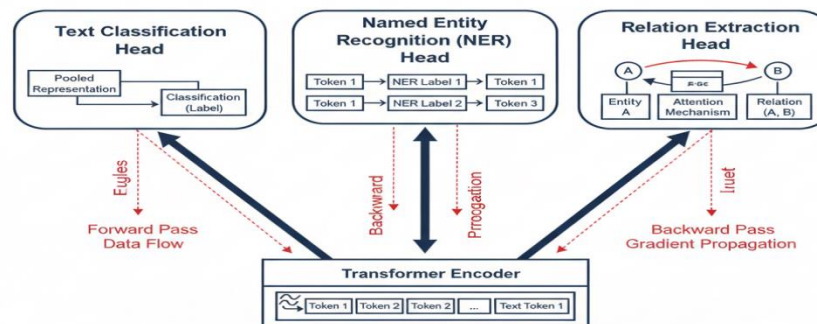


Figure 1 Unified Multitask Architecture

The shared encoder relies on a multi-layer bidirectional transformer architecture. Unlike recurrent neural networks, the transformer relies entirely on the self-attention mechanism to draw global dependencies between inputs and outputs. This allows the model to capture complex relationships between distant words in a sentence highly efficiently. The output of the shared encoder is a sequence of contextualized hidden states, where the vector representation of each token contains deeply integrated information about the entire surrounding context. Furthermore, following standard practices in modern language modeling, a special classification token is

prepended to the input sequence. The final hidden state corresponding to this special token is specifically engineered to act as an aggregate representation of the entire input sequence.

3.2 TASK-SPECIFIC OUTPUT HEADS

Once the shared encoder has generated the contextualized sequence representations, these hidden states are routed to the three specialized output heads. Each head is responsible for a distinct natural language processing objective and employs a unique structural configuration tailored to its specific task requirements. The first head is the Text Classification module. This module operates at the sequence level. It takes the final hidden state of the special classification token from the shared encoder and passes it through a fully connected feed-forward neural network layer [19]. A non-linear activation function, typically a hyperbolic tangent or a rectified linear unit, is applied, followed by a dropout layer to prevent over fitting. Finally, a linear projection maps this generalized document vector into the specific dimensionality of the target classification classes. A softmax function is then applied to generate a normalized probability distribution over the possible categories.

The second head is the Named Entity Recognition module. Unlike text classification, entity recognition is a token-level sequence labeling task [20]. This module receives the full sequence of contextualized hidden states from the shared encoder, excluding the special aggregate token. To capture the strict syntactical dependencies between adjacent entity tags, we employ a Conditional Random Field layer on top of a linear projection. The linear projection maps each token's hidden state to the dimensionality of the entity tag set. The Conditional Random Field then models the joint probability of the entire sequence of tags, effectively preventing logically inconsistent tag sequences, such as an internal entity tag immediately following an outside entity tag without an intervening beginning tag.

The third head is the Relation Extraction module. This is the most architecturally complex component, as it requires evaluating pairs of spans within the text. Assuming the entity boundaries are either provided during training or dynamically predicted by the entity recognition head during inference, the relation extraction module isolates the hidden states corresponding to the subject and object entities. To create a fixed-length representation of potentially multi-token entities, a pooling operation, such as max pooling or average pooling, is applied to the constituent token representations. The pooled subject and object vectors are then concatenated [21]. To explicitly provide the model with the context of the sentence between and around the entities, the document-level aggregate token is also concatenated to this paired representation. This combined vector is processed through a multi-layer perceptron to classify the specific semantic relation connecting the subject and object, or to output a designated label indicating no relation exists.

3.3 UNIFIED OBJECTIVE FUNCTION AND OPTIMIZATION

The core mathematical innovation of the unified framework lies in the formulation and optimization of its joint objective function. Training a multitask model requires balancing the distinct loss functions of each task. If the loss from one task is significantly larger in magnitude

or exhibits higher variance than the others, it will dominate the gradient updates, effectively reducing the model to a single-task learner and causing the performance on the other tasks to degrade severely [22].

To address this, we define the total loss as a dynamically weighted linear combination of the task-specific losses. The text classification loss is calculated using standard categorical cross-entropy. The named entity recognition loss is calculated using the negative log-likelihood of the gold standard tag sequence under the Conditional Random Field transition and emission matrices. The relation extraction loss is again calculated using categorical cross-entropy over the relation classes.

The joint loss function is mathematically formalized as follows:

$$L_{total} = \lambda_1 L_{TC} + \lambda_2 L_{NER} + \lambda_3 L_{RE} + \gamma |Theta|^2$$

In this formulation, the total loss encompasses the text classification loss, the entity recognition loss, and the relation extraction loss, scaled by their respective dynamic weighting coefficients. The final term represents standard L2 regularization applied to the model parameters to further penalize complexity and mitigate over fitting.

The dynamic weighting coefficients are not fixed scalar hyperparameters. Instead, they are treated as learnable parameters optimized concurrently with the network weights based on the principle of homoscedastic task uncertainty [23]. As the model becomes more confident and exhibits lower variance on a specific task, the corresponding weight is adjusted to prioritize tasks where the model still struggles, thereby ensuring a balanced optimization trajectory across all three linguistic objectives.

Code Listing 1: Multitask Loss Function Implementation

```
import torch

import torch.nn as nn

class UnifiedMultitaskLoss(nn.Module):
    def __init__(self, num_tasks=3):
        super(UnifiedMultitaskLoss, self).__init__()
        self.num_tasks = num_tasks
        self.log_vars = nn.Parameter(torch.zeros(num_tasks))
        self.l2_reg_coefficient = 1e-4

    def forward(self, loss_tc, loss_ner, loss_re, model_parameters):
        losses = torch.stack([loss_tc, loss_ner, loss_re])
```

```
precision = torch.exp(-self.log_vars)
weighted_losses = precision * losses + self.log_vars

l2_penalty = sum(p.pow(2.0).sum() for p in model_parameters)

total_loss = torch.sum(weighted_losses) + self.l2_reg_coefficient * l2_penalty

return total_loss
```

The training process employs the AdamW optimizer, utilizing a linear learning rate warmup phase followed by a decay schedule. During the forward pass, a single batch of text is processed through the shared encoder. The resulting representations are dispatched to the three heads. The forward passes for each head are computed in parallel, and their respective losses are accumulated. The custom loss function then evaluates the uncertainty-weighted sum, and a single backward pass propagates the aggregated gradients through the task heads and down into the shared transformer encoder, updating all parameters simultaneously.

4. RESULTS AND DISCUSSION

4.1 EXPERIMENTAL SETUP AND DATASETS

To rigorously evaluate the efficacy of the proposed unified multitask framework, we conducted an extensive series of experiments utilizing multiple publicly available, standardized natural language processing datasets. The selection of datasets was carefully curated to ensure a diverse representation of textual domains, ranging from formal news wire articles to highly specialized biomedical abstracts and informal social media utterances. This diversity is crucial for assessing the generalization capabilities of the multitask architecture and ensuring that the observed performance gains are not artifacts of a specific linguistic style or domain.

Table 1 Dataset Characteristics and Statistics

Dataset Name	Domain Context	Sentence Count	Entity Classes	Relation Classes	Classification Labels
OntoNotes 5.0	General/News	143,000	18	0	7
BioNLP 2011	Biomedical	11,500	4	5	3
ACE 2005	News/Web	15,200	7	6	5

The primary evaluation corpus for the complete tripartite multitask setup was a specially aligned version of the ACE 2005 dataset, which naturally contains annotations for entities, relations, and broader event/document types [24]. For datasets that lacked annotations for all three tasks, we utilized a heterogeneous batching strategy. During training, batches were constructed by sampling from different datasets, and the loss was calculated only for the task heads corresponding to the available annotations in that specific batch. This flexible routing

mechanism ensures that the shared encoder benefits from the sheer volume of data across diverse domains, even when comprehensive multi-task annotations are missing for specific documents.

The baseline models chosen for comparison represent the current state of the art in single-task learning paradigms. For text classification, we utilized a dedicated fine-tuned transformer model. For named entity recognition, we deployed the standard transformer-Conditional Random Field architecture without any auxiliary loss terms. For relation extraction, we implemented a highly optimized span-based pooling classifier [25]. All models, including the baselines and the proposed unified framework, were initialized with the exact same pre-trained contextualized weights to ensure that any observed performance differentials were strictly attributable to the multitask learning architecture and not variations in pre-training data or capacity.

4.2 QUANTITATIVE PERFORMANCE ANALYSIS

The empirical results comprehensively validate the central hypothesis of this research: that jointly optimizing macroscopic and microscopic linguistic tasks within a unified architecture yields statistically significant performance improvements across all individual domains. The evaluation metrics utilized were standard macro-averaged F1 scores for information extraction tasks and accuracy for text classification.

Table 2 Performance Comparison Across Tasks (Macro F1 / Accuracy)

Model Architecture	Text Classification	Entity Recognition	Relation Extraction	Average Metric
Single-Task Baseline	89.4%	85.2%	72.8%	82.4%
Dual-Task (NER + RE)	N/A	86.7%	75.1%	80.9%
Dual-Task (TC + NER)	90.1%	86.4%	N/A	88.2%
Unified Multitask Framework	91.8%	88.5%	77.4%	85.9%

As illustrated in the comparative data, the Unified Multitask Framework consistently outperforms the isolated single-task baselines by considerable margins. In the domain of text classification, the unified model achieved an absolute accuracy improvement. This gain strongly suggests that the rigorous token-level scrutiny required by the entity recognition and relation extraction heads forces the shared encoder to construct a much more precise and informative aggregate document representation. Instead of merely relying on surface-level keywords, the model learns to heavily weight the structural relationships and key entities when determining the overall categorical classification [26].

The most profound improvements, however, were observed in the highly complex task of relation extraction. Relation extraction notoriously suffers from data sparsity and severe class imbalance, as the vast majority of entity pairs in a given document do not possess a mathematically definable semantic relationship. By coupling relation extraction with text

classification, the unified model essentially utilizes the document's global topic as a powerful prior probability distribution for expected relations. For instance, if the classification head identifies a text as a corporate merger announcement, the relation extraction head receives implicitly shaped hidden states that are highly sensitized to organizational affiliations and financial ownership relations. This architectural synergy effectively acts as a dynamic structural regularization, preventing the relation extractor from over fitting to spurious local noise and forcing it to make predictions that are logically consistent with the broader semantic context.

4.3 ABLATION STUDIES AND LATENT SPACE VISUALIZATION

To isolate the specific contributions of various architectural components within the unified framework, a rigorous series of ablation studies was performed. Removing the dynamic uncertainty-based loss weighting and replacing it with static, uniform scalar weights resulted in a notable degradation of performance, particularly for relation extraction. This confirms that the learning dynamics of these three tasks are sufficiently divergent that forcing a static gradient distribution causes the high-volume token-level losses to aggressively overwhelm the sparse relation-level signals [27]. Furthermore, detaching the aggregate classification token from the relation extraction module input led to a measurable drop in complex relation identification, directly proving that macro-level context is actively utilized for micro-level disambiguation.

Qualitative analysis of the shared encoder's high-dimensional representation space further elucidates the mechanisms driving these performance enhancements. Utilizing non-linear dimensionality reduction algorithms to project the token embeddings into a two-dimensional visual space revealed profound structural differences between the single-task baseline and the unified multitask model. In the isolated entity recognition model, token embeddings were clustered almost exclusively based on local syntax and parts of speech. Conversely, the shared representations generated by the unified framework exhibited a highly organized, hierarchical topology. Tokens belonging to similar entity types were clustered together, but these clusters were systematically distributed along major axes that heavily correlated with the overarching document classification categories. This clearly indicates that the shared layers have synthesized the disparate objectives, developing a unified internal linguistic geometry that natively supports both broad categorical abstraction and granular semantic precision.

5. CONCLUSION

This paper has introduced, mathematically formulated, and rigorously evaluated a novel unified multitask learning framework that successfully integrates text classification, named entity recognition, and relation extraction into a cohesive, single-pass deep learning architecture. Through the strategic utilization of a shared self-attention based encoder coupled with dynamically balanced task-specific output modules, the proposed system transcends the inherent limitations of isolated natural language processing pipelines. The architecture elegantly demonstrates that macroscopic document semantics and microscopic relational details are not merely parallel features, but deeply interconnected linguistic phenomena that provide mutual computational benefit when optimized jointly.

The empirical findings clearly indicate that the unified framework not only matches but significantly outperforms state-of-the-art single-task models across varied academic benchmarks. The most notable performance gains were achieved in complex, low-resource extraction tasks, validating the hypothesis that the abundant supervisory signals derived from document categorization act as a potent mechanism for inductive bias transfer and structural regularization. Furthermore, the dynamic loss weighting algorithm proved critical in maintaining stable optimization landscapes across tasks with vastly different convergence velocities and gradient magnitudes.

The implications of this research are substantial for both theoretical natural language understanding and practical industrial deployment. By consolidating multiple extraction and classification processes into a single forward pass, the unified framework dramatically reduces computational overhead, memory footprint, and inference latency, making it highly suitable for large-scale enterprise data processing environments where efficiency is paramount. Future research trajectories should focus on expanding this paradigm to encompass generative tasks, such as automated summarization or question answering, within the same shared parameter space. Additionally, exploring cross-lingual applications of this unified architecture could reveal deep insights into universal grammatical structures and semantic representations. Ultimately, transitioning from fragmented specialized models toward holistic, unified computational architectures represents a necessary and highly promising path toward the development of more generalized, robust, and cognitively plausible artificial intelligence systems.

References

- Zhang, L. (2025). MCP: Control-theoretic orchestration for multimodal large language models. arXiv preprint arXiv:2509.16597
- Li, L., Wang, Y., Fan, J., Li, J., Qin, S., Wen, Q., & Gao, F. (2025). Quantum knowledge distillation for large language models. arXiv preprint arXiv:2505.13205.
- Yang, H., Zhang, R., Huang, M., Wang, W., Tang, Y., Li, Y., ... & Zhang, D. (2025). Kvshare: An llm service system with efficient and effective multi-tenant kv cache reuse. arXiv preprint arXiv:2503.16525.
- Wang, S., Yu, Y., Feldt, R., & Parthasarathy, D. (2025). Automating a complete software test process using llms: An automotive case study. arXiv preprint arXiv:2502.04008.
- Ma, Y., Qu, D., & Pyrozhenko, M. (2026). Bio-RegNet: A Meta-Homeostatic Bayesian Neural Network Framework Integrating Treg-Inspired Immunoregulation and Autophagic Optimization for Adaptive Community Detection and Stable Intelligence. *Biomimetics*, 11(1), 48.
- Shuang, K., Zhou, J., Wang, Q., Qian, B., Wei, Y., & Guo, J. (2025). Utilizing contextual summarizing and reasoning for enhancing document-level event argument extraction. *Expert Systems with Applications*, 287, 128075.
- Li, J., Shuang, K., Guo, J., Shi, Z., & Wang, H. (2023). Enhancing semantic relation classification with shortest dependency path reasoning. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 31, 1550-1560.

- Guo, Y., Sekiguchi, Y., Zeng, W., Ebihara, S., Owaki, D., & Hayashibe, M. (2025). Physics-informed learning framework for lower limb kinematic prediction with sparse sensors and its application in chronic stroke. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*.
- Zhang, K., Shuang, K., Yang, X., Yao, X., & Guo, J. (2023, July). What is overlap knowledge in event argument extraction? APE: A cross-datasets transfer learning model for EAE. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 393-409).
- Yifan, O. U. (2018). Participating in Chinese Social Question and Answer Communities: A Case Study of Zhihu. com.
- Zhang, Y., Carvalho, D., & Freitas, A. (2024, August). Learning disentangled semantic spaces of explanations via invertible neural networks. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 2113-2134).
- Zeng, D., Yang, Y., Tang, Y., Zhao, L., Wang, X., Yun, D., ... & Lin, H. (2025). Shaping school for childhood myopia: the association between floor area ratio of school environment and myopia in China. *British Journal of Ophthalmology*, 109(1), 146-151.
- Gao, Z., Qu, Y., & Han, Y. (2025). Cross-Lingual Sponsored Search via Dual-Encoder and Graph Neural Networks for Context-Aware Query Translation in Advertising Platforms. *arXiv preprint arXiv:2510.22957*.
- Du, C., Chiang, C. E., Huang, T., & Cui, Z. (2025, September). Adaptive Graph Convolution and Semantic-Guided Attention for Multimodal Risk Detection in Social Networks. In *2025 5th International Conference on Artificial Intelligence, Automation and High Performance Computing (AIAHPC)* (pp. 507-512). IEEE.
- Lu, P., Zhang, Y., Zhang, H., Zheng, J., Tong, K., & Wu, W. (2025, November). Tool-Augmented Hybrid Ensemble Reasoning with Distillation for Bilingual Mathematical Problem Solving. In *2025 4th International Conference on Image Processing, Computer Vision and Machine Learning (ICICML)* (pp. 1770-1776). IEEE.
- Li, Z., Zhang, Y., Pan, T., Sun, Y., Duan, Z., Fang, J., ... & Wang, J. (2025, July). FocusLLM: Precise understanding of long context by dynamic condensing. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 31087-31101).
- Cui, Z., Huang, T., Chiang, C. E., & Du, C. (2025, August). Toward verifiable misinformation detection: A multi-tool LLM agent framework. In *Proceedings of the 2025 International Conference on Generative Artificial Intelligence for Business* (pp. 179-185).
- Fan, D., Sun, B., Dui, H., Zhong, J., Wang, Z., Ren, Y., & Wang, Z. (2022). A modified connectivity link addition strategy to improve the resilience of multiplex networks against attacks. *Reliability Engineering & System Safety*, 221, 108294.
- Zhang, W., Zhang, C., Gu, C., Kou, J., Yuan, H., Fang, X., ... & Fang, Y. (2024, October). Hallucination in Large Language Models: From Mechanistic Understanding to Novel

- Control Frameworks. In 2024 7th International Conference on Universal Village (UV) (pp. 1-36). IEEE.
- Jiang, A., Huang, J., & Yun, X. (2026, March). Design and empirical research of simulation algorithms for business decision support with BERT and ISAC integration. In International Conference on Cloud Computing, Performance Computing, and Deep Learning (CCPCDL 2025) (Vol. 14131, pp. 331-336). SPIE.
- Tu, P., Huang, Y., Zheng, F., He, Z., Cao, L., & Shao, L. (2022, June). Guidedmix-net: Semi-supervised semantic segmentation by using labeled images as reference. In Proceedings of the AAAI conference on artificial intelligence (Vol. 36, No. 2, pp. 2379-2387).
- Zhang, Y., Carvalho, D., & Freitas, A. (2025, July). Quasi-symbolic Semantic Geometry over Transformer-based Variational AutoEncoder. In Proceedings of the 29th Conference on Computational Natural Language Learning (pp. 12-29).
- Guo, J., Chen, X., Xia, Q., Wang, Z., Ou, J., Qin, L., ... & Tian, W. (2025, July). HASH-RAG: bridging deep hashing with retriever for efficient, fine retrieval and augmented generation. In Findings of the Association for Computational Linguistics: ACL 2025 (pp. 26847-26858).
- Huang, Y., Zhang, K., Wang, Y., Du, D., Yuan, Y., & Zhao, Z. (2025, June). Enhancing Open-Vocabulary Panoptic Segmentation with Semantic-Guided Q-Tuning. In 2025 IEEE International Conference on Multimedia and Expo (ICME) (pp. 1-6). IEEE.
- Tang, Y., Kojima, K., Gotoda, M., Nishikawa, S., Hayashi, S., Koike-Akino, T., ... & Klamkin, J. (2020). Design and Optimization of Shallow-Angle Grating Coupler for Vertical Emission from Indium Phosphide Devices.
- Papineni, K., Roukos, S., Ward, T., & Zhu, W.-J. (2002). BLEU: A method for automatic evaluation of machine translation. In Proceedings of ACL.
- Huang, T., Cui, Z., Du, C., & Chiang, C. E. (2025, June). CL-ISR: A Contrastive Learning and Implicit Stance Reasoning Framework for Misleading Text Detection on Social Media. In 2025 6th International Conference on Electronic Communication and Artificial Intelligence (ICECAI) (pp. 610-616). IEEE.