Informatics Institute of Technology

In Collaboration With

University of Westminster, UK



*University of Westminster, Coat of Arms*

Generalized Abstractive Text Summarization Using Optimized Transformers

Literature Review

Mr. Nazhim Kalam

w1761265 / 2019281

Supervised by

Mr. Torin Wirasingha

September 2022

This Project Proposal is submitted in partial fulfilment of the requirements for

the BSc (Hons) Computer Science degree at

the University of Westminster.

Table of Contents

[List of Figures ii](#_Toc118908179)

[List of Tables ii](#_Toc118908180)

[1. Chapter Overview 1](#_Toc118908181)

[2. Concept Map 1](#_Toc118908182)

[3. Problem Domain 1](#_Toc118908183)

[3.1 User Reviews 1](#_Toc118908184)

[3.2 Corporate Advantage 2](#_Toc118908185)

[3.3 Text Summarization 2](#_Toc118908186)

[3.4 Abstractive and Extractive Techniques 3](#_Toc118908187)

[3.5 NLP with Deep Learning 4](#_Toc118908188)

[3.6 Transformers 4](#_Toc118908189)

[3.7 Hyperparameter Tuning 6](#_Toc118908190)

[3.8 Generalization 6](#_Toc118908191)

[3.9 Proposed architecture for the generalized text summarization system 6](#_Toc118908192)

[4. Existing Work 7](#_Toc118908193)

[5. Technological Review 7](#_Toc118908194)

[6. Evaluation Approaches 7](#_Toc118908195)

[7. Chapter Summary 7](#_Toc118908196)

[References I](#_Toc118908197)

[Appendix A – Concept Map Iii](#_Toc118908198)

# List of Figures

[Figure 12.1 - Prototype Feature Diagram (Self-composed) 13](#_Toc117550682)

[Figure 13.1 - Gantt Chart 16](#_Toc117550683)

[Figure 13.2 - Model development flow (Self-composed) 21](#_Toc117550684)

# List of Tables

[Table 5.1 - Related work in abstractive text summarization 3](#_Toc117584436)

[Table 11.1 - Research Objectives 9](#_Toc117584437)

[Table 13.1 - Research Methodology 13](#_Toc117584438)

[Table 13.2 - Deliverables and Dates 17](#_Toc117584439)

[Table 13.3 - Risk Mitigation Plan 19](#_Toc117584440)

**Acronyms**

|  |  |
| --- | --- |
| AI | Artificial Intelligence. |
| DL | Deep Learning |
| GUI | Graphical user Interface |
| ML | Machine Learning |
| NLP | Natural Language Processing |
| ROUGE | Recall-Oriented Understudy for Gisting Evaluation. |
| BLEU | Continuous-time Recurrent Neural Network. |
| T5 | Deep Learning. |
| BART | Graphics Processing Unit. |
| BERT | Long Short-Term Memory. |
| PEGASUS | Liquid Time-constant. |
| ILP | Machine Learning. |
| LSTM | Symmetric Mean Absolute Product Error. |
| RNN | Mean Absolute Scaled Error. |
| CNN  SEQ2SEQ | Mean Squared Error.  Sequence to Sequence |
| RoBERTa | Robustly Optimized BERT Pre-training Approach |
| GPT-3  REST | Third Generation Generative Pre-Trained Transformer  Representational State Transfer |

# CHAPTER OVERVIEW

In this chapter, the author presents critiques on prior relevant work about the use of abstractive text summarization in the domain of movie review summarization, along with the usage of advanced deep learning approaches such as transformers. Additionally, the author tries to create a generalized model that will handle several other domains in addition, not just to only the movie domain. Finally, the author determines the optimal transformer design that has been improved in order to produce the greatest outcomes by obtaining the optimum set of hyperparameters by model fine-tuning.

# CONCEPT MAP

The concept map illustrates the project scope that will be addressed in this literature review, and the nodes that are highlighted correspond to the project's primary study areas. The concept map was created to ensure that all necessary literature was covered. The concept map can be found in [**Appendix A – Concept Map**](#ConceptMap)**.**

# PROBLEM DOMAIN

The simplicity of selling products or services to customers is growing along with the usage of technology and the internet. Sellers utilize customer feedback to better decide how to improve sales and so attain customer satisfaction (Boorugu, Ramesh and Madhavi, 2019). When it comes to movies, people typically find it quite challenging to quickly determine whether a movie meets their demands by reading the reviews, which may occasionally be very lengthy and time-consuming (Khan et al., 2020).

## **User Reviews**

A user/customer review is typically referred to be written feedback from a customer who has used a product or service. Consumers frequently use user ratings and reviews to drive their purchasing decisions. Because the review data is unstructured, it becomes more challenging for consumers to compare and understand lengthier reviews (Lackermair, Kailer and Kanmaz, 2013).

User and customer reviews are extremely important to major corporations like tourism and hospitality as they constitute the primary engine for the country's economic growth and development. where tourists from over the world may blog about their experiences and share their reviews online in numerous formats (Mukherjee et al., 2020).

## **Corporate Advantage**

It is also known that it costs at least five times as much time and money to acquire a new customer as it does to keep an existing one, so it is important to learn how to foster customer loyalty to the brand, business, or service that is being offered. Customer satisfaction is essential to the survival of corporate industries. Understanding client expectations through their feedback or reviews helps business industries grow and fix faults (Pizam and Ellis, 1999).

On the other hand, companies like Netflix or Amazon Prime can use the movie summaries to help users and understand the watching pattern or their interest. Likewise, the movie-related industries need to allow the customers to quickly scan the summary and quickly decide whether they should be watching it or not (Khan et al., 2020).

## **Text Summarization**

With the massive accumulation of information/data on the internet nowadays, it is extremely difficult to extract relevant information from a large number of textual documents. The goal of text summarizing is to provide a condensed yet meaningful version of a lengthy textual content (Shi et al., 2020).

We all know that text summarization has several uses in a variety of internet-based fields, including search engines that are used for querying and e-commerce sites that utilize sentiment analysis to determine client satisfaction with items (Etemad, Abidi and Chhabra, 2021).

However, in the movie industry, consumers may utilize text summarization to simplify customer reviews of movies, which are often lengthy and time-consuming to read. This enables users to make better decisions when they decide whether or not to watch a certain movie (Khan et al., 2020).

## **3.4 Abstractive and Extractive Techniques**

Generally, text summarization is classified into two which are; abstractive text summarization and extractive text summarization, however the approach for creating a hybrid model for text summarization is possible (Alsaqer and Sasi, 2017). The abstractive text summarization technique aims to produce the sentences on its own and then uses them to provide a coherent summary. Therefore, the summary's content will vary from the original context yet still convey the same idea (Mahajan et al., 2021). Additionally, it is well recognized that a strong abstractive summary encompasses the input's key details and is linguistically fluent (Zhang et al., 2020).

The extractive text summarizing method focuses on picking out key phrases or groups of phrases from the original input content and combining them to produce a concise yet insightful text summary. It is determined which sentences should be included as parts of the summary based on the statistical and linguistic characteristics of the sentences (Gupta and Lehal, 2010). A hybrid system is one that combines various strategies to produce a single system. However, hybrid text summarizing systems do exist, for instance, using a combination of extractive and abstractive summarization can be utilized to generate a hybrid system that uses encoder-decoders (Kirmani et al., 2019; Abolghasemi, Dadkhah and Tohidi, 2022).

*Table 1: Comparison of Text Summarization Techniques*

|  |  |
| --- | --- |
| **Abstractive** | **Extractive** |
| Paraphrases content like humans do, meaning it creates its own context (Mahajan et al., 2021) | Doesn’t create its own context but uses the best possible phrases from the original document (Gupta and Lehal, 2010) |
| A vast number of datasets are available to experiment working in this domain. | Capable of visualizing sentence scores and investigating gradient-based ways to calculating the contribution of each input token to score prediction (Pai, 2014) |
| There is a probability of creating information which may be faulty or that gives a different in meaning compared to the original text. | There is a possibility that the combined sentences made from the extracted sentences will contain errors. |

## **3.5 NLP with Deep Learning**

NLP is a method for computers to intelligently and effectively analyze, comprehend, and derive meaning from human language, as opposed to other approaches that only focus on the interactions between human language and computers. Deep learning techniques are increasingly being used in the field of AI compared to traditional machine learning approaches due to their success rates in handling difficult high computing learning tasks (Lopez and Kalita, 2017; Mahajan et al., 2021).

In today's NLP, machine learning is prominent, but for the most part it only involves numerically optimizing the weights of characteristics and representations that have been created by humans. Deep learning aims to investigate how computers can utilize data to create features and representations suitable for challenging interpretation tasks (Socher, Bengio and Manning, 2012).

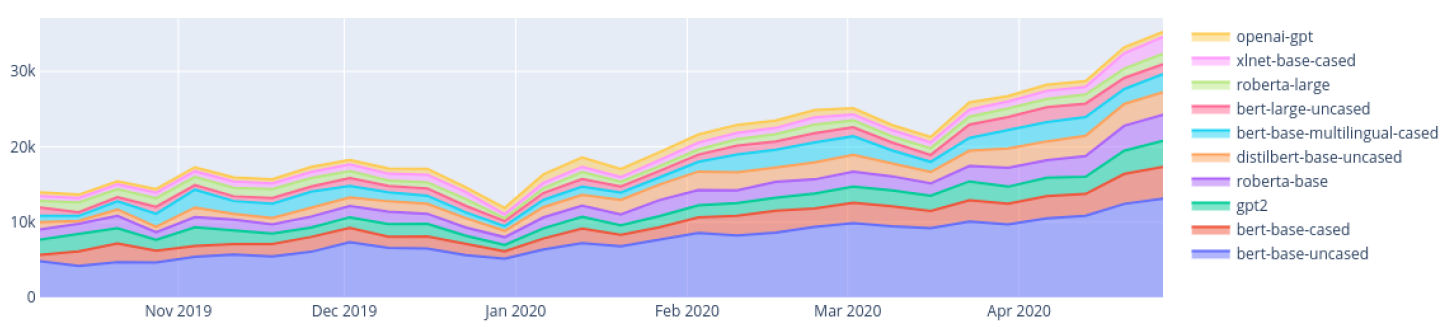
## **3.6 Transformers**

Open-source library Transformers contains modern transformer architectures that have been thoroughly developed and are integrated by a common API. Pretraining has enabled the efficient use of this capacity for a wide range of activities, and these designs have permitted the construction of higher-capacity models. Transformers are designed to be easy for practitioners, expandable for researchers, and quick and reliable in industrial deployments (Wolf et al., 2020).

It has been demonstrated that the modern generation of pre-trained language models based on transformers is rather competent at identifying syntactic signals like noun modifiers, possessive pronouns, prepositions, or co-referents, as well as semantic cues like entities and relations (Brasoveanu and Andonie, 2020).

Hugging Face Hub offers a variety of transformer designs, including BERT, GPT2, T5, PEGASUS, and many others. The figure below represents the daily average for unique downloads of the pretrained transformer model architectures between Oct 2019 to May 2020 (Wolf et al., 2020).

*Figure 3.1 – Transformer Architecture Downloads Rate (Wolf et al., 2020).*



(Etemad, Abidi and Chhabra, 2021) research compares various other researchers approaches taken in order to perform abstractive text summarization, these techniques includes the use of transformers and other neural network approaches such as CNN and LSTM RNN networks. The research comparison table below only includes the approaches of transformers used taken abstractive text summarization.

*Table 3.1 – Comparison table for abstractive text summarization using transformers (Etemad, Abidi and Chhabra, 2021).*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Researcher | Year | Type of model | Rouge 1 | Rouge 2 | Rouge L | Dataset |
| Haoyu Zhang et al. | 2019 | Transformer with BERT | 41.71 | 19.49 | 38.79 | CNN-Daily Mail |
| Andrew Hoang et al | 2019 | Transformer | 39.01  36.73  40.87 | 17.87  14.93  28.59 | 36.17  29.66  37.62 | CNN Daily Mail  Xsum  Newsroom |
| Kaiqiang Song et al. | 2019 | Transformer | 40.89  45.93 | 19.11  24.14 | 37.60  42.51 | Gigaword,  Newsroom |
| Mike Lewis et al. | 2019 | BART | 44.16  45.14 | 21.28  22.27 | 40.90  37.25 | CNN Daily Mail  Xsum |
| Itsumi Saito et al. | 2020 | RoBERTa Base | 45.80  45.42 | 22.53  22.13 | 42.48  36.92 | CNN Daily Mail  Xsum |
| Beliz Gunel et al. | 2020 | Transformer XL | 34.273 | 13.018 | 32.048 | CNN Daily Mail |
| Colin Raffel et al. | - | T5 | 43*.*52 | 21*.*55 | 40*.*69 | CNN Daily Mail |

## **3.7 Hyperparameter Tuning**

Finding the ideal collection of parameter values to train an algorithm using in order to build a model relevant to the dataset is known as hyperparameter tuning (Liu and Wang, 2021). The calculation of the performance improvement that may be obtained by changing the value of each of the considered hyperparameters from the original value to the value indicated in the target configuration set by the tuning strategy is where hyperparameters make the biggest contribution to improving algorithm performance (Joy and Selvan, 2022).

There are several hyperparameters that play a significant role in performance enhancement; however, not all of the parameters do so; just a select handful do, for example, learning rate, weight decay, number of epochs, batch size, and warmup ratio. As a result, giving critical hyperparameters a higher priority is crucial (aws.amazon.com, 2022).

Automated framework tools, such as Optuna, an open-source framework for hyperparameter optimization built on the Python programming language, does hyperparameter tweaking. The application of numerous hyperparameter optimization techniques, including Grid Search, Random Search, TPE, and CMA-ES algorithms, was made easier by this framework (Joy and Selvan, 2022).

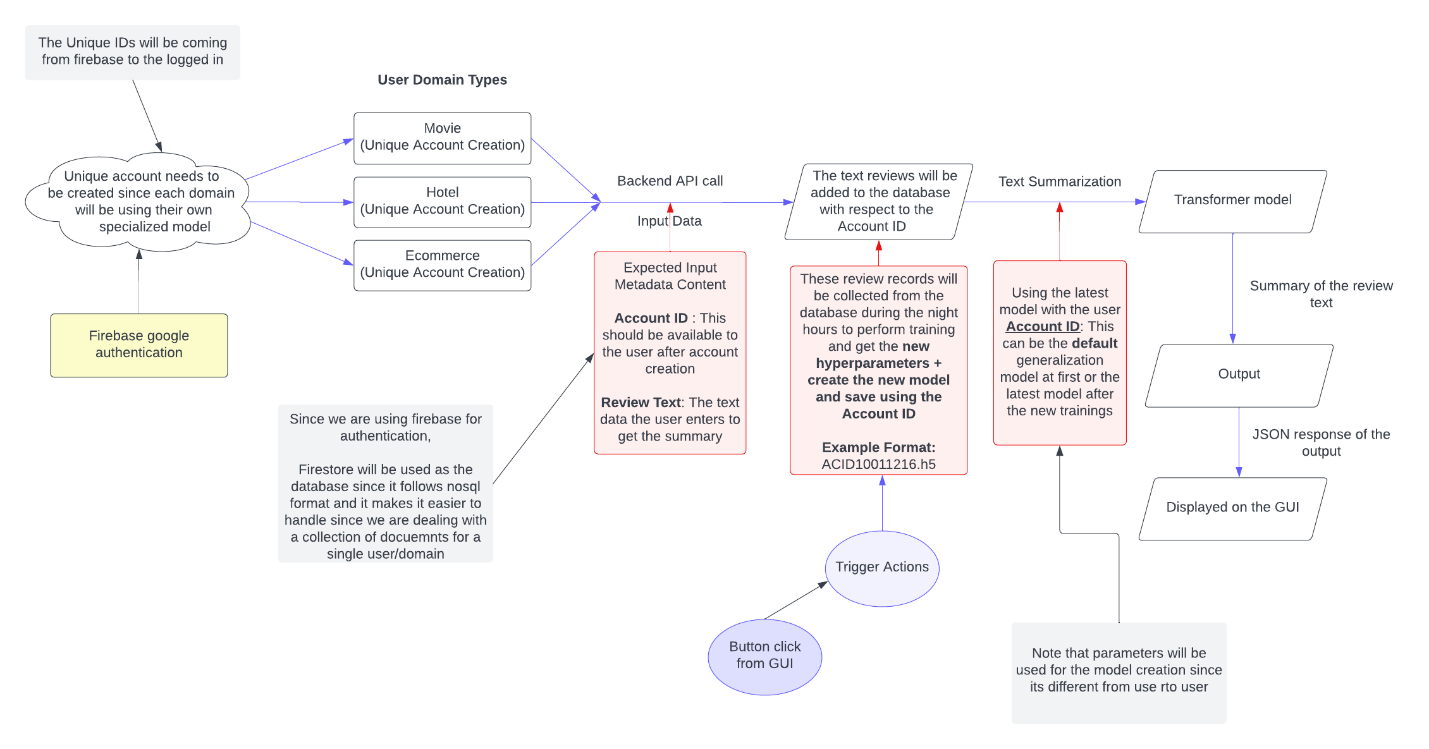
## **3.8 Generalization**

Generalization now plays a significant part in resolving issues in numerous fields that are linked to the same issue. The capacity of a model to generalize to new, previously unobserved data that comes from the same distribution as the model's original data is known as generalization (Neyshabur et al., 2017).

Generalization is also an effective method of solving a problem from the base constant at start and improves or specializes to one’s domain as new data is exposed.

## **3.9 Proposed architecture for the generalized text summarization system**

*Figure 3.2 – Proposed Generalized Abstractive Summarization System Process Flow*



# EXISTING WORK

There have been several works done on abstractive text summarization for the field of movie reviews, primarily using classic machine learning algorithms. However, there are several limitations that call for the inclusion of the most recent deep learning approaches to improve the **system's** **performance**.

(Khan et al., 2020) The author has done research on an automated method to condense long movie evaluations and enable viewers to swiftly distinguish a movie's good and bad points, by first focusing on feature extraction, transforming reviews into vector spaces, and then applying the Naive Bayes machine learning method for review classification utilizing an undirected weighted graph-based ranking algorithm to rank score for each review phrase in graph and then, in order to construct an **extractive** **summary**, the highest scoring sentences are selected. However, the author has limited the use of sophisticated deep learning algorithms to improve performance by solely using standard machine learning approaches to tackle the problem.

(Boorugu, Ramesh and Madhavi, 2019) In order to save the consumer time and provide a thorough summary of the reviews for him/her to decide if the product is what they are searching for, the author has concentrated on using customer reviews on items when making purchase decisions. Using the **seq2seq** model for summarization, the attention mechanism for improved accuracy, and the Concept net Number batch word embedding model, which is superior than Glove. Utilizing a 1D convolutional layer, a max pooling layer, an LSTM layer, and finally a fully connected layer at the very end. However, the author's use of generic deep learning algorithms to handle this problem introduces a new constraint that prevents performance from being improved using the most recent deep learning strategy for NLP-related problems, transformers.

(Mukherjee et al., 2020) By using an **extractive** **method** based on integer linear programming (**ILP** [Unsupervised method]) to choose an informative subset of opinions centered on the identified aspects, the author has investigated a solution for producing personalized aspect-based opinion summaries from large collections of online traveler reviews. The summary's attributes can also be customized based on the user's interest. Utilize ROUGE-based criteria to assess and contrast the summaries and get competitive outcomes. Since the dataset is also constrained, extractive summaries could not be particularly insightful; thus, utilizing an abstractive technique might produce superior results, despite the dataset's constrained size.

(Gupta et al., 2021) By employing pretrained models such Pipeline BART, BART modified, T5, and PEGASUS to deal with text summarization, the author has done extensive study on a comparison of a few transformer architecture-based pre-trained models. The ROUGE Scores were used as the evaluation measures. During the experiments, the author employed transformer designs; however, the **hyperparameters** used were **default** and might be tuned for a better performance. The constraints consist of concentrating on developing more reliable models that can further expand the method to produce summaries of varying length and applicable for multi-document summarization.

(Mahajan et al., 2021) The focus of the authors' study is utilizing the **encoder-decoder** model with the attention layer to produce text summaries with good syntax and no repeated words. the creation of an encoder-decoder model with gated recurrent units and training it to provide an abstract summary of a piece of writing. Although the author employed deep learning, its application in production required real-time training so that it could be updated with the most recent content over time.

(Etemad, Abidi and Chhabra, 2021)The author explores with deep learning methods in the broad text summarization domain to determine which method—among a collection that includes RNN, CNN, and Transformers—performs best. The author also considers metrics for model evaluations including BLEU and ROUGE, despite using sophisticated deep learning algorithms, the author was unable to undertake **hyperparameter** tuning to improve the method and obtain a better outcome.

# TECHNOLOGICAL REVIEW

# EVALUATION APPROACHES

# CHAPTER SUMMARY

# REFERENCES

Abolghasemi, M., Dadkhah, C. and Tohidi, N. (2022). HTS-DL: Hybrid Text Summarization System using Deep Learning. *2022 27th International Computer Conference, Computer Society of Iran (CSICC)*. 23 February 2022. Tehran, Iran, Islamic Republic of: IEEE, 1–5. Available from https://doi.org/10.1109/CSICC55295.2022.9780395 [Accessed 26 October 2022].

Alsaqer, A.F. and Sasi, S. (2017). Movie review summarization and sentiment analysis using rapidminer. *2017 International Conference on Networks & Advances in Computational Technologies (NetACT)*. July 2017. Thiruvanthapuram, India: IEEE, 329–335. Available from https://doi.org/10.1109/NETACT.2017.8076790 [Accessed 10 October 2022].

Boorugu, R., Ramesh, G. and Madhavi, K. (2019). Summarizing Product Reviews Using Nlp Based Text Summarization. *International Journal of Scientific & Technology Research Volume*, 8 (10), 1127–1133.

Brasoveanu, A.M.P. and Andonie, R. (2020). Visualizing Transformers for NLP: A Brief Survey. *2020 24th International Conference Information Visualisation (IV)*. September 2020. Melbourne, Australia: IEEE, 270–279. Available from https://doi.org/10.1109/IV51561.2020.00051 [Accessed 2 November 2022].

Etemad, A.G., Abidi, A.I. and Chhabra, M. (2021). A Review on Abstractive Text Summarization Using Deep Learning. *2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*. 3 September 2021. Noida, India: IEEE, 1–6. Available from https://doi.org/10.1109/ICRITO51393.2021.9596500 [Accessed 10 October 2022].

Gupta, A. et al. (2021). Automated News Summarization Using Transformers. *ArXiv*, abs/2108.01064.

Gupta, V. and Lehal, G.S. (2010). A Survey of Text Summarization Extractive Techniques. *Journal of Emerging Technologies in Web Intelligence*, 2 (3), 258–268. Available from https://doi.org/10.4304/jetwi.2.3.258-268.

Joy, J. and Selvan, M.P. (2022). A comprehensive study on the performance of different Multi-class Classification Algorithms and Hyperparameter Tuning Techniques using Optuna. *2022 International Conference on Computing, Communication, Security and Intelligent Systems (IC3SIS)*. 23 June 2022. Kochi, India: IEEE, 1–5. Available from https://doi.org/10.1109/IC3SIS54991.2022.9885695 [Accessed 24 October 2022].

Khan, A. et al. (2020). Movie Review Summarization Using Supervised Learning and Graph-Based Ranking Algorithm. *Computational Intelligence and Neuroscience*, 2020, 7526580. Available from https://doi.org/10.1155/2020/7526580.

Kirmani, M. et al. (2019). Hybrid Text Summarization: A Survey. In: Ray, K. Sharma, T.K. Rawat, S. et al. (eds.). *Soft Computing: Theories and Applications*. Advances in Intelligent Systems and Computing. Singapore: Springer Singapore, 63–73. Available from https://doi.org/10.1007/978-981-13-0589-4\_7 [Accessed 1 November 2022].

Lackermair, G., Kailer, D. and Kanmaz, K. (2013). Importance of Online Product Reviews from a Consumer’s Perspective. *Advances in Economics and Business*, 1 (1), 1–5. Available from https://doi.org/10.13189/aeb.2013.010101.

Liu, X. and Wang, C. (2021). An Empirical Study on Hyperparameter Optimization for Fine-Tuning Pre-trained Language Models. Available from http://arxiv.org/abs/2106.09204 [Accessed 24 October 2022].

Mahajan, R. et al. (2021). Text Summarization Using Deep Learning. *International Research Journal of Engineering and Technology (IRJET)*, 08 (05th May 2021), 1737–1740.

Mukherjee, R. et al. (2020). Read what you need: Controllable Aspect-based Opinion Summarization of Tourist Reviews. *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 25 July 2020. 1825–1828. Available from https://doi.org/10.1145/3397271.3401269 [Accessed 10 October 2022].

Neyshabur, B. et al. (2017). Exploring Generalization in Deep Learning. *undefined*. Available from https://www.semanticscholar.org/reader/d53fb3feeeab07a0d70bf466dd473ec6052ecc07 [Accessed 9 November 2022].

Pai, A. (2014). Summarizer Using Abstractive and Extractive Method. *International Journal of Engineering Research*, 3 (5), 5.

Pizam, A. and Ellis, T. (1999). Customer satisfaction and its measurement in hospitality enterprises. *International Journal of Contemporary Hospitality Management*, 11 (7), 326–339. Available from https://doi.org/10.1108/09596119910293231.

Shi, T. et al. (2020). Neural Abstractive Text Summarization with Sequence-to-Sequence Models. Available from http://arxiv.org/abs/1812.02303 [Accessed 10 October 2022].

Socher, R., Bengio, Y. and Manning, C.D. (2012). Deep Learning for NLP (without Magic). *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Tutorial Abstracts*. July 2012. Jeju Island, Korea: Association for Computational Linguistics, 5. Available from https://aclanthology.org/P12-4005 [Accessed 2 November 2022].

Wolf, T. et al. (2020). Transformers: State-of-the-Art Natural Language Processing. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*. 2020. Online: Association for Computational Linguistics, 38–45. Available from https://doi.org/10.18653/v1/2020.emnlp-demos.6 [Accessed 10 October 2022].

Zhang, J. et al. (2020). PEGASUS: Pre-training with Extracted Gap-sentences for Abstractive Summarization. Available from http://arxiv.org/abs/1912.08777 [Accessed 18 October 2022].

---External ref---

aws.amazon.com. (2022). *Hyperparameter optimization for fine-tuning pre-trained transformer models from Hugging Face | AWS Machine Learning Blog*. [online] Available at: https://aws.amazon.com/blogs/machine-learning/hyperparameter-optimization-for-fine-tuning-pre-trained-transformer-models-from-hugging-face/ [Accessed 9 Nov. 2022].

‌

# APPENDIX A – CONCEPT MAP

