ClimateGrantLLM: Benchmarking grant recommendation engines for natural language descriptions of climate resilient infrastructure capital projects

Stanford CS224N Custom Project

Bhumikorn Kongtaveelert
Department of Computer Science
Stanford University
bkongtav@stanford.edu

Auddithio Nag
Department of Computer Science
Stanford University
aunag@stanford.edu

Peter Li
Department of Computer Science
Stanford University
peterjl@stanford.edu

Abstract

Discovery of funding opportunities available for climate resilience infrastructure projects pose as a capacity challenge for local government officials, researchers, non-governmental organisations, and community-based organisations developing such projects. Towards that end, we propose a fine-tuned, LLM-based recommendation engine that suggests possible grants given a short description of an infrastructure project, and historical evidence of funding.

1 Key Information to include

- Mentor: Kaylee Burns
- External Collaborators (if you have any): N/A
- Sharing project: N/A

2 Introduction

While multiple federal grant efforts from the Inflation Reduction Act (IRA) and the Infrastructure Investment and Jobs Act (IIJA)/Bipartisan Infrastructure Law (BIL) are poised to catalyze climate action (Saha and Fazeli (2023)), too many small cities, townships, counties and villages in the United States opt out of pursuing these opportunities due to competing priorities and limited staff to scan for Notices of funding opportunities (NOFOs) and apply. (Digital Response (2023)). To improve distribution of funds to understaffed municipalities, knowing which grant to start applying for is critical for a more streamlined process. Therefore, a large language model (LLM) trained on such infrastructure grants and capable of suggesting possible funding opportunities, given a description of the application, would be invaluable for this domain and its practitioners.

Past work using large language models used for climate-related tasks remains limited. Firstly, climate-based corpora remain limited. For instance, Nicolas Webersinke (2022) developed ClimateBERT, a transformer-based language model that is pretrained on climate-related texts scrapped from online resources. On the other hand, past work on utilizing natural language processing (NLP) for grant recommendation systems also remains limited, although there is a growing body of work in using information retrieval system towards practical applications (Shin Kamada (2018a,b); Daniel E. Acuna (2022)). Most recently, Zhu et al. (2023) built a recommendation system, based on Bidirectional Encoder Representations from Transformers (BERT), for the National Institute of Health (NIH) grants using researchers' publications, to allow researchers to access these funds more easily. These work has shown that the problem of information asymmetry pervade across numerous sectors, and a well-trained grant recommendation system could be crucial in mitigating the gap.

Towards our goal for climate infrastructure projects, we are developing a similar text-based grant recommendation engine which returns possible grants for a short description of the project application. Our contributions in this paper are threefold:

- We have, for the first time, compiled an extensive dataset of 30,000 past climate infrastructure projects and their capital sources through Grants.gov. Moreover, we have augmented our dataset by hand through include all possible funding opportunities possible for each grant, thereby enhancing our dataset.
- We have built the first fine-tuned climate grant recommendation engine using BERT and other LLM-based techniques.
- We have benchmarked the performance of our recommendation engines using different approaches, across supervised, unsupervised, multi-shot and zero-shot techniques.

We hope that out work will provide a resource for training and fine-tuning future climate grant recommendation engines, and provide a informative idea of the landscape of model architectures that can be used for this domain.

3 Related Work

Work on LLMs trained on climate-related corpora is limited. As one of the first instances of such an LLM, Nicolas Webersinke (2022) developed ClimateBERT, a transformer-based language model that is pretrained on over 2 million paragraphs of climate-related texts, spanning from news articles to climate reports scraped from online resources, and which has been used to analyze companies' climate-risk disclosures.

On the other hand, there is some past work on utilizing NLP for grant recommendation systems. Grant recommendation engines can be characterised as a multi-label classification task, where each item corresponds to a project application, and each label qualifies as a funding opportunity. In the realm of rule-based learning methods, Shin Kamada (2018a,b) developed a Japanese grant recommender using keywords and association rules between researchers and grants, and further extended the system with TF-IDF technique. Another unsupervised system called EILEEN (Daniel E. Acuna (2022)) also adopted TF-IDF with Latent Semantic Analysis for topic extractions and used Rocchio Algorithm and Random Forest to predict potential matches of grants and publications (Daniel E. Acuna, 2022). With the advent of LLMs, more such possibilities arose. In particular, Zhu et al. (2023) built a recommendation system, based on BERT, for the NIHgrants using researchers' publications, to allow researchers to access these funds more easily. Zhu et al. (2023) used a multi-shot approach, where all of the class labels (here, funding opportunities) were known in advance to the model to be trained on, and each item was annotated with the correct labels.

However, a multi-shot approach may not be the most efficient characterisation of this problem space. Firstly, this domain suffers from the cold start problem in e-commerce: newer grants will have fewer applicant records, and will therefore be less likely to be suggested. Secondly, with the ever-expanding array of funding opportunities in different judicial territories, the label space becomes enormous and intractable for a softmax function (Farhad Pourpanah (2020)). Lastly, obtaining labelled data is laborious and difficult.

Thus, zero-shot multi-label text classification (ZMTC) methods came into the picture. ZMTC approaches can be used to solve the cold start problem (Jingjing Li and Huang (2019); Chang et al. (2021)) and can be deployed to tasks with millions and billions of classes (Tharun Kumar Reddy Medini and Shrivastava (2019); Tharun Medini and Shrivastava (2020); Yang Liu and Schaaf (2021b); Kunal Dahiya and Varma (2021)). ZMTC methods incorporate some form of representation learning, where the content can be cast into a multi-dimensional vector space for the nearest neighbor search problem. A few approaches have been attempted in this sphere. Chalkidis et al. (2020) and

Xiong et al. (2022) divide the structural documents into a set of sentences, and then perform sentence-level representation learning by modeling the pairwise similarity of input and label documents. However, this representation learning method neglects the paragraph-level information and the structural relationship between the title and contents. Other methods utilise trees (N. Gupta and Varma (2021)) or graphs (Chen et al. (2022)) to reduce the computation overhead. However, this procedure results in multi-stage training, which is generally hard to optimize end-to-end for ZMTC training.

Recently, Zhang et al. (2022) developed a way of bypassing some of these issues. In particular, they divided up the structural document into multiple chunks drawn from a uniformly random distribution, and then used contrastive learning to fine-tune an encoder model. In particular, the different permutations of the words in the document acted as positive pairs, and the final inference wss conducted by encoding all the labels into their respective vector representations, and finding the closest label neighbors for the application. Their model outperformed state-of-the-art ZMTC techniques, leading to precision and recall scores of 11.96 and 30.45 at 5 epochs. While their model relies on transformer models including BERT and MPNet, Zhang et al.'s algorithm presents a possibility at increasing LLM performance while not overly increasing the computational power required.

4 Approach

Firstly, we defined the task as *single-label, multi-shot* classification problem, where each project application corresponded to one grant that it was most suited for. Towards that end, we fine-tuned a BERT model for sequence classification (fine-tuning the transformers on the climate-related text, then using a softmax classifier across all known labels), inspired by Zhu et al. (2023). The training dataset was fully labelled for this purpose. The BERT sequence classifier uses a standard categorical cross-entropy loss, mathematically denoted as

$$\mathcal{L} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{C} y_{ij}\log(\hat{y}_{ij}),$$

where: N is the number of samples in the batch. C is the number of classes.y_{ij} is a binary indicator (0 or 1) if class label j is the correct classification for sample i. $\hat{y}_{ij}$ is the predicted probability that sample i belongs to class j.

Secondly, we extended the idea as a *multi-label* classification problem, where each project could be connected to multiple capital sources. In addition, given the vast and ever-growing nature of the label space, we looked at *zero-shot* strategies to tackle this problem. In that vein, following Zhang et al. (2022), we employed the following approach: (1) each long text description into smaller chunks via randomized text segmentation. (2) We constructed positive pairs (more information below) using the pieces of the documents, and used them for a contrastive representation learning task on a pre-trained encoder (here, BERT). (3) For inference, we encoded the application and each of the labels (grants) into their respective embedding vectors, and found the k nearest neighboring label embeddings to assign the top k label recommendations.

The contrastive loss function here is defined as

$$\mathcal{L} = -\frac{1}{N}\sum_{i=1}^{N}\log\frac{e^{f(E(x_i),E(\hat{x}_i))/\tau}}{\sum_{j=1}^{b} e^{f(E(x_i),E(\hat{x}_j))/\tau}}$$

where: N is the batch size. $E(\cdot)$ is the encoder function. x_i is the input sample. ($_i$ is the corresponding positive sample for x_i. τ is the temperature hyperparameter. $f(x,\hat{x}) = \frac{x\cdot\hat{x}}{\|x\|\|\hat{x}\|}$ is the cosine similarity function.

This is a more scalable approach (no need to retrain entire model to add new possible class), to a multi-label one-project-to-many-grants recommendation, compared to softmax-based supervised learning which is better fit for a fixed number of classes.

To generate the positive pairs, we first followed Zhang et al. (2022)'s approach to construct pairs from project-title-to-itself, project-title to project-description (segments), project-description to project-description description, and group grant title and grant description together as 'label', creating label-to-itself pairs. In the second part of our research, we also added grant-title-to-grant-description, project-description-to-grant-title, and project-description to grant-description.

5 Experiments

5.1 Data

We have compiled and augmented data on climate-related infrastructure projects from numerous sources. The following a tabular summary of all the data we have created and used in this paper.

Data type	Sources	Details
Project profiles	BIL Launchpad Successful Projects	3380 unique entries, with title and descriptions.[1] ***Title:*** *"Clearwater Multimodal Transit Center";* ***Content:*** *"Clearwater Multimodal Transit Center: This project in Downtown Clearwater will replace the existing Park Street Terminal with a new, more..."*
	OpenFEMA Dataset: Hazard Mitigation Assistance Projects	33000 entries, with no title and formulaic descriptions. ***Title:*** *"";* ***Content:*** *"Maui205.8: Retrofitting Public Structures - Wind"*
Grant profiles	Grants.gov	Filtered grants from a large archive to match the projects we have information for. ***Title:*** *"Public Transportation on Indian Reservations"* ***Description:*** *"Public Transportation on Indian Reservations Program: The Federal Transit Administration (FTA) announces...*
	Hazard Mitigation Assistance Grants | FEMA.gov	5 grants corresponding to the OpenFEMA Dataset ***Title:*** *Hazard Mitigation Assistance Program (HMGP)* ***Content:*** *"Hazard Mitigation Assistance Program (HMGP): Developing or updating a FEMA approved mitigation plan to help state, local, tribal, and terri*

We have preprocessed the data to remove any duplicates, since some grants have duplicates, or are reoccurring over several years of funding. We have simplified this by omitting the years to mitigate spurious factors in the model inference and picked the first description of a grant if it was reoccuring across fiscal years.

In addition, we have augmented the dataset by clustering grant title and description using word2vec to put semantically similar grant opportunities together, then reassigning each project to alternative class(es) within each cluster other than the original true grant it was paired with. This artificially allowed us to move from one-to-one project-grant pairs, to a one-to-many project-grant network, resembling how in reality one project may be eligible for a few grants with similar objectives. In future work, this could be based on human-feedback of experts in the area.

For the final results, we experimented with combinations of the different datasets: using the original data, using augmented data, and then a combination of both: **"Baby" dataset** (˜1100 entries, with 0.7:0.3 train-test split. [2] **Rebalanced dataset** (3700 entries, with 0.85:0.15 train-test split),[3] **Original dataset** (35000 entries, with 0.85:0.15 train-test split) and **original + synthetic dataset** (70000 entries, with 0.85:0.15 train-test split.)

[2]This omitted multiple classes during training.

[3]For this dataset, we significantly omitting around 99% of data from the following labels ['Hazard Mitigation Assistance Program (HMGP)', 'Flood Mitigation Assistance (FMA) grant program', 'Pre-Disaster Mitigation (PDM) Grant Program', ''Building Resilient Infrastructure and Communities (BRIC) grant program', 'Severe Repetitive Loss (SRL) Grant Program'], so they contribute around 300 entries, as supposed to disproportionately representing the dataset by >30000 entries in total.

5.2 Evaluation method

We adopt precision at k or $P@k$, and recall at k or $R@k$ for $k \in \{1, 3, 5\}$ as the evaluation metrics for modelling tasks. $P@k$ measures the accuracy of the top k results returned by a system, formulaically defined as

$$P@k = \frac{\sum_{i=1}^{n} \sum_{y \in Y_i^{\text{pred}}} \mathbf{1}_i^{\text{top-}k}(y)}{nk},$$

where n is the number of documents evaluated, Y_i^{pred} is the set of predicted labels for the i-th document, and $\mathbf{1}_i^{\text{top-}k}(\cdot)$ is an indicator function indicating whether a predicted label is a ground truth top-k label for the i-th document.

$R@k$ measures how well the top k results cover all possible relevant results, i.e. what proportion of the true pairs appear within what the algorithm determined to be the k most relevant results. Mathematically, it is defined as

$$R@k = \frac{1}{n} \sum_{i=1}^{n} \frac{\sum_{y \in Y_i^{\text{pred}}} \mathbf{1}_i^{\text{top-}k}(y)}{\sum_{y \in Y} \mathbf{1}_i^{\text{top-}k}(y)},$$

In addition, we make qualitative assessments to verify the output of our model. We inspect $R@5$ by label class, where we try to understand overarching trends, including possible overfitting. We also inspect changes in $R@k$ over the training period to determine if metrics improved over time or stagnated early on, which would in turn help us gauge the degree of fine-tuning we need.

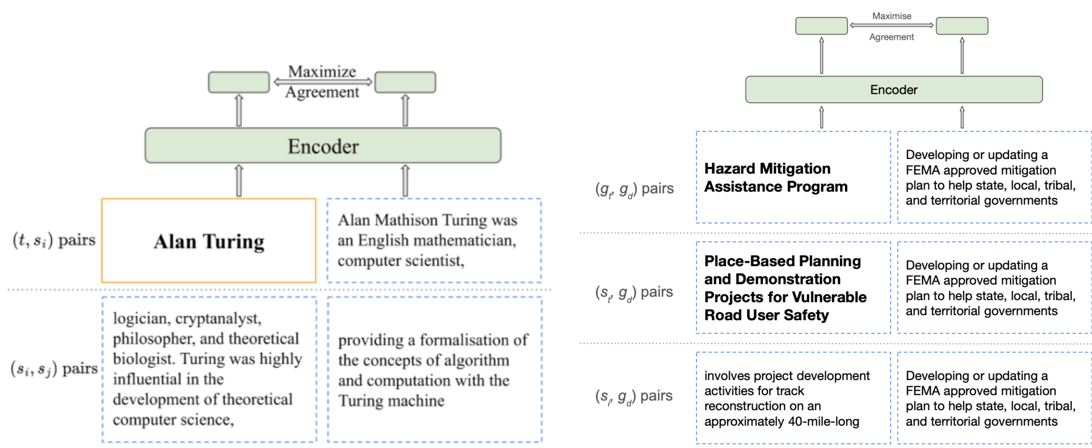


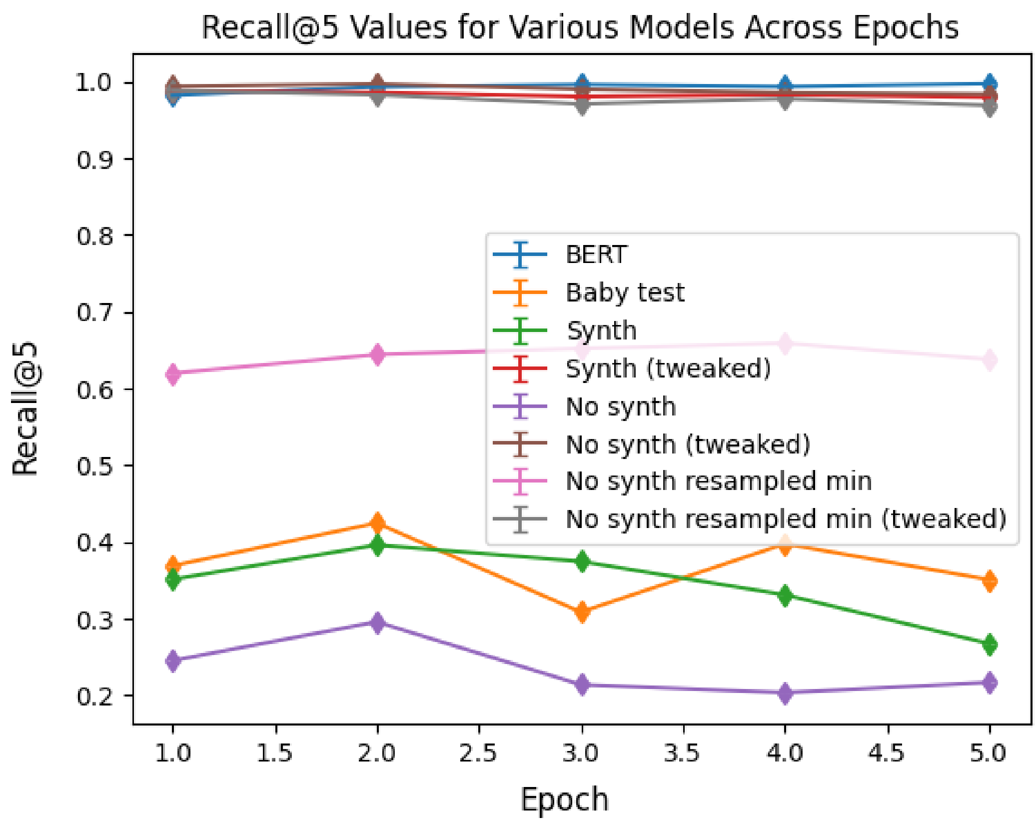
Figure 1: (Left) Zhang et al. (2022)'s approach in generating positive pairs for contrastive learning. (Right) Our approach in contrasting positive pairs.

5.3 Experimental details

- **BERT-based Sequence Classifier**: Learning rate = 1e-5, AdamW optimizer, 5 epochs, batch size = 3 (Colab GPU), BERT-based sequence classifier model and tokenizer. This model was trained on a list of BIL Launchpad's successful projects that we preprocessed, which are categorized by project profile names. The label/target index of each project profile corresponds to the grant that project is matched to.
- **Structured Contrastive Learning**: Learning rate = 5e-5, Adam optimizer, 5 epochs, batch size = 30 or 60 (depending on available RAM on Colab GPU), BERT-based uncased pretrained model, temperature parameter = 1.5. In this section, we have 8 experiments: varying different training 4 dataset compositions, and either Zhang's Structured Contrastive Learning or our tweaked method, which included more positive pairs for contrastive learning.

5.4 Results

Approaches	**Precision @ k**	**Recall @ k**
BERT-based sequence classification	P@1: 0.6623	R@1: 0.9818
	P@3: 0.7064	R@3: 0.9959
	P@5: 0.7195	R@5: 0.9970
Baby-test	P@1: 0.0572	R@1: 0.0572
	P@3: 0.0462	R@3: 0.1355
	P@5: 0.0379	R@5: 0.1852
Baby-test (tweaked; more positive pairs)	P@1: 0.4518	R@1: 0.4398
	P@3: 0.2570	R@3: 0.7455
	P@5: 0.1723	R@5: 0.8328
Rebalanced	P@1: 0.1111	R@1: 0.1111
	P@3: 0.0608	R@3: 0.1814
	P@5: 0.0500	R@5: 0.2483
Rebalanced (tweaked; more positive pairs)	P@1: 0.7691	R@1: 0.7674
	P@3: 0.3171	R@3: 0.9479
	P@5: 0.1962	R@5: 0.9774
Original (no synthetic)	P@1: 0.1121	R@1: 0.1088
	P@3: 0.0811	R@3: 0.2334
	P@5: 0.0622	R@5: 0.2959
Original (no synthetic) (tweaked; more positive pairs)	P@1: 0.7759	R@1: 0.7429
	P@3: 0.3429	R@3: 0.9552
	P@5: 0.2138	R@5: 0.9853
Original + Synthetic	P@1: 0.0428	R@1: 0.0394
	P@3: 0.0382	R@3: 0.1014
	P@5: 0.0340	R@5: 0.1512
Original + Synthetic (tweaked; more positive pairs)	P@1: 0.7802	R@1: 0.7462
	P@3: 0.3406	R@3: 0.9491
	P@5: 0.2121	R@5: 0.9795



The table in the previous page shows the precision and recall values we obtained from our experiments, while the graph above shows Recall@5 values across 5 epochs for different models we tested. Please see section 6 for qualitative analyses of the results.

6 Analysis

There are a few salient observations to be gleaned from our results. Firstly, many of our precision and recall values are comparable to state of the art results Zhang et al. (2022); Xiong et al. (2022). In particular, our recall@k test scores reach relatively high levels for $k \leq 5$, indicating that model is able to capture real world examples well. While our precision@k scores may seem low, there is a caveat explaining this phenomenon. Most of the information scraped from online resources report one grant awarded to each possible application. Therefore, this imposes an upper ceiling of $\frac{1}{k}$ for the precision@k scores, which explains the seeming deflation. Given the sparse nature of our datasets, precision@k is probably not a helpful metric of our performance at this stage, but could be helpful once we gained more one-project-to-many-grant labels.

Incorporating the relationship between query and label dramatically improves the ability for long-text comprehension and embedding for BERT-based transformers. This is evident by how the tweaked method (generating alternative pairs for contrastive learning using our labelled data) performed much better on all metrics across all datasets. This is likely because while it is taking advantage of the unsupervised learning to improve embedding representation of long sentence descriptions, which Zhang et al. (2022) provides, the tweaked-method to structured contrastive learning for ZMTC is also learning to place embedding of projects to cluster around their corresponding grant titles and implicitly learn to cluster.

Interestingly, un-augmented data has the best $R@k$ scores. It is likely that the addition of synthetic data or removal of overrepresented data does not meaningfully change the $R@k$ score, when compared to original data with augmentation. However, this might be due to the face that there is meaningfully more weight placed on the overrepresented grant types, so misclassification in other classes does not meaningfully affect the overall score.

For the experiments with zhang's Structural Contrastive Representation Learning for Zero-shot Multi-label Text Classification, the top 4 most represented categories in original+synthetic and original dataset 'Hazard Mitigation Assistance Program (HMGP),' 'Pre-Disaster Mitigation (PDM) Grant Program', 'Flood Mitigation Assistance (FMA) grant program,' 'Building Resilient Infrastructure and Communities (BRIC) grant program' consistently underperforms (average recall@5 is below 0.7, sometimes below <0.01). This might be because title of these grants (e.g. "Hazard Mitigation Assistance Program (HMGP)") are drastically different from the short description and format associated with each of these FEMA-related projects ("Maui205.8: Retrofitting Public Structures - Wind"), so when Zhang-method fine tune the label and the description of query separately, due to different semantic structures, the fine-tuned BERT-based embeddings actually drifted apart, worsening the performance of nearest neighbor classification. This would explain why the tweaked method performs better, because we are establishing relationship between the unique format of FEMA-related projects by generating positive pairs between the project content and title with the grant opportunity content.

The tweaked-method worked well across all classes, except 1-2 classes depending on the dataset. Instances where it underperforms (recall@5 is below 0.7) can likely be attributed to a lack of training and testing samples. For example, in the original dataset with tweaked-methods, 'Rural Surface Transportation Grant Program' was the only class thatthe model underperformed. This class is severely underrepresented (showed up twice in the testing set) and 10 times in training. Instead, the recommendation engine recommended the overrepresented first few classes (e.g.Hazard Mitigation Assistance Program (HMGP)) and "Safe Streets and Roads for All Funding Opportunity," which is similar grant type to 'Rural Surface Transportation Grant Program'

We also examined how the scores varied with each epoch over the training regime to understand the possible computational overhead of fine-tuning. It seems most such models plateau on the second epoch of fine-tuning on climate-related texts.

7 Conclusion

For the first time, we have aggregated and processed an extensive dataset of >35,000 past climate infrastructure projects, developed a climate grant recommendation engine using BERT sequence classification and LLM-based techniques, and benchmarked the performance across numerous established metrics. Our results indicate that LLMs can indeed provide a solution to the recommendation engines for this domain, with minimal computation overhead for fine-tuning. We have shown, in the presence of labelled data, that our unique augmentations for the generation of positive pairs in the zero-shot contrastive learning framework can provide high recall scores, taking advantage of baseline sequence classification but provides flexibility for encountering new labels without having to finetune the model from scratch. Even in the absence of labelled data, our model provides results comparable or better than state of the art models for this specific domain.

There are a few limitations to our work. Firstly, our dataset contains a disproportionate representation of each label and possible applications. Applications from certain grants recur more frequently over the dataset, repeated over several years, whereas others are fewer, as sparse as one per dataset. This may skew our results, particularly by overfitting to the more popular grants and applications, but can be mitigated by constructing positive pairs across application-grant labels. Secondly, our architecture still rewards the usage of labelled datasets, which might be more difficult to obtain when deployed into the real world, but the zero-shot multi-label classification is promising. Lastly, handling the datasets from grants often involved very careful and precise data handling, which might be intractable at scale.

A few options lie ahead in the future. Aside form having a standardized repository for this information, more zero-shot unsupervised/semi-supervised methods should be explored to understand how to better devise grant recommendation engines. There should be greater systematic examination into the individual grants/labels and performances across them to improve semantic understanding the transformers has over this domain, to identify which particular grants are under and over-prescribed by our model for these capital projects. Lastly, human-feedback will be critical to construct alternative pairs and negative pairs to further improve a zero-shot recommendation tasks in the future iterations.

8 Ethics Statement

Firstly, one potential ethical challenge that may arise with poor performance of the model. If implemented, it may cause large scales of financial and/or environmental loss e.g. poor recommendation may delay start of grant funding process amongst underserved communities if adopted. However, we believe that this is not greater than the degree of inefficiency that persists in the climate infrastructure ecosystem for this segment; therefore our proposed technology is still a net good in this regard.

Secondly, the usefulness of the model is only as good as its input. The maintenance of up-to-date data on new successful (and also hopefully unfit, to construct negative pairs) projects and upcoming grants is critical for holistic recommendation. Similarly, the sourcing of human-feedback from overworked resilience officers or other users, most of them serving medium to small cities, has to be done with caution, transparency and respect to user's privacy about the different ways their data is potentially recorded and how they will be utilised, as well as maintaining safe encryption of personal/key information if any is recorded.

Thirdly, as a climate-related project, we acknowledge that the use of transformer-based large language models pose enormous carbon footprints on the planet. We hope that, upon its public release, the benefits towards expanding climate infrastructure project offsets and ultimately cancels out the cost that this incurs. We also pivoted towards a zero-shot multi-label classification model, because the model is more suited to evolving number of grants and infrasturcture projects, without having to retrain regularly to incorporate new classes.

9 Contributions

BK was responsible for project conception, data scraping and cleaning, implementation of baseline BERT sequence classification model, generating synthetic data using nearest neighbor clustering of embedding, adapting the Zhang et al. 2022 model for semi-supervised contrastive learning with alternative positive pair generation, qualitative and quantitative analysis, and conclusion. AN was responsible for the project conception, literature review, implementation of the word2vec baseline, and qualitative analyses, conclusion and ethics statement. PL was responsible for the implementation of BERT sequence classification models, researching the usage of various large language models, and creating diagrams, and related analyses. All authors have equal contribution to the report and poster.

References

Ilias Chalkidis, Manos Fergadiotis, Sotiris Kotitsas, Prodromos Malakasiotis, Nikolaos Aletras, and Ion Androutsopoulos. 2020. An empirical study on large-scale multi-label text classification including few and zero-shot labels. pages 7503–7515.

Wei-Cheng Chang, Daniel Jiang, Hsiang-Fu Yu, Choon Teo, Jiong Zhang, Kai Zhong, Kedarnath Kolluri, Qie Hu, Nikhil Shandilya, Vyacheslav Ievgrafov, Japinder Singh, and Inderjit Dhillon. 2021. Extreme multi-label learning for semantic matching in product search. pages 2643–2651.

Qi Chen, Wei Wang, Kaizhu Huang, and Frans Coenen. 2022. Zero-shot text classification via knowledge graph embedding for social media data. *IEEE Internet of Things Journal*, 9(12):9205–9213.

Priya Matnani Daniel E. Acuna, Kartik Nagre. 2022. Eileen: A recommendation system for scientific publications and grants. Online. arXiv.

U.S. Digital Response. 2023. No front door: Barriers to access for small communities seeking federal grants.

Yuxuan Luo Xinlei Zhou Ran Wang Chee Peng Lim Xi-Zhao Wang Q. M. Jonathan Wu Farhad Pourpanah, Moloud Abdar. 2020. A review of generalized zero-shot learning methods. arXiv.

Grants.gov. Grants.gov: Home.

Ke Lu Lei Zhu Yang Yang Jingjing Li, Mengmeng Jing and Zi Huang. 2019. From zero-shot learning to cold-start recommendation. In *AAAI conference on artificial intelligence*, volume 33, pages 4189–4196.

Anshul Mittal Ankush Shaw Kushal Dave Akshay Soni Himanshu Jain-Sumeet Agarwal Kunal Dahiya, Deepak Saini and Manik Varma. 2021. Deepxml: A deep extreme multi-label learning framework applied to short text documents. pages 31–39. 14th ACM International Conference on Web Search and Data Mining.

Y. Prabhu S. Purohit N. Gupta, S. Bohra and M. Varma. 2021. Generalized zero-shot extreme multi-label learning. ACM SIGKDD Conference on Knowledge Discovery and Data Mining.

Julia Anna Bingler Markus Leippold Nicolas Webersinke, Mathias Kraus. 2022. Climatebert: A pretrained language model for climate-related text. Online. arXiv.

Devashree Saha and Sandy Fazeli. 2023. How the us federal government can help states and cities capitalize on historic climate funding.

Takanobu Watanabe Shin Kamada, Takumi Ichimura. 2018a. Recommendation system of grants-in-aid for researchers by using jsps keyword. Online. arXiv.

Takanobu Watanabe Shin Kamada, Takumi Ichimura. 2018b. A recommendation system of grants to acquire external funds. Online. arXiv.

Yiqiu Wang Vijai Mohan Tharun Kumar Reddy Medini, Qixuan Huang and Anshumali Shrivastava. 2019. Extreme classification in log memory using count-min sketch: A case study of amazon search with 50m products. Advances in Neural Information Processing Systems.

Beidi Chen Tharun Medini and Anshumali Shrivastava. 2020. Solar: Sparse orthogonal learned and random embeddings. In International Conference on Learning Representations.

Yuanhao Xiong, Wei-Cheng Chang, Cho-Jui Hsieh, Hsiang-Fu Yu, and Inderjit Dhillon. 2022. Extreme Zero-Shot learning for extreme text classification. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 5455–5468, Seattle, United States. Association for Computational Linguistics.

Russell Klopfer Matthew R Gormley Yang Liu, Hua Cheng and Thomas Schaaf. 2021b. Effective convolutional attention network for multi-label clinical document classification. pages 5941–5953. 2021 Conference on Empirical Methods in Natural Language Processing.

Tianyi Zhang, Zhaozhuo Xu, Tharun Medini, and Anshumali Shrivastava. 2022. Structural contrastive representation learning for zero-shot multi-label text classification. In *Findings of the Association for Computational Linguistics: EMNLP 2022*, pages 4937–4947, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Jie Zhu, Braja Gopal Patra, Hulin Wu, and Ashraf Yaseen. 2023. A novel nih research grant recommender using bert. Online. National Library of Medicine.