# Cross-Domain Multi-Task Learning with Emotion-Tuned BERT for Financial and Biomedical Texts

**Xiong Luo**
xiongluo@umass.edu

**Isaac Zhong**
bzhong@umass.edu

**Julia St.Germain**
jstgermain@umass.edu

## 1 Introduction

Transformer-based architectures, such as BERT, have revolutionized natural language processing (NLP) by providing strong generalization across diverse linguistic tasks. Despite their versatility, most models are still trained in a single-task or single-domain setting. This limits their ability to capture shared linguistic and semantic structures that exist across related fields, such as finance and biomedicine, where emotional or affective language plays a critical role in interpreting meaning. Moreover, building and maintaining separate models for each domain is inefficient and often results in redundant learning and inconsistent behavior.

In this project, we aim to bridge these gaps by developing a **multi-domain, multi-task BERT framework** capable of performing emotion and sentiment classification across multiple specialized domains. Our approach builds upon an emotion-aware foundation: we begin with a BERT model pre-trained on the GoEmotions dataset, which contains 27 fine-grained emotion categories derived from social media text. This initialization provides a rich affective prior that we then adapt to the financial and biomedical domains through unsupervised domain-adaptive pretraining and parameter-efficient fine-tuning.

To preserve the emotional knowledge learned from GoEmotions while adapting to new domains, we employ **adapter modules**, which are lightweight parameter-efficient layers inserted within BERT. These adapters enable specialization without catastrophic forgetting, allowing the model to retain affective understanding while learning domain-specific terminology and structures. Furthermore, we train the model under a **multi-task learning** (MTL) framework, where financial and biomedical datasets jointly contribute to shared emotional representations. This setup encourages knowledge transfer across domains and provides a natural setting for **zero-shot cross-domain evaluation**, where the model is trained on two domains and tested on a third unseen one.

Beyond technical efficiency, this research also offers linguistic and psychological insights. Emotionally charged words such as “recovery,” “loss,” or “growth” appear in both biomedical and financial contexts but carry distinct affective nuances. By aligning these expressions through a shared emotion-tuned representation, our model enables a deeper understanding of how emotions manifest across domains. Such analysis can reveal how affective patterns, such as fear, confidence, or hope, transfer or diverge between domains, shedding light on the cognitive and semantic parallels that underlie human communication in specialized contexts.

we include a third, general-domain dataset (e.g., social-media or reviews) to probe true cross-domain generalization beyond specialized text. This setup encourages knowledge transfer across domains and provides a natural setting for **zero-shot cross-domain evaluation**, where the model is trained on two domains and tested on a third unseen one.

Ultimately, our goal is to demonstrate that a unified, adapter-based, multi-domain model can outperform isolated single-domain systems in both accuracy and generalization, while remaining computationally efficient. The proposed approach contributes to the broader vision of creating emotionally grounded language models that generalize robustly across diverse, real-world applications.

## 2 Related work

(Zhao et al., 2021) proposed a BERT-based sentiment analysis and key-entity detection framework suitable for online financial texts. Their method is that they fine-tune RoBERTa to classify sentiment polarity and extract salient financial entities by using sentence-matching and machine-reading-comprehension formulations, and the performance is better than conventional NER and lexicon-based approaches. This work points out the limitations of traditional rule- or dictionary-based sentiment and entity recognition systems in financial contexts, which emphasizes the advantage of contextualized representations. Furthermore, they demonstrate that transformer models can capture domain-specific cues. For example, risk expressions and company events. However, their focus remains single-task and domain-specific, without exploring cross-domain or multi-task transfer.

(Zhang and Yang, 2022) presented a comprehensive survey of multi-task learning (MTL), categorized algorithms into several categories such as feature-learning, low-rank, task-clustering, task-relation, and decomposition approaches. Furthermore, they explained that multi-task learning improves model generalization by allowing related tasks to share useful representations. This shared learning process helps the model make better use of limited labeled data and alleviates the challenges caused by data scarcity. In addition, the survey also compared homogeneous and heterogeneous task settings and analyzed parameter-sharing strategies such as hard and soft sharing. Their conclusions indicate that properly structured MTL can reduce overfitting and promote robust knowledge sharing across domains. This is a great insight, which is applicable to combining financial and biomedical NLP tasks, where both domains face data imbalance and task interdependence.

(Chen et al., 2024) further analyzed multi-task learning architectures for NLP. They categorize them into parallel, hierarchical, modular, and generative-adversarial structures. They demonstrate how MTL mitigates overfitting, enhances data efficiency, and enables parameter-efficient adaptation in large pre-trained language models. In addition, they have reviewed several techniques such as parallel feature fusion, hierarchical pipelines, and multi-level supervision. The paper demonstrates that MTL can unify related NLP tasks within a shared transformer backbone. These insights can motivate our proposed fine-tune BERT by handling financial and biomedical tasks simulta. Even with two different fields, it can still preserve domain-specific subtasks.

(Chan et al., 2023) reviewed sequential transfer learning (STL) techniques for sentiment analysis and compared them with multi-task learning frameworks. They describe STL as the classic "pre-train then fine-tune" paradigm, where knowledge from a large source domain is transferred to a smaller target domain. While STL improves data efficiency, it may lead to catastrophic forgetting or limited cross-domain generalization when the tasks or domains diverge significantly. In contrast, multi-task learning jointly optimizes multiple objectives to learn task-invariant features, which enhances robustness and knowledge transfer across related tasks. This distinction directly motivates our design choice: rather than relying solely on sequential transfer (e.g., GoEmotions $\rightarrow$ domain fine-tuning), we adopt a multi-task BERT framework that learns sentiment and classification tasks concurrently across finance and biomedical domains.

Recent advances in parameter-efficient fine-tuning (PEFT) have enabled large pre-trained language models to adapt to multiple domains without full retraining. Adapter modules (Houlsby et al., 2019) introduce small trainable bottlenecks within each Transformer layer, achieving competitive performance while updating only a few percent of parameters. AdapterFusion (Pfeiffer et al., 2021) extends this idea by learning to combine multiple pre-trained adapters, effectively integrating knowledge from different tasks or domains. Similarly, Low-Rank Adaptation (LoRA) (Hu et al., 2021) updates only low-rank decompositions of weight matrices, providing a memory- and compute-efficient alternative to full fine-tuning. Beyond PEFT, multi-task learning approaches (Liu et al., 2019) share a common encoder across tasks with lightweight, task-specific output heads, promoting transfer of general linguistic features while mitigating catastrophic forgetting. Together, these methods form the foundation for efficient cross-domain adaptation explored in this work.

## 3 Our approach

Our goal is to develop a **multi-domain, multi-task BERT framework** capable of performing sentiment and emotion classification across three distinct domains: *financial*, *biomedical*, and one other domain that still needs to be determined. This approach differs from prior work in that it jointly learns representations across multiple tasks and domains, rather than training separate single-task or domain-specific models. By sharing a common encoder and leveraging complementary information between domains, the model aims to achieve robust cross-domain generalization, particularly in low-resource scenarios.

Our proposed pipeline consists of five stages:

1. **Emotionally Pretrained BERT Model:** We begin with a pre-trained BERT model that has already been fine-tuned on the GoEmotions dataset, which contains 27 emotion categories derived from social media text. This initialization provides the model with a rich affective understanding that serves as a foundation for subsequent domain adaptation.

2. **Domain-Adaptive Pretraining (Financial & Biomedical):** The emotion-pretrained model will undergo further unsupervised domain adaptation using large unlabeled corpora such as financial news, company reports, PubMed abstracts, and clinical notes. We will continue pretraining the model using masked language modeling (MLM) to capture domain-specific vocabulary and syntax. To mitigate catastrophic forgetting of emotional features, we will employ **adapter modules** and low learning rates, allowing efficient parameter updates while keeping the base model stable. Adapters allow the model to retain previously learned domain knowledge while efficiently specializing to new ones.

3. **Parameter-Efficient and Multi-Task Fine-Tuning:** We will integrate domain-specific **adapter layers** and lightweight classification heads for each dataset. These adapters enable parameter-efficient fine-tuning and preserve knowledge from previous domains. Training will follow a **multi-task learning** paradigm where shared representations are learned across emotion recognition tasks in all three domains, improving generalization. To evaluate true adaptability, we will also perform **zero-shot cross-domain testing**, where the model is trained on two domains (financial and biomedical data) and evaluated on the third unseen one, to assess transfer robustness.

4. **Cross-Domain Evaluation:** The model will be evaluated across all three domains using accuracy and macro F1-score. We will report both in-domain and cross-domain performance to measure transfer robustness. Ablation studies will compare full fine-tuning, adapter-based tuning, and multi-task configurations to understand trade-offs in efficiency and generalization.

5. **Structured Error Analysis and Insight Extraction:** Time permitting, we will conduct a detailed error analysis to categorize model failures by *emotion type*, *domain*, and *linguistic phenomenon*. This analysis will provide qualitative insights into model limitations and guide future improvements in emotion-aware domain adaptation.

This multi-domain and parameter-efficient approach aims to move beyond traditional binary domain transfer, offering a unified framework capable of robust, interpretable, and efficient emotion classification across diverse text sources.

**What baseline algorithms will you use?**

### 3.1 Baseline Algorithms

To evaluate the effectiveness of our proposed multi-domain, adapter-based approach, we will compare it against the following baselines:

1. **Majority-class baseline:** Predict the most frequent emotion class across all samples. This establishes a naive lower bound for performance evaluation.

2. **Single-domain BERT:** Fine-tune separate BERT models on each individual dataset (financial, biomedical, and social-media) without any shared parameters. This measures the benefit of unified representation learning compared to isolated, domain-specific fine-tuning.

3. **Sequential transfer learning (STL):** Pretrain on the GoEmotions dataset, then sequentially fine-tune on the financial and biomedical domains. This baseline tests traditional transfer learning and highlights issues such as catastrophic forgetting, against which our adapter-based method is designed to be more stable.

4. **Multi-task BERT without adapters:** Train a shared BERT model jointly on all three domains using standard full fine-tuning. This baseline isolates the impact of parameter-efficient tuning mechanisms (adapters or LoRA) on cross-domain performance and generalization.

5. **Zero-shot cross-domain evaluation:** Train the model on two domains and test on the third unseen domain. This baseline quantifies the model's ability to generalize emotional understanding to out-of-domain text.

6. **Adapter variants (ablation):** Replace adapters with LoRA or prefix-tuning under the same multi-task setup to compare PEFT choices and their effect on stability and efficiency.

These baselines provide a comprehensive comparison to assess whether multi-task learning and parameter-efficient strategies (adapters, LoRA) deliver measurable gains in accuracy, efficiency, and robustness for cross-domain emotion understanding.

### 3.2 Schedule

We will begin by acquiring and preprocessing both labeled and unlabeled datasets from the financial, biomedical, and social-media domains. Each team member will handle data preparation and initial experiments for one domain, after which we will collaborate on integrating the datasets into a unified multi-task framework. The model building and fine-tuning process will be performed jointly, with adapters and task heads trained under the same setup. Individual members will focus on evaluating domain-specific performance and conducting qualitative error analyses. We will finalize results, discussion, and presentation materials together.

1. Acquire and preprocess datasets (1 week)

2. Domain-adaptive pretraining using unlabeled corpora (1 week)

3. Multi-task fine-tuning with adapters on labeled datasets (2 weeks)

4. Cross-domain and zero-shot evaluation, structured error analysis (1 week)

5. Final report and presentation preparation (1 week)

## 4 Data

For this project, we will begin from an existing BERT model fine-tuned on GoEmotions (Stage 1). This is a large-scale emotion dataset developed by Google that contains over 58,000 Reddit comments annotated with 27 fine-grained emotion labels. In Stage 2 (domain adaptation), we will adapt this model to both financial and biomedical language using masked language modeling (MLM) on large unlabeled domain corpora. For the financial domain, we will use the lukecarlate/english_finance_news dataset, which contains a wide range of English financial sentences from news and reports. For the biomedical domain, we will use PubMed abstracts and BioASQ, which together provide large-scale unlabeled biomedical text along with structured question–answer and classification tasks that can later support fine-tuning. At this stage, we will also include text from labeled datasets such as Financial PhraseBank, FiQA, and Shekswess/AI-Healthcare-Biomedical-Sentiment, but use them without their labels, so that the model learns domain-specific terminology and context purely through self-supervised learning.

In Stage 3 (multi-task fine-tuning), we will fine-tune the model using labeled datasets for sentiment or emotion classification in each domain (Financial PhraseBank, FiQA, and Shekswess biomedical sentiment),under a shared multi-task framework with domain-specific adapters. This enables the model to retain emotional understanding from GoEmotions while learning task-relevant distinctions in each field.

In Stage 4 (cross-domain evaluation and generalization), we will introduce a **third domain**—a general-purpose corpus such as Amazon Polarity Reviews or social-media data—to evaluate the model's ability to generalize affective understanding beyond its training domains. The model will be trained on two domains (e.g., financial + biomedical) and tested on the third (e.g., general) in a zero-shot configuration. This setup will help quantify cross-domain transfer performance and reveal whether emotional features learned in specialized contexts extend to general language.

In Stage 5 (structured error analysis), we will analyze model performance across domains by categorizing misclassifications by emotion type, domain, and linguistic complexity (e.g., negation, figurative language, technical terminology). This will identify which emotions are domain-specific and where the model's generalization fails, guiding future refinement.

## 5 Tools

We will use Python as our main programming language. We will rely on deep learning frameworks such as PyTorch and Transformers from Hugging Face for implementing and fine-tuning BERT-based models. Preprocessing tasks, including finding the biomedical and financial datasets, and related utilities from the Transformers library. For evaluation and baseline models, we may use Scikit-learn to train traditional classifiers like logistic regression and SVM for comparison. We will also use pandas and NumPy for data manipulation and Matplotlib for visualization. Training will be conducted on mainly in GPU-enabled environments at Google Colab to efficiently handle fine-tuning. We do not need crowdsource because our selected datasets are already annotated and well-maintained.

## 6 AI Disclosure

- Did you use any AI assistance to complete this proposal? If so, please also specify what AI you used.
  - Yes, CHATGPT

*If you answered yes to the above question, please complete the following as well:*

- If you used a large language model to assist you, please paste *all* of the prompts that you used below. Add a separate bullet for each prompt, and specify which part of the proposal is associated with which prompt.
  - Summarize this paper
  - a short Adapters, AdapterFusion, LoRA, multi-task learning related essay which i need for related work. please provide the actual link
  - what is adapter in transformer
  - Can you find several financial- and biomedical-related BERT datasets
  - Draft a short introduction for the proposal based on the pretrained BERT model in GoEmotion and finetuning in the bio and financial areas by using transfer learning and multitask learning
  - im trying to understand from this article how starting from just unsupervised bio data it is able to preform sentiment analysis
  - i need help to get organized to get a clearly view of the pipeline for this project. we have this data for pretraining?: GoEmotions (27 emotions, Reddit) and then we would fine-tune on both finance/bio data? is that the multitaks model set up? is that different from the sequential transfer learning? is it worth exploring associative learning?
  - can you alter this intro so that i brings up deep associative learning in a way that we would add in the end potentially and why
  - help me reorganizing this data section to be more clear following the stages
  - so for stage 2 i don't need labeled data? why how does that work? what is MLM? because the data for finace we found is labeled, but not 27 emotions labels? is that useable? what about for the bio data ?
  - so our approach is changed, here is the new plan,Extend our current from binary domain transfer (finance biomedical) to a three-domain generalization by including another public dataset (for example, a social-media emotion dataset like GoEmotions).We integrate adapter modules to mitigate catastrophic forgetting during fine-tuning. And evaluate in zero-shot cross-domain settings — for example, train on two domains and test on a third unseen domain.We'll add a structured error analysis, which breaks down errors by emotion type and domain, with special focus on figurative language and maybe others please revise that
  - help me make this correct so it align witht he project and the stages we came up with: For this project, we plan to use published text datasets from both the financial and biomedi- cal domains to evaluate our multi-task BERT framework. For the financial part, we will use datasets such as the Financial PhraseBank and FiQA datasets, which pro-

vide labeled financial sentences for sentiment classification and aspect- based analysis. For the biomedical side, we will use ** NEED TO BE FILLED**. In addition, because our model needs to support cross-domain emotion understanding, we will also use GoEmotions. This is a large-scale emotion dataset devel- oped by Google that contains over 58,000 Reddit comments annotated with 27 fine-grained emotion labels. GoEmotions will be used as the source dataset for sequential transfer learning (STL) and as a shared auxiliary task for multi-task learning (MTL). These datasets are open-source and can be directly accessed through public repositories such as Kaggle and the Hugging Face Datasets Hub, so there are no licensing barriers. We do not plan to manually annotate data because existing datasets already provide high-quality annotations. Collec- tively, these sources are sufficient and appropri- ate for training, fine-tuning, and evaluating BERT- based models across both domains.

  – help me put this overleaf so the links are attached to the dataset name

- **Free response:** For each section or paragraph for which you used assistance, describe your overall experience with the AI. How helpful was it? Did it just directly give you a good output, or did you have to edit it? Was its output ever obviously wrong or irrelevant? Did you use it to generate new text, check your own ideas, or rewrite text?
  – The answer was pretty close to what I wanted, I just needed to edit it slightly, and it was ready to use. Furthermore, to summarize the paper, it was very helpful.
  – For helping rework the data section it got confused about what data we will use in which stage, so I had to rewrite it myself to correct it. However, it provide useful advice initially about which type of data should be used in each stage and why.

## References


Chan, J. Y.-L., Bea, K. T., Leow, S. M. H., Phoong, S. W., and Cheng, W. K. (2023). State of the art: a review of sentiment analysis based on sequential transfer learning. *Artificial Intelligence Review*, 56:749–780.

Chen, S., Zhang, Y., and Yang, Q. (2024). Multi-task learning in natural language processing: An overview.

Houlsby, N., Giurgiu, A., Jastrzebski, S., Morrone, B., de Laroussilhe, Q., Gesmundo, A., Attariyan, M., and Gelly, S. (2019). Parameter-efficient transfer learning for nlp.

Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., and Chen, W. (2021). Lora: Low-rank adaptation of large language models.

Liu, X., He, P., Chen, W., and Gao, J. (2019). Multi-task deep neural networks for natural language understanding.

Pfeiffer, J., Kamath, A., Rücklé, A., Cho, K., and Gurevych, I. (2021). Adapterfusion: Non-destructive task composition for transfer learning.

Zhang, Y. and Yang, Q. (2022). A survey on multi-task learning. *IEEE Transactions on Knowledge and Data Engineering*, 34(12):5586–5609.

Zhao, L., Li, L., Zheng, X., and Zhang, J. (2021). A bert based sentiment analysis and key entity detection approach for online financial texts. In *2021 IEEE 24th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*, pages 1233–1238.