

# Text Generation to Auto-complete using RNNs

Mukund Dhar      Jitesh Parapoi      Chetan Choudhary  
University of Central Florida

March 28, 2024

## 1 Introduction

Auto-completion is a widely used feature in email and messaging applications, enhancing user experience by predicting and suggesting the next word or phrase based on the context of the input. Recurrent Neural Networks (RNNs) such as LSTM and GRU networks, have shown promising results in various natural language processing (NLP) tasks, including text generation. In this project, we aim to leverage RNNs to develop an auto-complete system for a given text input.

## 2 Project Goals

The primary goal of this project is to create an efficient and accurate text auto-completion system using RNNs. Specifically, our objectives are:

- Collect and prepare text data to process it into their appropriate representations.
- Train an RNN model on a large corpus of text data to learn the relationships between words and their contexts.
- Implement a mechanism for generating probable next words or phrases given a partial input text.

## 3 Proposed Methods

We plan to achieve the project goals through the following steps:

1. Data Collection: Gather a diverse dataset of text from various sources such as books, articles, and online content.
2. Preprocessing: Clean and tokenize the text data, including removing special characters, converting text to lowercase, and splitting it into individual sequences of the desired size.

3. Model Architecture: Design and implement an RNN architecture suitable for sequence prediction tasks. We will experiment with different types of RNNs, such as Vanilla RNNs, Long Short-Term Memory (LSTM) networks, and Gated Recurrent Units (GRUs).
4. Training: Train the RNN model on the preprocessed text data using appropriate optimization algorithms and hyperparameters. We will evaluate the model's performance using metrics such as perplexity and accuracy.

## 4 Related Works

Several studies have explored text generation and auto-completion providing valuable insights and methodologies that we will consider and possibly extend upon in our project. Some notable works include:

1. K. Padmanandam, J. Nikhitha, P. P. Sri, G. Pavithra and C. S. Megaha, "Machine Learning Powered Text Auto-Completion and Generation", *2023 7th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, Coimbatore, India, 2023, pp. 511-516, doi: 10.1109/ICECA58529.2023.10394873. [link](#)
2. Sunitha Devi, P., Tejaswini, C.S., Keerthana, M., Cheruvu, M., Srinivas, M. (2023). Prediction of Next Words Using Sequence Generators and Deep Learning Techniques. In: *Seetha, M., Peddoju, S.K., Pendyala, V., Chakravarthy, V.V.S.S.S. (eds) Intelligent Computing and Communication. ICICC 2022. Advances in Intelligent Systems and Computing*, vol 1447. Springer, Singapore. [link](#)
3. Kulkarni, A., Shivananda, A., Kulkarni, A. (2022). Text Generation: Next Word Prediction. In: *Natural Language Processing Projects. Apress, Berkeley, CA*. [link](#)
4. Sourabh Ambulgekar, Sanket Malewadikar, Raju Garande and Bharti Joshi "Next Words Prediction Using Recurrent NeuralNetworks" *ITM Web Conf., 40 (2021) 03034* [link](#)
5. Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., Stoyanov, V., & Zettlemoyer, L. (2019). BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension. In *arXiv preprint arXiv:1910.13461*. [link](#)

## 5 Conclusion

In conclusion, our project aims to develop an effective text auto-completion system using RNNs and other related techniques. By leveraging the capabilities of recurrent neural networks and drawing from related works in the field, we seek to enhance user experience in text editing and messaging applications.