# Extending Generative NLP: Incorporating Diversity, Context, and Inclusivity in Neural Text Generation

Kaushal Kumar Maurya

CS18RESCH11003

A Synopsis Submitted to

Indian Institute of Technology Hyderabad

In Partial Fulfillment of the Requirements for

The Degree of Doctor of Philosophy

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Department of Computer Science and Engineering

Advisor: Dr. Maunendra Sankar Desarkar

Dept. of Computer Science and Engineering

August, 2024

# Introduction

In recent years, Deep Learning [1] has witnessed remarkable advancements, revolutionizing the landscape of artificial intelligence and machine learning. These breakthroughs are primarily attributed to the development of sophisticated neural network architectures, increased computational power, and the availability of large datasets. It has a profound impact on many real-life applications, from Natural language processing (NLP) to Computer Vision and many more. The primary driving force for such remarkable advancement in NLP research has been propelled by large pre-trained language models (PLMs; [2, 3]) based on self-supervised training objectives [3]. These PLMs are developed on top of transformer neural network [4] and have millions or billions of parameters. They undergo training on large monolingual data for thousands of compute hours, yielding high-quality *out-of-box* representations. Subsequently, these models can be fine-tuned for downstream tasks, leading to superior accuracy.

However, there is a notable disparity in NLP research, with the majority of studies being conducted on English data [6, 5], despite the fact that the vast majority of the global population, approximately 95%, do not speak English as their primary language, and a staggering 75% does not speak English at all[1]. Further, there are around 7,000 spoken languages, with approximately 400 languages having over 1 million speakers and about 1,200 languages having more than 100,000 speakers [7]. Fig. 1 illustrates the six-class distribution of language resources, and it can be observed that class 0 has a large set of languages with no or limited resources. Notably, 88% of the world's languages fall into class 0, spoken by 1.2 billion people, and are untouched by the benefits of language technology. A study presented at ACL 2008 [8] revealed that 63% of all papers focused only on English. A more recent study during ACL 2021 [9] concluded that nearly 70% of the papers were evaluated on English datasets. 10 years of progress and language coverage in NLP research is almost unchanged due to the limited availability of datasets for low-resource languages (LRLs), aka. the long tail of languages. To put it succinctly, the scarcity of data, lack of linguistic tools and resources, and absence of representation from PLMs [10] leads to performance gaps or hinders advancement of language technology for LRLs[2].

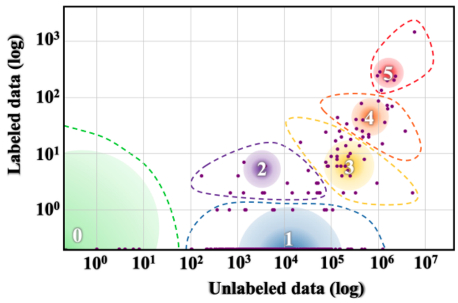


Figure 1: Language resource distribution illustrated by Joshi et al. [5] for six classes. The gradient circle represents the number of languages. Color spectrum VIBGYOR, represents the total speaker population.

This thesis is a step towards enabling language technologies for tailored low-resource languages (LRLs) characterized by limited available data. The primary focus is on natural language generation (NLG), a field concerned with *the automated generation of human-like text from a given input context.* The context can be a nat-

[1] https://en.wikipedia.org/wiki/List_of_languages_by_total_number_of_speakers
[2] https://labs.theguardian.com/digital-language-divide/?ref=ruder.io

ural language text, an image, a video, etc. NLG consists of a wider range of tasks, including machine translation, abstractive text summarization, headline generation, question generation, and many more. The issue of data scarcity is more pronounced for NLG tasks as task-specific data availability for LRLs is even rare. Current multilingual language models support only around 100 languages [11]. Moreover, their adaptability to various generative applications, even for 100 languages, poses a significant challenge [12]. The scope of the thesis extends beyond text generation in LRLs and extends NLG technology to two other frequently observed scenarios: *diverse text generation* and *text generation with limited context*.

**Definition (Low-resource Language):** *It is hard to define what low-resource language is. As a "resource-ness" continuum and criteria must be arbitrary. With the ability to crawl and gather more data, some languages are no longer considered resource-poor. To establish a clear definition, we consider research focused on "low-resource languages" to be those that propose methods to overcome the lack of data in languages [13].*

**Thesis Objectives:** This thesis aims to advance generative NLP by addressing three key challenges: *diverse text generation*, *text generation with limited context*, and *text generation with limited available data*, often case for LRLs.

**Thesis Outline:** We summarize our research efforts to extend generative NLP in Fig. 2. This thesis is divided into two parts: *advancing the frontier of NLG* and *enabling low-resource language generation*. In the first part, we extend generative NLP to an application where diverse text needs to be generated, i.e., the distractor generation task [14]. It is the task of generating multiple incorrect options for a given Multiple-Choice Question (MCQ) for reading comprehension. Next, we mitigate issues of limited context in personalized query completion [15] with Retrieval-Augmented Generation (RAG; [16]) type of modeling. The second part focuses on enabling language technology for LRLs, which has limited available data. We first mitigate the issues of catastrophic forgetting problems in cross-lingual modeling to enable well-formed zero-shot generation in LRLs [17]. Next, on top of [17], we explore a meta-learning approach [18] to transfer cross-lingual supervision from HRL to LRLs uniformly. This improves cross-lingual transfer, even for LRLs that exhibit less similarity with HRL in the zero-shot setting. Finally, we propose two novel noise augmentation approaches [19, 20] to enable zero-shot machine translation for extremely LRLs.

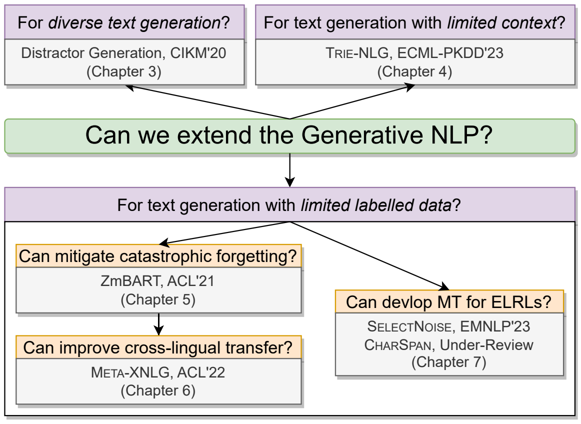


Figure 2: An overview of our research efforts.

These efforts have been grouped into five threads. Next, we will briefly explain each of them.

# 1 Advancing Frontiers of NLG: Distractor Generation

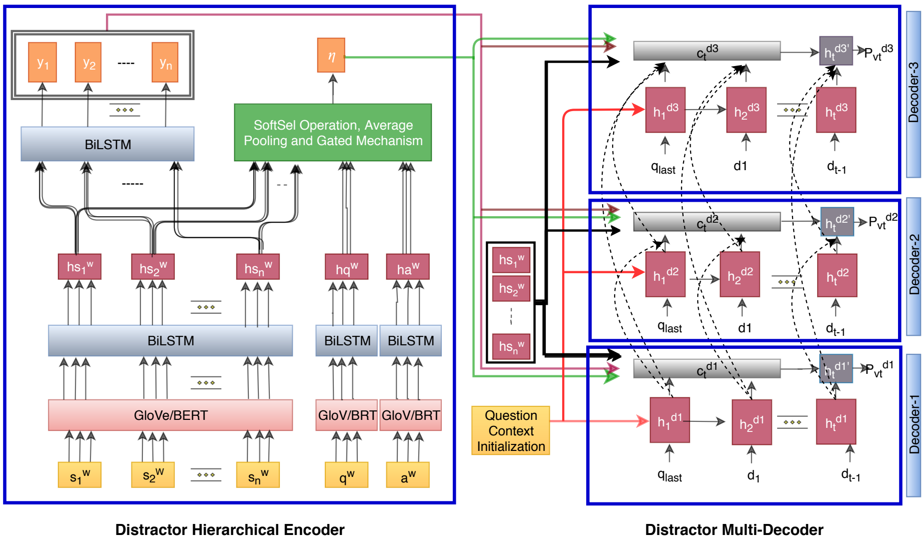


Figure 3: Architectural diagram of the proposed HMD-Net Model [14].

*Diverse text generation* in NLG is crucial for real-world applications. For example, diverse headlines for input news articles can give the author/publisher few options to select and broaden the audience's engagement [21]. Similarly, generating diverse responses for frequently occurring monotonous contexts in a dialogue system enhances user engagement. Another application is the distractor generation (DG), which involves generating multiple distractors or incorrect options for a given Multiple-Choice Question (MCQ) for reading comprehension, i.e., a triplet of `<passage, question, correct answer>`. This capability is particularly valuable for saving time for course instructors or Intelligent Tutoring Systems (ITS; [22]) for creating diverse incorrect options. The task of distractor generation is one of the problems addressed in the thesis. The ideal distractors should possess the following properties: (i) contextual relevance to the question, (ii) semantic dissimilarity to the answer, (iii) diversity from each other, and (iv) confusion-inducing. Additionally, diverse text generation necessitates a one-to-many setup, where the model generates multiple distractors, all derived from a single input triplet.

Considering these, we propose a novel Hierarchical Encoder-Decoder, LSTM-based neural network (HMD-Net; [14]) model for the distractor generation. On the encoder side, it employs *softsel* and *Gated Mechanism* to identify candidate (decouple) sentences in the input passage that are not semantically similar to the answer and maintain context with the question. The decoder conditions on these candidate sentences, the question statement, and the correct answer in a multi-decoder setup (interconnected with each other) to generate lexically diverse and confusing multiple distractors. An overview of HMD-Net is presented in Fig. 3. We further utilize linguistic features and BERT contextual token embedding representations to boost the model's performance. We prepared a new DG dataset from the existing RACE MCQ dataset.

## 2 Advancing Frontiers of NLG: Personalized Query Auto-Completions

Query formulation can be time-consuming for naive or users with complex information needs. Modern search engines, therefore, have a Query Auto-Completion (QAC) module to assist users in efficiently expressing their information needs as a search query. Due to advancements in NLG, QAC is formulated as an NLG problem, which involves generating top-$m$ completions given the user-specific `<query prefix, session>`. The session contains personalized data - previously typed queries - making them Personalized QAC (PQAC) systems. While research in query completion spans over many decades, the challenge of *limited context*, particularly for short and unseen prefixes, persists. Short prefixes typically consist of a few characters and unseen prefixes are those which have never been recorded previously (new query prefixes typed by the user). The traditional Trie-based model [23] offers the most popular completions (MPC) for short prefixes and provides no completions for unseen prefixes. Modern NLG-based models overcome the limitations by generating completions for unseen prefixes. However, since short prefixes have few characters and unseen prefixes are rarely typed, their context in the session is limited, leading to poor and non-relevant completions.

To address the issue of limited context in PQAC, we leverage insights from both Trie and NLG and proposed TRIE-NLG model [15]. In TRIE-NLG, we first provide a quantitative analysis to motivate the need for incorporating both popularity signals from the trie and personalization signals from session queries for effective QAC. Then, we create two tries: $\text{MPC}_{Main}$ and $\text{MPC}_{Synth}$ for short and unseen prefixes, respectively. Finally, we explore the Retrieval-Augmented Generation (RAG; [16]) type of framework to augment top completions from the tries and fine-tune a pre-trained language model. To the best of our knowledge, this is the first attempt of trie knowledge augmentation in NLG models for personalized QAC. We have achieved state-of-the-art performance on two real *prefix-to-query* click behavior QAC datasets from Bing and AOL. Fig. 4 presents sample generations with TRIE-NLG for the short and unseen prefixes.

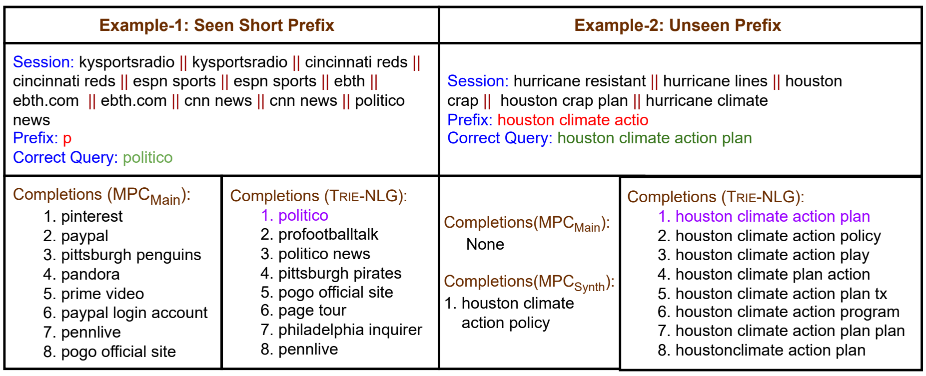


Figure 4: Sample generations with TRIE-NLG. Here, we consider two examples: a seen short prefix and an unseen prefix. Session queries are separated with '||'.

# 3 Mitigating Catastrophic Forgetting to Enable Zero-Shot Cross-Lingual Generation

The remarkable progress in NLP is primarily driven by large annotated datasets. However, most low-resource languages (LRLs) lack such annotated datasets. To address this issue, cross-lingual transfer has emerged as a popular technique for enabling language technology in LRLs with limited supervision or *limited annotated data*. In this modeling approach, a multilingual pre-trained language model (mPLM) is trained with large task-specific data in high-resource languages (HRL) and then evaluated for the task on unseen LRLs (zero-shot) or on LRLs with limited examples (few-shot). This modeling regime transfers supervision from HRLs to LRLs, extending language technology in many LRLs. For example, let us consider a sentiment analysis task. Initially, a mPLM is trained using a large HRL dataset, often in English, where the input consists of English sentences, and the target is the class label. Subsequently, when a sentence in LRL (e.g., in Hindi) is fed, the model generates an appropriate class label.

This modeling recipe presents additional challenges for the NLG tasks. One challenge is the issue of *catastrophic forgetting*. In the zero-shot evaluation phase of the NLG task, the text must be generated in LRLs. For instance, the abstractive text summarization model (developed with the above recipe) is expected to generate a zero-shot summary in LRL when given an LRL input article. However, it is observed that [11] the zero-shot generations are in HRL or code-mixed with HRL and LRL. This occurs because the model forgets the multilingual pre-training, which is referred to as the catastrophic forgetting/off-target/accidental translation problem. This issue has impeded the extension of existing NLG techniques to a wide range of LRLs.

To address the *catastrophic forgetting* problem and enable well-formed zero-shot generation in LRLs, we propose an unsupervised cross-lingual framework - ZmBART [17]. ZmBART is developed on top of the mBART pre-trained language model and does not require parallel data/pseudo-parallel or back-translated data. This framework employs (1) intermediate unsupervised adaptive training, (2) freezing the model component (inspired by continual learning approaches), and (3) adding language tags. Adaptive unsupervised training is done with novel auxiliary task that requires only small monolingual data from LRLs. Here, we consider four NLG tasks and three typologically diverse languages. The proposed approach is scalable to multiple NLG tasks (as the model does not modify any hyper-parameter values across the tasks) and LRLs (operates in a zero-shot setting). Additionally, we have also created HiDG, a high-quality distractor generation dataset for the Hindi language.

# 4 Meta-Learning Approach to Improve Zero-Shot Cross-Lingual Transfer and Generation

Another major challenge in cross-lingual modeling is *uneven supervision transfer*. As supervision is transferred from HRL to LRLs, for LRLs that are similar (close) to the

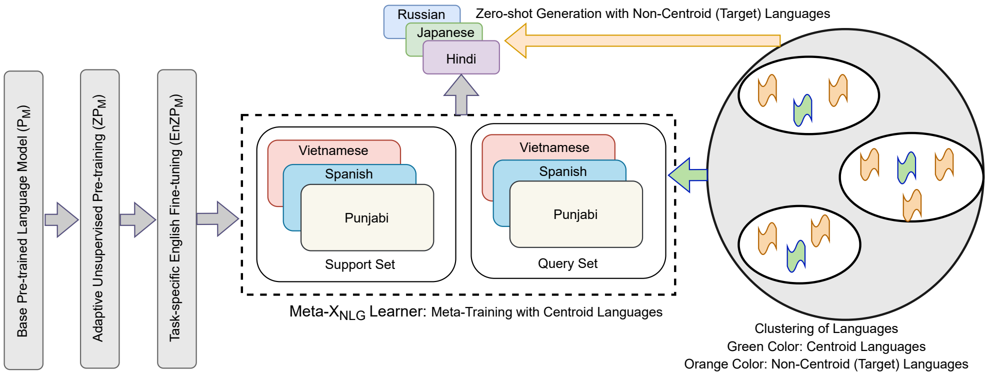


Figure 5: An overview of Meta-$X_{XNLG}$ framework

considered HRLs, the transfer strength is high. However, for languages that are less similar to HRL, the transfer is often weak. This creates issues of uneven supervision transfer, which directly impacts the zero-shot performance for LRLs; that is, the better the supervision transfer, the better the performance, and vice versa.

To address this issue, we propose a novel cross-lingual transfer and generation framework, Meta-$X_{NLG}$[18], based on *Model-Agnostic Meta-Learning (MAML)*, and *language clustering*. In Meta-$X_{NLG}$, we first cluster languages and identify the centroid language for each cluster. Subsequently, the MAML algorithm is trained using centroid languages and evaluated with non-centroid languages in a zero-shot setting. Training with a single centroid language facilitates *intra-cluster* generalization, while training with multiple centroid languages enables *inter-cluster* generalization. This way, the proposed approach exhibits more uniform cross-lingual transfer. The framework is developed on top of the mBART model. It is the first attempt, to the best of our knowledge, to explore meta-learning techniques for cross-lingual NLG. We evaluate the model's performance across two NLG tasks, 30 LRLs, and 5 popular datasets. The proposed model outperformed all the strong baseline models across most of the LRLs, indicating the impact of uniform cross-lingual transfer.

# 5 Zero-Shot Machine Translation for Extremely Low-resource Languages

There are approximately 7,000 spoken languages worldwide, ranging from HRLs like English to LRLs such as Hindi or Japanese. Within the spectrum of LRLs, there exists a large subset known as Extremely Low-Resource Languages (ELRLs), exemplified by languages like Bhojpuri or Sundanese. These ELRLs face unique challenges in the development of NLG applications. Unlike HRLs and some LRLs, ELRLs lack parallel or pseudo-parallel data, have limited monolingual resources, and are not represented in multilingual pre-trained language models. This scarcity of learning resources makes the task of developing NLG applications for ELRLs more challenging.

With this effort, we enable NLG technology for extremely low-resource languages (ELRLs). Specifically, we have addressed the task of *ELRLs to English* machine translation (MT) by utilizing surface-level lexical similarity between *closely related* ELRLs and HRL. There are many ELRLs that are lexically similar to HRLs; for example, Bhojpuri is lexically similar to Hindi. The training instance for ELRL (say, $E_i$) is a noisy version of the related HRL instance (say, $H_i$). In other words, $E_i = \eta(H_i)$ where $\eta$ is noise function. We propose two novel noise augmentation strategies and apply them to the source side (i.e., HRL) of HRL-to-English parallel data to obtain noisy proxy training data for the ELRL-to-English MT task. The noise augmentation in HRL improves lexical similarity between HRL and ELRLs. We learn the vocabulary, train a stranded transformer neural network with this augmented data, and perform the evaluation in a zero-shot setting with ELRLs. Noise augmentation acts as a regularizer to account for lexical variances between HRL and ELRLs and improve cross-lingual transfer. Fig. 6 illustrates a sample example where the noise argumentation in HRL (Hindi) improves the lexical similarity with ELRLs (Bhojpuri).

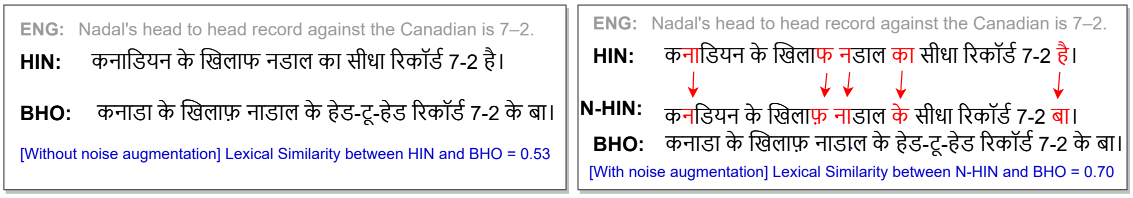


Figure 6: Noise argumentation improves the lexical similarity. The character level lexical similarity computed with the longest common subsequence ratio (LCSR; [24]). ENG (HRL): English, HIN (HRL): Hindi, N-HIN (HRL): Noisy Hindi and BHO (ELRL): Bhojpuri.

We have proposed two novel noise augmentation approaches: (i) CharSpan [19]: This approach randomly augments character span noise and does not require any training resources in ELRLs other than alphabets. (ii) SelectNoise [20]: In this approach, noise augmentation character candidates are extracted with Byte Pair Encoding (BPE) merge operations and edit operations. Decoding algorithms are then used for noise augmentation. This approach is systematic and linguistically inspired but requires small monolingual data (1000 examples) in ELRLs. These models are evaluated with multiple ELRLs across different typologically diverse language families. They outperformed strong baselines by large margins and emerged as state-of-the-art models for ELRLs to English machine translation.

# 6 Publications

## Included in Thesis

1. **Kaushal Kumar Maurya**, and Maunendra Sankar Desarkar."Learning to distract: A hierarchical multi-decoder network for automated generation of long distractors for multiple-choice questions for reading comprehension." Proceedings of the 29th ACM international conference on information & knowledge management. **CIKM** 2020.

2. **Kaushal Kumar Maurya**, Maunendra Sankar Desarkar, Yoshinobu Kano, and Kumari Deepshikha."ZmBART: An Unsupervised Cross-lingual Transfer Framework for Language Generation." In Findings of the Association for Computational Linguistics: **ACL** 2021, pages 2804–2818, Online.

3. **Kaushal Kumar Maurya** and Maunendra Desarkar. "Meta-XNLG: A Meta-Learning Approach Based on Language Clustering for Zero-Shot Cross-Lingual Transfer and Generation." In Findings of the Association for Computational Linguistics: **ACL** 2022, pages 269–284, Dublin, Ireland.

4. **Kaushal Kumar Maurya**, Maunendra Sankar Desarkar, Manish Gupta, and Puneet Agarwal. "Trie-NLG: trie context augmentation to improve personalized query auto-completion for short and unseen prefixes." Joint European Conference on Machine Learning and Knowledge Discovery in Databases (Data Mining and Knowledge Discovery). **ECML-PKDD (DAMI)** 2023.

5. Maharaj Brahma **Kaushal Kumar Maurya**[3], and Maunendra Desarkar. "SelectNoise: Unsupervised Noise Injection to Enable Zero-Shot Machine Translation for Extremely Low-Resource Languages." In Findings of the Empirical Methods in Natural Language Processing: **EMNLP** 2023, Singapore.

6. **Kaushal Kumar Maurya**, Rahul Kejriwal, Maunendra Desarkar, and Anoop Kunchukuttan. 2024. CharSpan: Utilizing Lexical Similarity to Enable Zero-Shot Machine Translation for Extremely Low-Resource Languages. In Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics: **EACL** 2024, Malta.

## Not Included in Thesis

1. Sreekanth Madisetty, **Kaushal Kumar Maurya**, Akiko Aizawa, and Maunendra Sankar Desarkar. "A neural approach for detecting inline mathematical expressions from scientific documents." **Expert Systems** 38, no. 4 (2021): e12576.

2. Arkadipta De, Venkatesh E, **Kaushal Kumar Maurya**, Maunendra Sankar Desarkar, "Coarse and FineGrained Hostility Detection in Hindi Posts using Fine Tuned Multilingual Embeddings." CONSTRAINT workshop, **AAAI-w** 2021. **Shared task best paper honorable mention**.

3. Aditi Bagora, Kamal Shrestha, **Kaushal Kumar Maurya**, and Maunendra Sankar Desarkar. "Hostility Detection in Online Hindi-English Code-Mixed Conversations." In Proceedings of the 14th ACM Web Science Conference 2022, pp. 390-400. 2022. **WebSci** 2022.

4. Venkatesh, E., **Kaushal Kumar Maurya**, Deepak Kumar, and Maunendra Sankar Desarkar. "DivHSK: Diverse Headline Generation using Self-Attention based Keyword Selection." In Findings of the Association for Computational Linguistics: **ACL** 2023, pp. 1879-1891. 2023.

[3]Equal contribution with Maharaj

5. **Kaushal Kumar Maurya** and Maunendra Desarkar. “Towards Low-resource Language Generation with Limited Supervision.” In the BigPicture workshop. Empirical Methods in Natural Language Processing: **EMNLP-w** 2023, Singapore.

6. Aishwarya Maheswaran, **Kaushal Kumar Maurya**[4], Manish Gupta, and Maunendra Sankar Desarkar. ”DQAC: Detoxifying Query Auto-completion with Adapters.” In Pacific-Asia Conference on Knowledge Discovery and Data Mining (**PAKDD**), pp. 108-120. Singapore: Springer Nature Singapore, 2024.

7. Aishwarya Maheswaran, **Kaushal Kumar Maurya**, Manish Gupta, and Maunendra Sankar Desarkar. ”DAC: Quantized Optimal Transport Reward-based Reinforcement Learning Approach to Detoxify Query Auto-Completion.” In Proceedings of the 47th International **ACM SIGIR** Conference on Research and Development in Information Retrieval, pp. 608-618. 2024.

# References


[1] I. Goodfellow, Y. Bengio, and A. Courville. Deep Learning. MIT Press, 2016. http://www.deeplearningbook.org. 1

[2] W. X. Zhao, K. Zhou, J. Li, T. Tang, X. Wang, Y. Hou, Y. Min, B. Zhang, J. Zhang, Z. Dong et al. A survey of large language models. *arXiv preprint arXiv:2303.18223* . 1

[3] J. Yang, H. Jin, R. Tang, X. Han, Q. Feng, H. Jiang, B. Yin, and X. Hu. Harnessing the Power of LLMs in Practice: A Survey on ChatGPT and Beyond . 1

[4] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin. Attention is all you need. *Advances in neural information processing systems* 30. 1

[5] P. Joshi, S. Santy, A. Budhiraja, K. Bali, and M. Choudhury. The State and Fate of Linguistic Diversity and Inclusion in the NLP World. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics, Online, 2020 6282–6293. 1

[6] E. Bender. The# benderrule: On naming the languages we study and why it matters. *The Gradient* 14. 1

[7] D. van Esch, T. Lucassen, S. Ruder, I. Caswell, and C. Rivera. Writing system and speaker metadata for 2,800+ language varieties. In Proceedings of the Thirteenth Language Resources and Evaluation Conference. 2022 5035–5046. 1

[8] E. M. Bender. On achieving and evaluating language-independence in NLP. *Linguistic Issues in Language Technology* 6. 1


[4]Equal contribution with Aishwarya

[9] S. Ruder, I. Vulić, and A. Søgaard. Square One Bias in NLP: Towards a Multi-Dimensional Exploration of the Research Manifold. In Findings of the Association for Computational Linguistics: ACL 2022. Association for Computational Linguistics, Dublin, Ireland, 2022 2340–2354. 1

[10] S. Doddapaneni, G. Ramesh, A. Kunchukuttan, P. Kumar, and M. M. Khapra. A Primer on Pretrained Multilingual Language Models. *CoRR* abs/2107.00676. 1

[11] L. Xue, N. Constant, A. Roberts, M. Kale, R. Al-Rfou, A. Siddhant, A. Barua, and C. Raffel. mT5: A Massively Multilingual Pre-trained Text-to-Text Transformer. In Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Association for Computational Linguistics, Online, 2021 483–498. 2, 5

[12] K. Ahuja, R. Hada, M. Ochieng, P. Jain, H. Diddee, S. Maina, T. Ganu, S. Segal, M. Axmed, K. Bali et al. Mega: Multilingual evaluation of generative ai. *arXiv preprint arXiv:2303.12528* . 2

[13] B. Haddow, R. Bawden, A. V. M. Barone, J. Helcl, and A. Birch. Survey of low-resource machine translation. *Computational Linguistics* 48, (2022) 673–732. 2

[14] K. K. Maurya and M. S. Desarkar. Learning to Distract: A Hierarchical Multi-Decoder Network for Automated Generation of Long Distractors for Multiple-Choice Questions for Reading Comprehension. In Proceedings of the 29th ACM International Conference on Information & Knowledge Management. 2020 1115–1124. 2, 3

[15] K. K. Maurya, M. S. Desarkar, M. Gupta, and P. Agrawal. trie-nlg: trie context augmentation to improve personalized query auto-completion for short and unseen prefixes. *Data Mining and Knowledge Discovery* 37, (2023) 2306–2329. 2, 4

[16] P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Küttler, M. Lewis, W.-t. Yih, T. Rocktäschel et al. Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in Neural Information Processing Systems* 33, (2020) 9459–9474. 2, 4

[17] K. K. Maurya, M. S. Desarkar, Y. Kano, and K. Deepshikha. ZmBART: An Unsupervised Cross-lingual Transfer Framework for Language Generation. In Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021. Association for Computational Linguistics, Online, 2021 2804–2818. 2, 5

[18] K. Maurya and M. Desarkar. Meta-X$_{NLG}$: A Meta-Learning Approach Based on Language Clustering for Zero-Shot Cross-Lingual Transfer and Generation. In Findings of the Association for Computational Linguistics: ACL 2022. Association for Computational Linguistics, Dublin, Ireland, 2022 269–284. 2, 6

[19] K. K. Maurya, R. Kejriwal, M. S. Desarkar, and A. Kunchukuttan. Utilizing Lexical Similarity to Enable Zero-Shot Machine Translation for Extremely Low-resource Languages. In arXiv preprint arXiv:2305.05214. 2023 . 2, 7

[20] M. Brahma, K. K. Maurya, and M. S. Desarkar. SelectNoise: Unsupervised Noise Injection to Enable Zero-Shot Machine Translation for Extremely Low-resource Languages. In Findings of EMNLP. 2023 . 2, 7

[21] V. E, K. K. Maurya, D. Kumar, and M. S. Desarkar. DIVHSK: Diverse Headline Generation using Self-Attention based Keyword Selection. In Findings of the Association for Computational Linguistics: ACL 2023. Association for Computational Linguistics, Toronto, Canada, 2023 . 3

[22] E. Wenger. Artificial intelligence and tutoring systems: Computational and cognitive approaches to the communication of knowledge. Morgan Kaufmann, 2014. 3

[23] B. Mitra and N. Craswell. Query auto-completion for rare prefixes. In Proceedings of the 24th ACM international on conference on information and knowledge management. 2015 1755–1758. 4

[24] I. D. Melamed. Automatic Evaluation and Uniform Filter Cascades for Inducing N-Best Translation Lexicons. In Third Workshop on Very Large Corpora. 1995 . 7