A Review of Classification of Insincere Questions in Quora Using Deep Learning Approaches

Nishad Ahamed
Department of Information Technology
University of Moratuwa
Katubedda, Sri Lanka
ahamedmnn.19@uom.lk

Supunmali Ahangama
Department of Information Technology
University of Moratuwa
Katubedda, Sri Lanka
supunmali@uom.lk

2023 IEEE 17th International Conference on Industrial and Information Systems (ICIIS) | 979-8-3503-2363-4/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICIIS58898.2023.10253595

Abstract— **The rise of online question-and-answer platforms like Quora and Stack Overflow has transformed knowledge acquisition, offering a vast repository of collective information. However, this has also led to the proliferation of insincere questions, which lack a neutral undertone and are not intended for learning. The posting of such queries can lead to chaos, necessitating the need for monitoring and discerning which questions should be published to maintain the platform's quality. The classification of insincere questions is crucial for maintaining the quality of these platforms. Detecting insincere questions can enhance user experience and foster a healthy and safe environment for online discussions. While Quora previously relied on manual reviews, the application of algorithms, particularly those from the field of Deep Learning, a subset of Machine Learning, has made this process more efficient and accessible. This paper reviews various approaches and algorithms used in classifying insincere questions on Quora, their benefits, and drawbacks. It provides insights into how Deep Learning can enhance the accuracy and effectiveness of insincere question classification on online platforms. The review findings indicate that class imbalance is a common issue in this type of research, highlighting the need for further exploration in future work.**

Keywords - Text Classification, Deep Learning, Convolutional Neural Network, Quora, Insincere, Sincere, Machine Learning, Question

I. Introduction

Internet use has grown over the period in leaps and bounds. This popularity stems from the Internet's capacity to simplify complex issues. Internet users frequently search for answers. This swiftly expanded into posting questions on a website or in a forum where global contributors could answer. The simplicity of question forums makes them popular. Quora, Stack Overflow, and Yahoo Answers are some of the question-and-answer (Q&A) platforms available for communities.

Online question-and-answer platforms such as Quora and Stack Overflow have become indispensable sources of information, facilitating diverse individuals to discuss and share knowledge [1]. However, this has been exploited by some individuals who insert meaningless content amid valuable discussions, leaving the platform vulnerable to attacks and frustrating participants. It is necessary to classify "insincere" questions to address this issue. Insincere questions are characterized by incorrect assumptions, promotion of violence, explicit content, or targeting a specific audience. Such questions are intended to make a point rather than elicit genuine responses, and they are often non-neutral, insulting, provocative, unfounded, and explicit.

Q&A forums are meant to share expertise. However, some users submit inappropriate and destructive content. Actual learners on these networks have little user engagement. These questions often produce pandemonium, undermining these websites' mutual aid mission. These requests must be classified and eliminated before they damage the website's reputation [2]. Thus, these questions must be categorized and eliminated before they damage the website's reputation. This page categorizes website-responsible questions.

A. *Quora questions classification*

Quora is a well-known Q&A community with over 300 million monthly active members and facilitates collaborative learning. Like any other social or Q&A forum, Quora has high and low-quality content. It also has a variety of categories in which to ask questions. This helps the user understand the category of the question. However, some individuals take unfair advantage of this website. The presence of false information on Quora puts the site and its users at risk. Manually removing mock questions from a large dataset is difficult and dangerous. Therefore, it is better to screen automatically the incoming data before posting [1].

Previously, Quora filtered incoming data using Machine Learning (ML) algorithms and manual review. As the number of users and complexity grew, developing a more scalable and reliable approach for recognizing current fraudulent inquiries and screening new ones became imperative. As a result, the Quora team generated a tagged dataset with questions and labels indicating whether or not they were genuine.

Furthermore, the classification process was enhanced by employing ML models. The ML models such as random forest, logistic regression, Support Vector Machine (SVM), and K Nearest Neighbors (KNN) are implemented with various feature extraction techniques [3]. Although an ML strategy can produce decent results, meticulous feature extraction for text data is still required. The effectiveness of a system is significantly influenced by the retrieved attributes [4], [5].

Deep Learning (DL) techniques have been suggested as a potential solution to address the issue of text representation in ML approaches and utilize word order data [4] and inspired by how the brain functions, these models require fewer engineering features to achieve high accuracy, making it easier for people to gather features for classifiers. To address this issue, researchers have employed DL architecture such as

Recurrent Neural Networks (RNN), Long Short-Term Memory networks (LSTM), and Convolutional Neural Networks (CNN) [6]. Among these DL architectures, CNN is the most commonly used for text classification due to its ability to gather local information. A simple CNN model can perform remarkably well compared to more complex DL models that use intricate pooling strategies [6]. These models have shown great promise in accurately classifying insincere questions and have become a popular choice for researchers tackling this problem.

The use of CNN for text classification has been successful in accurately classifying insincere questions. However, one of its limitations is information loss through pooling. Capsule Networks have emerged as a promising alternative for addressing this challenge and providing spatial meaning to the text. Like other DL models, Capsule Networks require significant data to be compelling, and feature engineering is often necessary to improve their performance. Although the reliance on feature engineering has been identified as a potential limitation, Capsule Networks have shown great promise in accurately classifying insincere questions, offering an exciting area for further research in this field.

Pre-trained models, known as the Transfer Learning Method, were developed to get around the DL models' limitations. With the innovation of Transfer Learning, it has been shown that language model pretraining helps make universal language representations [7]. Several DL-based architectures, including Transformers and BERT family classifiers, were applied for this procedure. This study examines how various DL techniques may classify inquiries that are not sincere.

II. Methodology

A systematic literature search followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines to identify relevant studies on insincere question classification in Quora using DL Approaches. Electronic databases such as PubMed, Google Scholar, and IEEE Xplore were used, and search terms such as "insincere questions classification," "Quora," "deep learning," and "natural language processing" were included. Out of 218 papers identified, 20 duplicates were removed during the initial filtration. The inclusion criteria were studies published in English between 2015 and 2023 that focused on insincere question classification in Quora and used DL algorithms, resulting in 92 papers. The selected studies were then reviewed and analyzed for their contributions to the current state-of-the-art insincere question classification in Quora using DL approaches and their limitations.

During the screening process, the classification of insincere questions in Quora was categorized into linguistic, contextual, and behavioral features. This categorization was taken into account during the selection of relevant studies. According to [3], [4], [6] DL models for insincere questions Classification in Quora can be categorized into RNNs, CNNs, attention-based models, and transformer-based models. This study is focused on these model-based approaches, as it enables a comparison of the different DL models used, their architecture, and performance, to examine the applicability and effectiveness of DL models in insincere question classification in Quora.

The findings from the selected studies were synthesized and discussed in the context of the research objectives. This review mainly focuses on using natural language processing (NLP) techniques for insincere question classification in Quora. This review paper aims to provide insights and guidance for researchers and practitioners working in this field by analyzing the current state-of-the-art DL approaches used for this task.

III. Insincere Question Classification

The widespread expansion of the Internet has significantly increased the volume of textual information available to users. However, with such a vast repository of information, Internet surfers often struggle to identify trustworthy sources. In response to this challenge, Kaggle, a platform that allows data scientists to apply ML ideas and compete globally, hosted an online competition to classify insincere questions on Quora [8]. This competition demonstrated the growing interest among academics in automating identifying insincere questions on online platforms such as Quora. By automating the classification of insincere questions, researchers hope to improve the quality of online discourse and protect users from harmful content.

Improving the community experience requires the development of a classifier that can automatically distinguish between authentic and misleading user-generated questions. A combination of Random Forest, Naive Bayes, and SVM algorithms was employed to achieve this. Despite the challenges posed by text data feature extraction, ML algorithms have shown promising results in this domain [9]. However, the choice of features dramatically affects the performance of these models. The results revealed that decision trees outperformed other ML models in this task. Additionally, the ability of ML models to predict long document sequences was low. While these models performed well on the training dataset, their accuracy decreased on new and evolving datasets due to the limitations imposed by predefined features [9].

To achieve accurate text representation and word order data, DL is necessary, as noted in [4]. While CNN pooling is a popular technique for text categorization that reduces computational complexity, it has limitations. Specifically, pooling can result in misclassification by orientation or percentage, as highlighted in [9] and [10]. Capsule Networks (CapsNets) have emerged as a solution to these pooling difficulties. Capsules consist of activity vectors that express various properties of an object or object component, as explained in [9]. By incorporating Capsule Networks into CNN, text categorization performance can be improved. Dynamic routing is a crucial aspect of Capsule Networks that considers the spatial connections between capsules, as stated in [11]. In particular, dynamic routing assesses the connection between lower- and upper-level capsules using coupling coefficient-based recurrent routing, as noted in [10]. This method has been shown to improve capsule network routing and benchmarking.

Researchers used pre-trained models to improve performance and accuracy with less processing power when problem-solving methods changed. BERT's multi-layer, bidirectional Transformer encoder employs Transformer [7]. Many BERT designs improved performance. The original model is slower than RoBERTa, DistilBERT, and ALBERT

[7]. The model proposed by Vaswani et al. [12] outperformed the CNN and RNN models in NLP and synthesis—training data scale design. Realistic parallel training captures long-range sequence features with model size [12]. Users can access large-scale pre-trained models, develop and experiment on them, and deploy them in high-performance downstream applications with the open-source Transformers community and library.

This paper explores the different text classification methods utilized based on the nature of the dataset, accuracy, and performance. This paper examines the various ML models utilized for this task and explores why DL models have become increasingly necessary. The focus of this study will be on the DL approaches that are being used, as well their advantages and disadvantages.The Quora team utilized manual review and artificial intelligence (AI) to eliminate fraudulent questions from the platform. To enhance the identification of harmful or misleading content in a more scalable and dependable manner, Quora furnished a dataset comprising over 1.3 million records. Furthermore, the public can access this dataset for research purposes through Kaggle competitions. The dataset includes labels such as "Qid," "Question Text," and "Target," while Quora also provides 375,806 test data identified by "qid" and "question text." Our experiment employed training data [6], which was analyzed using pandas and graphing. The dataset contained 80,810 dishonest questions and 1,225,312 truthful ones. As indicated in Fig.1, 93.81% of the questions were sincere, while 6.18% were not, indicating a class imbalance issue within the dataset [6].

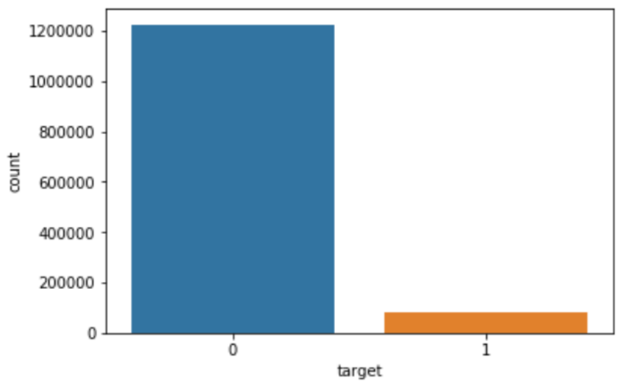

Fig. 1. Number of sincere and insincere questions. Here 1 represents insincere, and 0 illustrates sincere questions *[3].*

IV. Deep Learning Approaches

A. *Why Deep Learning Algorithms?*

Artificial neural networks (ANN) served as the basis for DL techniques, which are currently a key component of ML and have been effectively used to achieve exceptional performance in several study fields [13]. However, this section evaluates four distinct DL models for classification-related challenges.

Although an ML strategy can produce decent results, meticulous feature extraction for text data is still required. While classic ML techniques like naive Bayes, logistic regression, and decision trees can legitimately solve this issue, they have significant flaws in how their constructions work [7]. The effectiveness of a system is significantly influenced by the retrieved features [4]. The two most prevalent approaches to text representation are bag-of-words (also known as BOW) and n-grams representation. Features may consist of the frequency of terms (or words), variants such as term frequency-inverse document frequency (TF-IDF), information gain or entropy, mutual information x2 statistic, and generalized singular value decomposition of values [14]. These methods transform a text document into a high-dimensional feature vector. These techniques keep track of the frequency with which terms are used in a document but disregard information regarding word order.

However, ignoring the dependence on feature extraction and pre-processing data methods is impossible. Only with correctly extracted features do ML models function well. Additionally, they fail to vectorize sequences while maintaining their semantic models' significance. Domain-specific expertise and several trials are needed to remove features for ML models, as Wankmuller et al. noted [14].

DL approaches are suggested to address text representation and use word order information [4]. These techniques take their cues from how the human brain functions. High precision can be attained with fewer developed characteristics. As a result, it makes it easier for humans to extract features for classifiers. DL network-based models have gained popularity recently [15].

B. *Deep Learning Algorithms*

Neural networks, a class of algorithms and methods used in DL, are models of the human brain's functions. DL architectures perform incredibly high accuracy with lower-level engineering and processing, which has significant advantages for text classification.

CNN, which has the potential to capture local information, is the most often used DL architecture in text classification [16]. CNN was initially designed to address computer vision problems. It is the preferred method for extracting features from text, audio, or video data. The idea of utilizing sliding windows to perform convolution on text data was first proposed. In [16], it was further enhanced by implementing filters with different window widths. As mentioned in [4], performing phrase semantic modeling using Dynamic K-Max Pooling. To take advantage of word interactions, Lie at el. Uses n-gram characteristics [17]. Furthermore, it has then used a taxonomy knowledge base to enhance the performance of short text.

RNN is another significant DL model in addition to CNN. A multitask learning approach based on RNNs addresses the problem of limited training data by learning collaboratively across several related tasks [18]. An LSTM recurrent neural network and CNN phrase representation are used for text categorization. To enhance the semantic model, use LSTM with a tree structure. The representation of text data is a critical issue in DL [4]. Currently, word embedding is the best option.

(a) Recurrent Neural Networks (RNN)

For insincere question classification on Quora, RNNs have been widely used to capture word dependencies and structures by viewing the text as a sequence of words. However, feed-forward neural networks can outperform pure RNN models [9]. To overcome the limitations of standard RNNs in capturing long-term dependencies, LSTM has become the most popular RNN variant design [9]. Using a memory cell and three gates (input gate, output gate, and forget gate), LSTM can store values for a more extended period, and control data flow into and out of the cell, thus, solving the vanishing gradient problem in standard RNNs [9].

To improve the classification of insincere questions on Quora, recent studies have suggested incorporating natural language tree structures, long-span word relationships, and document subjects into Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) models, which have shown promising results [3]. However, despite their success, RNNs are still limited in their ability to recall long-distance interdependence between words. Latent topic models, which focus on describing only the semantic structure of a document, also fail to capture the word order [4]. Therefore, more advanced DL techniques are needed to effectively capture the complex relationships between words and the overall structure of the text, thereby improving the classification of insincere questions on Quora. It should be noted that the accuracy of RNNs for insincere question classification on Quora has been reported as 85.5% [19].

(b) Convolution neural network (CNN)

In the realm of insincere question classification on Quora, Convolutional Neural Networks (CNNs) have been frequently employed due to their ability to capture local information in DL architectures for text categorization [5]. While originally developed for computer vision challenges, CNNs have since been utilized for analyzing text, audio, and video data. By detecting spatial patterns and comprehending local and position-invariant patterns in text, CNNs have proven to be effective for categorizing short text inputs such as questions [4]. In CNN architectures for text classification, convolution, max pooling, and fully connected layers are employed to classify text requests [6]. However, traditional CNNs have difficulty interpreting complex local text data, and the sliding window semantics only analyze continuous language [8]. Furthermore, the pooling algorithms employed in CNNs discard spatial relationship information, leading to the potential misclassification of objects based on their orientation or proportion [4]. Despite these limitations, CNNs have exhibited superior performance over other models in text classification tasks [6]. Notably, a reported accuracy of 90.49% for CNNs was achieved in insincere question classification on Quora [20].

(c) Gated Recurrent Unit (GRU)

In recent years, the use of DL models in natural language processing (NLP) has significantly improved the accuracy of text classification tasks, including insincere question classification on Quora. Among these models, the Gated Recurrent Unit (GRU) has emerged as a superior variant of conventional recurrent neural networks (RNNs) [21]. Unlike RNNs, GRUs can effectively solve the vanishing gradient problem and retain knowledge from the past while discarding irrelevant data, leading to better performance than LSTM and RNN models. This makes GRUs an attractive choice for insincere question classification on Quora. CuDNNGRU, a fast implementation of GRU backed by CuDNN (a GPU-accelerated library for deep neural networks), has been utilized in several studies [17]. GRU was first introduced by Cho et al. (2014) as a simplified and improved version of LSTM, using update and reset gates to replace separate memory units [18]. With its streamlined architecture, GRU can effectively capture the temporal mapping relationship between sequential data [22].

Applying GRU in classifying insincere questions on Quora can benefit the task by allowing the model to capture better the temporal relationship between words and sentences in the question, leading to improved accuracy in detecting insincere content. Another DL model that has been extensively used for text classification tasks, including insincere question classification on Quora, is the Convolutional Neural Network (CNN) [5]. CNNs were originally developed for computer vision tasks but have since been applied to text, audio, and video analysis. CNNs can detect spatial patterns and understand local and position-invariant patterns in text, making them practical for categorizing short text inputs like questions [4]. Despite some limitations in interpreting complex local text data, CNNs have been shown to outperform other models in text classification tasks, with an accuracy of 89.49% reported in one study [23].

(d) Capsule Neural Networks

Capsule networks have recently been used in insincere question classification on Quora, where they have shown promising results. Unlike traditional CNNs, capsule networks use capsules, groups of neurons that encode visual object properties and represent them as vectors using a non-linear squashing function. Capsules can also retain spatial relationship information, often lost in the pooling algorithms of CNNs, allowing them to capture better objects' attributes in images and text [4].

In text classification, capsule networks have been used to represent sentences and documents as vectors, with promising results. For instance, one study proposed a CapsNets-based text classification model that employs dynamic routing, allowing capsules to find lower-level elements by examining their spatial interactions. The model outperformed traditional CNNs in text classification with an accuracy of 87.17% [9]. These findings suggest that capsule networks could be a valuable tool for insincere question classification on Quora, as they can capture the temporal relationships between sequential data and retain spatial information, which could improve the accuracy of classification models.

V. Discussion

The implications of this paper are significant for the field of NLP and Q&A websites like Quora. The study demonstrates that DL models outperform ML models in accurately classifying serious and insincere inquiries with fewer errors. Moreover, using convolution layers has enhanced model performance, and implementing complicated designs and modifications has resulted in improved prediction precision. These findings suggest that employing pure DL designs and complex architecture can significantly improve performance.

The study's results also have practical implications for Quora and other similar websites, as accurately classifying questions is crucial for maintaining a high-quality community Q&A platform (Table 1). DL models and their advanced architectures can help ensure the platform can accurately categorize inquiries, providing users with relevant, high-quality answers. Overall, this research highlights the potential of DL models for improving the accuracy of question classification and has important implications for developing future NLP applications.

TABLE I. PERFORMANCE OF DL MODELS

Reference	Model	Accuracy
[24]	RNN	85.5%
[20]	CNN	90.49%
[25]	XLNet	95.51%
[26]	CapsuleNet	92.60%
[23]	GRU	89.49%

VI. LIMITATIONS & FUTURE RESEARCH

DL models have shown promise in classifying insincere questions and improving question classification. Pre-trained DL models, such as BERT and its variations, have demonstrated their ability to predict words from context using large text corpora [27]. These models can be fine-tuned for downstream operations by adding task-specific layers. However, due to their large number of parameters (often in the hundreds of millions), fine-tuning and online serving of these models can be challenging in real-life applications [28]. Most DL models need lots of memory for training and inference [9]. In recent years, the demand for small models that can operate on the edge has led to the development of model compression or knowledge distillation techniques, which can produce minor "student" models from larger, pre-trained "teacher" models [28]. Pre-trained models have been shown to enhance accuracy while reducing computational power. This has brought about a significant change in problem-solving approaches. For example, Vaswani et al. [12] proposed a Transformer architecture, which powers BERT's multi-layer, bidirectional encoder, which has become a cornerstone in NLP research [7].

As a result, various performance-enhanced BERT architectures have been proposed, including RoBERTa, DistilBERT, and ALBERT, which offer remarkable improvements in performance and speed [29]. The Transformer model, introduced by Vaswani et al. [12] in 2017, outperforms other neural models, such as convolutional and recurrent models for NLP and synthesis. Its manageable model size allows for realistic parallel training, efficient sequence feature collection, and customizable architecture [7]. The open-source Transformers community and library allow users to design and implement large-scale pre-trained models for high-performance applications. Creating an ensemble of pre-trained language models with less data and processing capacity can improve future studies, helping to discover critical questions and manage Q&A platforms.

VII. CONCLUSION

In our exploration of Deep Learning (DL) techniques as a solution to address the issue of text representation in Machine Learning (ML) approaches, we found that these techniques, inspired by the functioning of the human brain, effectively utilize the inherent structure of language, specifically the order of words, to generate more accurate models.

Among the various DL architectures employed, Recurrent Neural Networks (RNN), Long Short-Term Memory networks (LSTM), and Convolutional Neural Networks (CNN) were prominently used. Our research indicated a particular effectiveness of CNNs in the field of text classification due to their ability to capture local information effectively. We found that even a simple CNN model could outperform more complex DL models that use intricate pooling strategies, demonstrating their efficacy in accurately classifying insincere questions.

However, our research also highlighted a limitation of CNNs, specifically the potential for information loss during the pooling process. To address this, we explored Capsule Networks as a promising alternative. These networks aim to provide spatial meaning to the text, thereby preserving more information during the classification process.

Like other DL models, we found that Capsule Networks require substantial data to be effective. Additionally, feature engineering often proved necessary to enhance their performance. Our research indicates that with the right amount of data and appropriate feature engineering, Capsule Networks can significantly improve the accuracy of insincere question classification.

An important aspect of our research that warrants discussion is the issue of class imbalance in our dataset, as depicted in Figure 1. Class imbalance is a common problem in machine learning tasks, where one class of data significantly outnumbers the other. In the context of insincere question classification, the insincere questions, which are our class of interest, often constitute the minority class. This imbalance can lead to a bias in our models towards the majority class, resulting in poor performance on the minority class. Various strategies have been employed in the literature to address this issue, including balanced batch generators, oversampling techniques, and adjusting class weights in machine learning models. Despite these strategies, dealing with class imbalance remains a challenging aspect of insincere question classification, and further research is needed to develop more effective strategies for handling this issue.

In conclusion, our findings suggest that while CNNs are effective tools for text classification, Capsule Networks, with the right data and feature engineering, can provide a valuable alternative that mitigates some of the limitations of CNNs. The issue of class imbalance presents a significant challenge in this field, and future work should focus on developing novel strategies to address this problem.

REFERENCES

[1] S. Gandhi, “Classifying the Insincere Questions using Transfer Learning,” *NORMA EResearch NCI Libr.*, p. 24, Nov. 2021.

[2] A. Mungekar, N. Parab, P. Nima, and S. Pereira, “Quora Insincere Questions Classification,” *Natl. Coll. Irel.*, p. 8.

[3] K. Harsh, S. Gajanand, and H. Jatin, “Classification of Online Toxic Comments Using Machine Learning Algorithms,” *Int. Conf. Intell. Comput. Control Syst. ICICCS 2020*, p. 5, 2020.

[4] J. Wang, Y. Li, J. Shan, J. Bao, C. Zong, and L. Zhao, “Large-Scale Text Classification Using Scope-Based Convolutional Neural Network: A Deep Learning Approach,” *IEEE Access*, vol. 7, pp. 171548–171558, 2019, doi: 10.1109/ACCESS.2019.2955924.

[5] S. Qaiser and R. Ali, “Text Mining: Use of TF-IDF to Examine the Relevance of Words to Documents,” *Int. J. Comput. Appl.*, vol. 181, no. 1, pp. 25–29, Jul. 2018, doi: 10.5120/ijca2018917395.

[6] B. Gaire, B. Rijal, D. Gautam, S. Sharma, and N. Lamichhane, *Insincere Question Classification Using Deep Learning*. 2019. doi: 10.13140/RG.2.2.30392.49925.

[7] A. Rachha and G. Vanmane, “Detecting Insincere Questions from Text: A Transfer Learning Approach.” arXiv, Dec. 07, 2020. Accessed: Aug. 09, 2022. [Online]. Available: http://arxiv.org/abs/2012.07587

[8] X. Li and D. Roth, “Learning Question Classifiers,” in *COLING 2002: The 19th International Conference on Computational Linguistics*, 2002. Accessed: Oct. 03, 2022. [Online]. Available: https://aclanthology.org/C02-1150

[9] S. Minaee, N. Kalchbrenner, E. Cambria, N. Nikzad, M. Chenaghlu, and J. Gao, “Deep Learning--based Text Classification: A Comprehensive Review,” *ACM Comput. Surv.*, vol. 54, no. 3, p. 62:1-62:40, Apr. 2021, doi: 10.1145/3439726.

[10] R. Katarya and Y. Arora, “Study on Text Classification using Capsule Networks,” in *2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS)*, Mar. 2019, pp. 501–505. doi: 10.1109/ICACCS.2019.8728394.

[11] S. Sabour, N. Frosst, and G. E. Hinton, “Dynamic Routing Between Capsules,” in *Advances in Neural Information Processing Systems*, Curran Associates, Inc., 2017. Accessed: Aug. 18, 2022. [Online]. Available: https://proceedings.neurips.cc/paper/2017/hash/2cad8fa47bbef282badbb8de5374b894-Abstract.html

[12] A. Vaswani *et al.*, “Attention is All you Need,” *Adv. Neural Inf. Process. Syst.*, vol. 30, p. 11.

[13] S. Wankmüller, *Neural Transfer Learning with Transformers for Social Science Text Analysis*. 2021. [Online]. Available: https://doi.org/10.48550/arXiv.2102.02111

[14] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015, doi: 10.1038/nature14539.

[15] T. Lei, R. Barzilay, and T. Jaakkola, “Molding CNNs for text: non-linear, non-consecutive convolutions,” in *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, Lisbon, Portugal: Association for Computational Linguistics, Sep. 2015, pp. 1565–1575. doi: 10.18653/v1/D15-1180.

[16] P. Liu, X. Qiu, and X. Huang, “Recurrent Neural Network for Text Classification with Multi-Task Learning,” *Twenty-Fifth Int. Jt. Conf. Artif. Intell.*, pp. 2873–2879, Jul. 2016, doi: 10.5555.

[17] K. S. Tai, R. Socher, and C. D. Manning, “Improved Semantic Representations From Tree-Structured Long Short-Term Memory Networks,” in *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, Beijing, China: Association for Computational Linguistics, Jul. 2015, pp. 1556–1566. doi: 10.3115/v1/P15-1150.

[18] M. Coavoux and B. Crabbé, “Incremental Discontinuous Phrase Structure Parsing with the GAP Transition,” in *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, Valencia, Spain: Association for Computational Linguistics, 2017, pp. 1259–1270. doi: 10.18653/v1/E17-1118.

[19] C. Chen, “Insincere Question Classification by Deep Neural Networks,” in *2021 IEEE International Conference on Computer Science, Artificial Intelligence and Electronic Engineering (CSAIEE)*, Aug. 2021, pp. 15–23. doi: 10.1109/CSAIEE54046.2021.9543161.

[20] X. Zhang, J. Zhao, and Y. LeCun, “Character-level Convolutional Networks for Text Classification,” in *Advances in Neural Information Processing Systems*, Curran Associates, Inc., 2015. Accessed: Nov. 11, 2022. [Online]. Available: https://papers.nips.cc/paper/2015/hash/250cf8b51c773f3f8dc8b4be867a9a02-Abstract.html

[21] Z. Han, J. Gao, H. Sun, R. Liu, C. Huang, and H. Qi, “An Ensemble Learning-based model for Classification of Insincere Question,” *Forum Inf. Retr. Eval.*, vol. 2517, p. 7, 2019.

[22] M. Zulqarnain, A. K. Z. Alsaedi, R. Ghazali, M. G. Ghouse, W. Sharif, and N. A. Husaini, “A comparative analysis on question classification task based on deep learning approaches,” *PeerJ Comput. Sci.*, vol. 7, p. e570, Aug. 2021, doi: 10.7717/peerj-cs.570.

[23] D. Mediratta and N. Oswal, “Detect Toxic Content to Improve Online Conversations,” *ArXiv*, vol. abs/1911.01217, p. 10, 2019, doi: 10.48550.

[24] Q. Chen, X. Zhu, Z.-H. Ling, S. Wei, H. Jiang, and D. Inkpen, “Recurrent Neural Network-Based Sentence Encoder with Gated Attention for Natural Language Inference,” in *Proceedings of the 2nd Workshop on Evaluating Vector Space Representations for NLP*, Copenhagen, Denmark: Association for Computational Linguistics, Sep. 2017, pp. 36–40. doi: 10.18653/v1/W17-5307.

[25] Z. Yang, Z. Dai, Y. Yang, J. Carbonell, R. R. Salakhutdinov, and Q. V. Le, “XLNet: Generalized Autoregressive Pretraining for Language Understanding,” *Curran Assoc. Inc*, vol. 32, p. 11, 2019.

[26] W. Zhao, J. Ye, M. Yang, Z. Lei, S. Zhang, and Z. Zhao, “Investigating Capsule Networks with Dynamic Routing for Text Classification,” Mar. 2018.

[27] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, Minneapolis, Minnesota: Association for Computational Linguistics, Jun. 2019, pp. 4171–4186. doi: 10.18653/v1/N19-1423.

[28] W. Wang, F. Wei, L. Dong, N. Yang, and M. Zhou, “MINILM: Deep Self-Attention Distillation for Task-Agnostic Compression of Pre-Trained Transformers,” *Adv. Neural Inf. Process. Syst.*, vol. 33, pp. 5776--5788, 2020.

[29] T. Wolf *et al.*, “Transformers: State-of-the-Art Natural Language Processing,” in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, Online: Association for Computational Linguistics, Oct. 2020, pp. 38–45. doi: 10.18653/v1/2020.emnlp-demos.6.