Detection of AI-written and Human-written Text using Deep Recurrent Neural Networks

Ghazanfar Latif [1, *], Nazeeruddin Mohammad [1], Ghassen Ben Brahim [1], Jaafar Alghazo [2], Khaled Fawagreh [1]

[1] Department of Computer Science, Prince Mohammad Bin Fahd University, Khobar, Saudi Arabia.

[2] Artificial Intelligence Research Initiative, College of Engineering and Mines, University of North Dakota Grand Forks, ND 58202, USA.

*Corresponding author: glatif@pmu.edu.sa

ABSTRACT

With the development of Artificial Intelligence (AI)-based language models, it is becoming pertinent and will become even more pertinent in the future to be able to distinguish between AI-based generated text and human-based generated text. The implications of having humans present work generated by AI language models and claiming it their own have serious implications at all levels with the basic being the ethical implication. In this paper, we propose the use of modified deep learning models using the Deep Recurrent Neural Network (DRNN) for the classification of text to be either AI-generated or human-written. Two modified architectures are proposed DRNN-1 and DRNN-2. This led to the second contribution of this work which is the development of a dataset containing short answers to simple questions in Information Technology (IT), Cybersecurity, and Cryptography given to junior and senior students in Computer Engineering & Science, and IT to produce a total of 450 answers. The same questions were given to ChatGPT for a total of 450 answers. The combined dataset consisted of 900 answers in the three domains. Though both proposed architectures produced good results, the DRNN-2 achieved better results with a test accuracy of 83.78% using the cybersecurity questions alone and 88.52% using the combined total dataset. This is considered one of the very excellent results achieved in this new emerging field of research.

Keywords: AI-written Text, Human Written Text, ChatGPT, Bard, Recurrent Neural Network, RNN, OpenAI, NLP.

1. INTRODUCTION

The future of AI-generated text is promising and rapidly evolving. As technology advances and AI models become more sophisticated, AI-generated text will likely play an increasingly important role in various industries, including journalism, marketing, and customer service, among others. Some potential future applications of AI-generated text include Content creation, Personalized writing, Translation, Customer service, and Improved accessibility. There are different AI based text generative models are proposed such as ChatGPT and GPT4 by Microsoft/OpenAI [1], Bard by Google [2] and LLaMA by Meta [3]. Elon Musk recruited a team to counter OpenAI ChatGPT [4].

ChatGPT is a language model developed by OpenAI, a research organization focused on advancing artificial intelligence [5]. It is a type of Generative Pretrained Transformer (GPT) model, which means it has been trained on a large corpus of text data to generate text based on a given prompt [6]. ChatGPT has been trained on a diverse range of text data, including web pages, books, and other sources, and has been fine-tuned to generate high-quality text for a variety of language tasks, including text completion, question-answering, and text generation.

ChatGPT uses a mixture of supervised and reinforcement learning approaches and applies a set of 175 Billion parameters to train the model [7]. It generates text by using a transformer-based neural network language model trained on a large corpus of text data [8]. The model uses an attention mechanism to consider the context and generate a sequence of words, one word at a time until the desired length of the text is generated or a termination token is reached. The generated text is based on the input prompt and the patterns learned from the training data.

In some cases, AI-generated text can be of high quality and indistinguishable from text written by a human. However, it's important to note that the text generated by AI is based on patterns learned from the training data, so it may not always reflect the nuances and complexities of human language. Additionally, the model may generate text that is biased or offensive, depending on the training data it was exposed to. In general, AI-generated text can be useful for tasks such as text completion, text summarization, and text generation from a prompt, where the model can leverage its knowledge of language patterns to produce coherent and relevant text. However, for tasks that require a deeper understanding of human emotions, motivations, and experiences, human-generated text may still be superior [9].

With the advancement of AI-based models, it is becoming increasingly important to distinguish between human-based and AI-generated text. Classifying AI-generated text versus human-written text is important for ensuring the authenticity, transparency, and fairness of information, as well as for regulatory and policy purposes [10]. As the use of AI-generated text continues to grow, it will become increasingly important to develop methods for accurately and reliably classifying AI-generated text. Classifying AI-generated text versus human-written text is important for several reasons as given below.

- Verification and authenticity: Classifying AI-generated text versus human-written text can help to verify the authenticity and accuracy of the information, especially in the context of news articles, scientific publications, or other types of information that may have important implications for individuals or society.
- Trust and transparency: Classifying AI-generated text can also help to build trust and transparency with audiences, who may want to know whether the information they are reading has been generated by a human or by an AI system.
- Bias and fairness: AI-generated text systems may reflect biases that are present in the training data used to train them, or may perpetuate harmful stereotypes or prejudices. By classifying AI-generated text, it becomes possible to identify these biases and address them, helping to promote fairness and reduce harm.
- Regulation and policy: In some contexts, such as political advertising or financial disclosure, it may be important to distinguish between AI-generated text and human-written text for regulatory or policy reasons.

The motivation for this work is thus apparent to be part of the original contributing research to distinguish between human-based and AI-generated text. This line of research will continue to grow as the complexity of the problem continues to grow in parallel. ChatGPT is one such example and many more variants and versions will appear in future.

The aim of this research is to develop modified machine-learning algorithms that can achieve binary classification to classify AI-based and human-based text. In this paper, we propose two modified deep recurrent neural network (DRNN) models we name DRNN-1 and DRNN-2 [11-12]. Several machine learning algorithms can be used in combination or isolation, depending on the specific requirements of the task and the resources available to classify AI-generated text versus human-written text. Statistical analysis of the language used in the text can be used to distinguish between AI-generated text and human-written text. For example, features such as word frequency, sentence length, or syntactic complexity can be used to identify patterns that are characteristic of AI-generated text. Model-based approaches can be used to train machine learning algorithms to distinguish between AI-generated text and human-written text. These models can be trained on a large corpus of text and use features such as word frequency, n-grams, or other linguistic features to make predictions. Syntactic tree analysis can be used to examine the structure of sentences in the text and distinguish between AI-generated text and human-written text [13]. For example, AI-generated text may be more likely to produce parse trees with repetitive or unnatural structures. Human evaluation can also be used to classify AI-generated text versus human-written text by having human annotators examine a set of texts and make judgments about whether each text was generated by a human or an AI system. Content analysis can also be used to classify AI-generated text versus human-written text by examining the content of the text and making judgments about its sourceIt is important to keep in mind that no single method is perfect and that different methods may have different strengths and weaknesses, so it is often useful to use multiple methods to get a more comprehensive view of the text and its origin. However, we hope that our proposed models will be able to achieve better results than those reported in the extant literature and will be a stepping stone toward the future development of solution to this pertinent problem.

1.1 Applications of Large Language Models

Software systems developed based on large language models such as GPT3 and GPT4 have many useful practical benefits [14]. These systems can be used for the for various purposes including the following:

- Text generation: They can be used to generate well-written and coherent texts in a wide range of topics, styles, and languages easily and swiftly. For instance, news summaries, stories, poems, music, product descriptions, etc. can be generated
- Problem resolver: These systems have the ability to analyze and find the solution to problems covering a wide variety of topics. For instance, finding bugs in software programs, solving math questions, grading exams, etc.
- Universal answer generator: They can generate consistent and appropriate answers in an extremely large range of contexts. For instance: find answers to students' homework and assignment questions, translate a particular text, write

code to implement some functionality, etc.

- Social network support: They can help in this area due to its ability in generating messages for social networks as well as attractive posts, etc.
- Platform Emulator: these systems can be used as an emulator for some particular existing platforms. For instance, ChatGPT may be instructed to pretend and act as a high-level UNIX emulator executing shell commands, creating directories and files, typing and compiling code, etc.

Like any technology, language models can be misused. It is important to be aware of these potential misuses and to use language models responsibly, with a focus on ethical and responsible practices [15]. This may include using high-quality, diverse training data, being transparent about the source and limitations of the model's text, and being vigilant about the potential for biases and misinformation.

1.2 AI-written Text impact on Content Writers

AI-generated text systems have the potential to automate certain aspects of the writing process, which may have an impact on the demand for certain types of writing jobs and can also have impact on the learning trends [16-17]. However, it is important to note that while AI-generated text systems may be able to produce text quickly and efficiently, they do not have the same level of creativity, nuance, and critical thinking skills as experienced human writers. In many cases, AI-generated text may be used to supplement or support the work of human writers, rather than replace them completely. For example, AI-generated text may be used to generate outlines or summaries of information, which can then be refined and expanded upon by human writers. That being said, the use of AI-generated text may lead to some changes in the job market for content writers, as well as changes in the types of skills and expertise that are in demand. In this sense, experienced content writers may need to adapt and evolve their skills to stay competitive in an increasingly automated industry.

Overall, AI-generated text systems have the potential to impact the job market for content writers, but they are unlikely to replace the creativity and critical thinking skills of human writers. Experienced content writers who are able to adapt to new technologies and continue to develop their skills and expertise will likely continue to play an important role in the writing and content creation industry.

2. MATERIALS AND METHODS

The proposed machine learning-based model aims to identify genuine students' answers from those generated by the AI systems. The methodology uses deep-learning-based neural networks for classification. The classification process starts by training the model with our dataset, where each entity is labelled as either a "AI-based answer" or a "human-based answer". This makes the problem falls under the category of binary classification. It is worth noting that the built dataset is evenly partitioned in both class labels. Fig. 1 depicts the workflow of the proposed system. It shows all model building blocks capturing all steps involved in the classification process starting from dataset labeling to text vectorization, to dataset partition into training and validation subsets, to finally classifiers.

The preprocessing phase consists of manually annotating the dataset with either of the two labels. During the next phase, text vectorization is performed on full sentences through the tokenization of words. In the following step, text features are extracted then the dataset is partitioned into training and validation sets. The data partition was performed such that 80% of the data is allocated for training and the remaining (20%) for model testing. The chosen classifiers then, classify the text into either of the classes. The envisioned system was designed to be accurate, autonomous, and with the option of being linked to a cloud-based service for real-time processing.

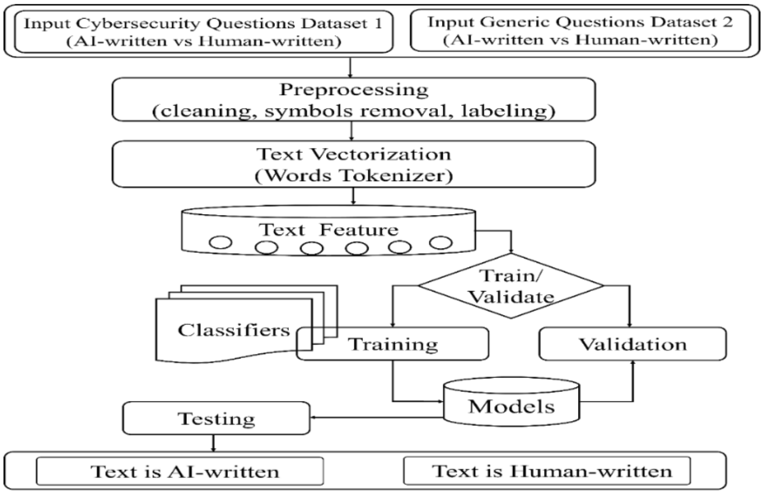
Input Cybersecurity Questions Dataset 1
(AI-written vs Human-written)
Input Generic Questions Dataset 2
(AI-written vs Human-written)
Preprocessing
(cleaning, symbols removal, labeling)
Text Vectorization
(Words Tokenizer)
Text Feature
Train/
Validate
Classifiers
Training
Validation
Models
Testing
Text is AI-written
Text is Human-written

Figure 1. Workflow of the proposed system for text classification.

2.1 Dataset Description

The dataset of the present study is prepared at Prince Mohammad Bin Fahd University (PMU). The questions for this study are generic questions related to information technology, cyber security, and cryptography. The questions were intentionally kept simple to allow any computing and IT major student to answer from his/her understanding. A total of 150 questions were created. The sample questions are shown in Table 1. Students from four different sections are chosen to answer the questions. All students are at the junior level (3rd year) or senior level (4th year) of their degree. The student population included both male and female students and screening is done based on their course grades.

To eliminate any external influences on the data, each student is assigned only three random questions from the pool and given sufficient time (24 hours) to answer them. For each sample question, on average 3 answers from three different students are collected. Similarly, three auto-generated answers are obtained from ChatGPT for each question. Thus, the dataset has 900 responses, which included 450 generative AI answers and 450 human-written answers. Both students and ChatGPT are instructed to keep their answers short.

Table 1: Sample Questions used to get the AI-written and Human-written text for the experiments

Sample Cybersecurity Questions	**Sample Generic Questions**
What is CIA triad in cybersecurity?	What is Operating System?
How do you keep your computer secure from hackers?	Why we used to have RAM memory in Computers?
How to keep safe our passwords?	What is Robot?
Explain what is a good password	What is Machine Learning?
What are the risks of using free software?	What is Cloud Computing?
How to protect ourselves from spam?	What to do when my computer crashes?
What is one-time pad cipher?	What is BIOS?
Describe monoalphabetic cipher?	What is Computer Vision?
Why iPhones are considered more secure?	What is benefit of Graphics Processing Unit (GPU)?
What is defense in depth?	What is the use of Motherboard in Computer?
What is Spam?	What is Object Oriented Programming?
What is cryptanalysis?	What is a processor in computer systems?
Explain difference between passive and active attacks?	What is the meaning of a programming language?
What is brute force attack?	What is Multitasking in Computers?
Why wireless communications are not safe?	What is a microprocessor?
What is Eavesdropping?	What is an array?
Why CAPTCHA is necessary?	Why we used to have variables in programming?
Why is it important to have a strong password?	What is URL?
What is VPN?	What is Internet of Things?

2.2 Preprocessing and Word Embedding Layer

The preprocessing stage plays a crucial role in improving the accuracy of the dataset under study. During this stage, irrelevant data was removed and diacritics, symbols, special characters, etc. were stripped away. A dictionary was created to map similar Unicode representations to fixed characters.

Each text of every answer (student-based or AI based) was processed by applying a default tokenizer to remove symbols and was then converted to tensors for numerical data suitable for neural networks. Each dataset entity was accurately labelled to either of the classes (human versus AI). The labels were converted to tensors or float data types, with a default batch size of 32.

For word embedding, the models selected were Word2Vec and GloVe, which allowed the system to learn word vectors based on their co-occurrence in sentences [18-19]. This representation in a lower-dimensional vector space enabled deep learning algorithms to map words based on semantic similarity. Fig. 2 shows the two architectures for Word2Vec: a continuous bag of words (CBOW) and skip-gram (SG). CBOW predicts the target word using its contextual context, while SG predicts the surrounding words based on the target word. The data was split into 60% training, 20% validation, and 20% testing, as previously mentioned.

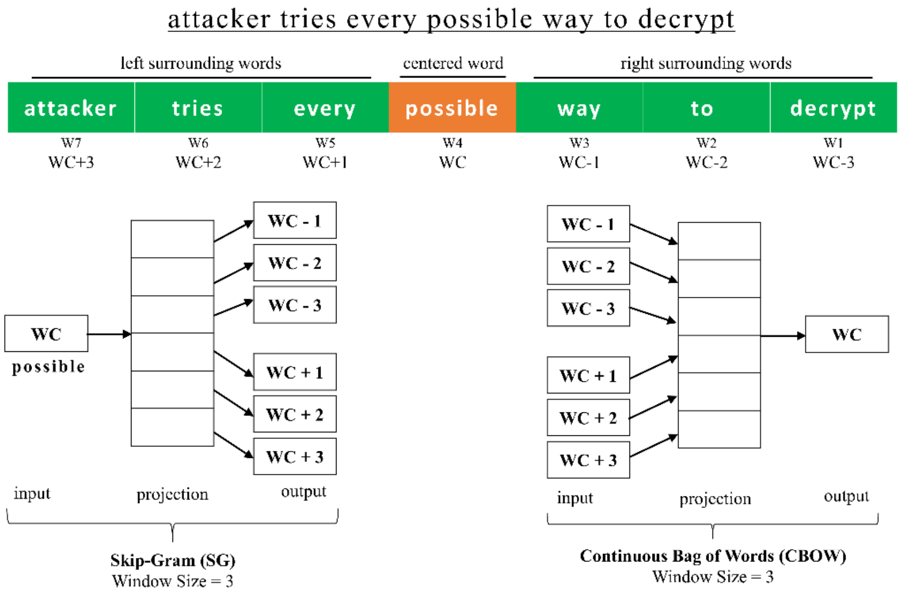

Figure 2. Word2Vec architecture based on CBOW and SG (example text: attacker tries every possible way to decrypt).

2.3 Proposed Deep Recurrent Neural Network Model for text classification

The widespread adoption of deep learning in almost all data processing problems was mainly due to the tremendous existence of today's computing machine processing power. A Convolutional Neural Network (CNN) is a deep neural network that can take advantage of the inherent properties of data, especially from images and text [20-22]. Recurrent Neural Networks (RNNs) are similar to CNNs but also manage time-related data [23]. In a unidirectional (feedforward) RNN, the model learns word by word in a sequential manner, from the start to the end.

In the text, the meaning of words and sentences can change based on the context in which they are used and the surrounding words in a sentence, as well as every human, has their writing style where specific words are used more frequently. To cope with the different meanings or semantics, or words in various contexts, the bidirectional RNN as shown in Fig. 3 is used which considers two sequences: a forward sequence (left to right) and a backward sequence (right to left). In text classification, the time-related aspect of the data, which represents the order of words in the input text, plays a crucial role in improving the accuracy of the classification results.

The proposed deep bidirectional RNN model for text classifications consists of ten layers, including the input layer, bidirectional RNN layers, flatten layer, dense layers, dropout layers, and output layers. The model uses the embedding layer produced by Word2Vec and GloVe, which is input into the backward RNN layer, followed by the forward RNN layer [19]. The output of the RNN is then combined, flattened, and fed into the dense and dropout layers. The data is finally classified into either “AI” or “Human” based.

The proposed model as shown in Fig. 4 uses the tanh activation function for the dense layers and the sigmoid activation function for the output layer. The experiments were run using a learning rate of 0.001 with 32 batch sizes and 50 iterations. The best model, based on validation accuracy, was saved out of the 50 iterations and used on the test dataset to generate results and make comparisons.

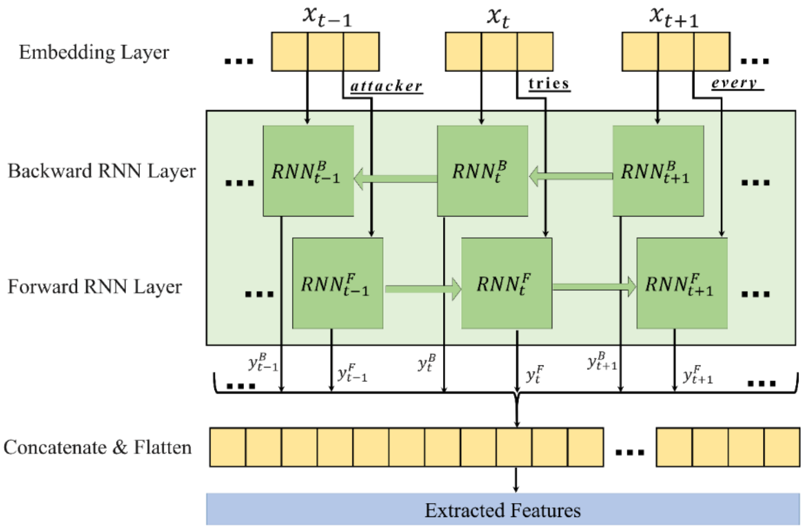

Figure 3. Proposed architecture of the deep recurrent neural network model for feature extraction

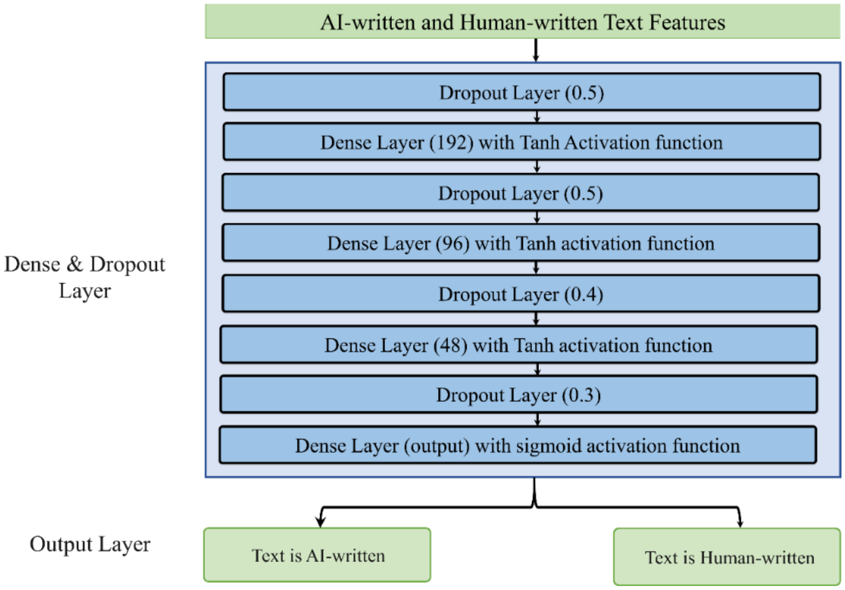
AI-written and Human-written Text Features
Dropout Layer (0.5)
Dense Layer (192) with Tanh Activation function
Dropout Layer (0.5)
Dense Layer (96) with Tanh activation function
Dropout Layer (0.4)
Dense Layer (48) with Tanh activation function
Dropout Layer (0.3)
Dense Layer (output) with sigmoid activation function
Dense & Dropout Layer
Output Layer
Text is AI-written
Text is Human-written

Figure 4. Architecture of the convolutional operations for text classification from the extracted features.

3. RESULT AND DISCUSSION

Three different types of experiments were performed using the newly built dataset of AI-written and Human-written text. The experiments were performed using hardware with a Nvidia 1080TI GPU and 32-gigabyte memory; the software was implemented using Python programming. A set of five metrics were considered to evaluate the proposed classification model in terms of its ability in identifying AI-based answers from genuine human-based ones which includes: (1) accuracy, (2) confusion matrix, (3) recall, (4) precision, and (5) F1 measure. The experimental dataset was divided into 80% training and 20% validation. The 20% validation data were also split into 80% validation and 20% verification. Different machine learning methods were also used to compare the results with various deep recurrent neural networks models, such as decision trees (DT), multilayer perceptron (MLP), naïve Bayes (NB), and linear regression (LR) [24]. The experimental results were evaluated based on their accuracy, precision, recall, F1 score, and confusion matrices.

3.1 AI-written and Human-written Text Classification from Cybersecurity-Related Questions

The first experiment was based on the binary classification of comments as AI-based or human-based. Table 3 shows the results of using the DT, MLP, NB, and LR classifiers. It can be seen from Table 2 that the highest test accuracy (72.97%) was achieved using DT; however, this was still a very low percentage. The precision, recall, and F1 score were the highest for DT as well, with values of 0.74, 0.74, and 0.73.

Table 2: Comparison of experimental results using typical classification methods for Cybersecurity related Questions.

Method	Accuracy	Precision	Recall	F1 Score
DT	72.97%	0.74	0.74	0.73
MLP	60%	0.60	0.59	0.60
NB	67.5%	0.69	0.66	0.66
LR	52%	0.51	0.51	0.51

A repeat of the same experiment using the proposed deep RNN model showed a significant improvement in the binary classification, as shown in Table 3. There were two proposed architectures: one consisting of seven deep layers with 436,322 parameters, referred to as DRNN-1, and the other with 10 deep layers and 359,090 parameters, referred to as DRNN-2. Using the DRNN-1 architecture, the training accuracy increased to 99.37%, with a loss of 0.013 and a validation loss and validation accuracy of 1.71 and 77.77%.

Table 3. Experimental results using the proposed DRNN-based classification methods for Cybersecurity related Questions.

Model	Iterations	Train Loss	Train Acc.	Val. Loss	Val. Acc.
7 deep layers with 436,322 parameters (DRNN-1)	10	0.4839	78.48%	0.5966	61.11%
	25	$1.18e^{-5}$	100%	0.6667	66.67%
	50	0.2041	90.51%	0.8676	72.22%
	Best	0.0134	99.37%	1.7152	77.77%
10 deep layers with 359,090 parameters (DRNN-2)	10	0.1188	95.89%	0.5676	75.76
	25	$4.707e^{-6}$	100%	1.7129	84.85
	50	$1.041e^{-4}$	100%	2.3070	81.82
	Best	0.0013	100%	0.9205	87.88

With DRNN-2, the highest training accuracy achieved was 100% after 25 iterations with a training loss of 0.0013, while the validation loss and validation accuracies were 0.9205 and 87.88%, respectively It was observed that DRNN-2 performed better than the DRNN-1, while both achieved better results than the classifiers listed in Table 2.

3.2 AI-written and Human-written Text Classification from Combined Questions

The third experiment was based on combining all questions (Information Technology and Cybersecurity). Table 5 shows the results of using the DT, MLP, NB, and LR classifiers. It can be seen from Table 4 that the accuracy of DT is not the best, and the highest accuracy (72%) was achieved using RF. Nonetheless, this was still a very low percentage. The precision, recall, and F1 score were the highest for RF as well, with values of 0.80, 0.72, and 0.63, respectively.

Table 4. Comparison of experimental results using the typical classification methods for combined Generic IT and Cybersecurity related Questions.

Method	Accuracy	Precision	Recall	F1 Score
DT	67%	0.66	0.65	0.67
MLP	62%	0.59	0.62	0.60
NB	52%	0.57	0.52	0.54
LR	57%	0.59	0.59	0.58
RF	72%	0.80	0.72	0.63

A repeat of the same experiment using the proposed deep RNN models showed significant improvement in the classification of the binary classes using the more complex compared to those in table 4 dataset, as shown in Table 5 Both DRN-1 and DRNN-2 reached a training accuracy of 100% with the DRNN-1 performing slightly better on the other metrics. The training loss for DRNN-1 shows at 1.58e^(-4) while for the DRNN-2 it is 1.081e^(-9). The training loss and validation accuracy are 0.8202 and 89.09% respectively for DRNN-1 and 2.2182 and 90.91% for DRNN-2. It can be observed that the proposed model showed better results than those listed in Table 5.

Table 5. Experimental results using the proposed DRNN-based classification methods for combined Generic IT and Cybersecurity related Questions.

Model	Iterations	Train Loss	Train Acc.	Val. Loss	Val. Acc.
7 deep layers with 436,322 parameters (DRNN-1)	10	0.0235	99.19%	0.7174	85.45%
	25	$1.52e^{-5}$	100%	1.0976	83.64%
	50	$1.94e^{-5}$	100%	1.1686	85.45%
	Best	$1.58\ e^{-4}$	100%	0.8202	89.09%
10 deep layers with 359,090 parameters (DRNN-2)	10	0.0236	99.19%	1.3536	83.64%
	25	$1.01e^{-7}$	100%	1.4615	85.45%
	50	$5.79e^{-9}$	100%	2.3763	87.27%
	Best	$1.081e^{-9}$	100%	2.2182	90.91%

Table 6 shows the test accuracies of the proposed models when applying the test dataset portion using the dataset for cybersecurity and then applying the combined dataset. The highest test accuracy for the dataset containing the cybersecurity questions was 83.78% using the DRNN-2 architecture. It is also observed that the combined dataset achieved even better results with a test accuracy of 88.52% again using the DRNN-2 architecture. The DRNN-2 proposed architecture performed better than the DRNN-1 which is easily explained because of the added deep layers.

Table 6. Comparison of Test Accuracies for the different proposed DRNN-1 and DRNN-2 models.

Model	Dataset	Test Accuracy	Test Loss
7 deep layers with 436,322 parameters (DRNN-1)	Cybersecurity Questions only	70.00%	1.1529
	Combined	81.97%	0.9586
10 deep layers with 359,090 parameters (DRNN-2)	Cybersecurity Questions only	83.78%	0.8609
	Combined	88.52%	0.3628

The results presented in Tables 2 to 5 demonstrate that machine learning techniques can differentiate between AI-generated text and human-written text. It is evident that deep learning models are better at classifying text. This implies that there are certain features that distinguish human writing from AI-generated text. This is specifically true for the text generated by the average university student. However, further analysis is needed to understand the distinguishing features to achieve high accuracy. Similarly, there is a need for a dataset with large samples of text, written by experts from different fields.

4. CONCLUSION

Overall, the future of AI-generated text holds great potential for improving efficiency, accessibility, and personalization in many different industries. However, it is also important to be mindful of the potential drawbacks and biases that may arise from the use of AI-generated text and to develop and implement ethical and responsible practices for its use. The implications of having humans present work generated by AI language models and claiming it their own have serious implications at all levels with the basic being the ethical implication. In this paper, we propose the use of modified deep learning models using the Deep Recurrent Neural Network (DRNN) for the classification of text to be either AI-based or human-based. Two modified architectures are proposed DRNN-1 and DRNN-2. This led to the second contribution of this work which is the development of a dataset containing short answers to simple questions in Information Technology (IT) and Cybersecurity given to junior and senior students in Computer Engineering & Science, and IT to produce a total of 450 answers. The same questions were given to ChatGPT for a total of 450 answers. The combined dataset consisted of 900 answers in the three domains. Though both proposed architectures produced good results, the DRNN-2 achieved better results with a test accuracy of 83.78% using the cybersecurity questions alone and 88.52% using the combined total dataset. This is considered one of the very excellent results achieved in this new emerging field of research.

The authors envision that the field of research to distinguish between AI-based text and human-based text will continue to grow and will become more complex in the future as the technology that drives the AI-based language models continues to develop. It also becomes more complicated when you try to distinguish other variations of text such as solutions to mathematical problems, or machine-generated programming code. Machine Learning researchers will need to come up with innovative ways to tackle these research areas.

REFERENCES

[1] Wang, F. Y., Miao, Q., Li, X., Wang, X., & Lin, Y. (2023). What does chatGPT say: the DAO from algorithmic intelligence to linguistic intelligence. *IEEE/CAA Journal of Automatica Sinica*, *10*(3), 575-579.

[1] Rahaman, M., Ahsan, M. M., Anjum, N., Rahman, M., & Rahman, M. N. (2023). The AI Race is On! Google's Bard and OpenAI's ChatGPT Head to Head: An Opinion Article. *Mizanur and Rahman, Md Nafizur, The AI Race is on*.

[2] LLaMA: Open and Efficient Foundation Language Models - Meta Research | Meta Research (2023) Meta Research, 2023. Available from: https://research.facebook.com/publications/llama-open-and-efficient-foundation-language-models/.

[3] Reuters (2023) Reuters, Elon Musk recruits team to develop OpenAI's ChatGPT rival - The Information, 2023. Available from: https://www.reuters.com/technology/elon-musk-recruits-team-develop-openai-rival-information-2023-02-28/.

[4] OpenAI Team. ChatGPT: Optimizing language models for dialogue. https://openai.com/blog/chatgpt , November 2022. Accessed: 2023-03-14.

[5] Luciano Floridi and Massimo Chiriatti. GPT-3: Its nature, scope, limits, and consequences. Minds and Machines, 30(4):681–694, 2020.

[6] Gilson, A., Safranek, C. W., Huang, T., Socrates, V., Chi, L., Taylor, R. A., & Chartash, D. (2023). How does CHATGPT perform on the United States Medical Licensing Examination? the implications of large language models for medical education and knowledge assessment. *JMIR Medical Education*, *9*(1), e45312.

[7] Roshanzamir, A., Aghajan, H., & Soleymani Baghshah, M. (2021). Transformer-based deep neural network language models for Alzheimer's disease risk assessment from targeted speech. *BMC Medical Informatics and Decision Making*, *21*, 1-14.

[8] Rapp, A., Curti, L., & Boldi, A. (2021). The human side of human-chatbot interaction: A systematic literature review of ten years of research on text-based chatbots. *International Journal of Human-Computer Studies*, *151*, 102630.

[9] King, M. R., & chatGPT. (2023). A conversation on artificial intelligence, chatbots, and plagiarism in higher education. *Cellular and Molecular Bioengineering*, 1-2.

[10] Popoola, S. I., Adebisi, B., Ande, R., Hammoudeh, M., Anoh, K., & Atayero, A. A. (2021). smote-drnn: A deep learning algorithm for botnet detection in the internet-of-things networks. *Sensors*, *21*(9), 2985.

[11] Anezi, F. Y. A. (2022). Arabic Hate Speech Detection Using Deep Recurrent Neural Networks. *Applied Sciences*, *12*(12), 6010.

[12] Quan, Z., Wang, Z. J., Le, Y., Yao, B., Li, K., & Yin, J. (2019). An efficient framework for sentence similarity modeling. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *27*(4), 853-865.

[13] Carlini, Nicholas, Florian Tramer, Eric Wallace, Matthew Jagielski, Ariel Herbert-Voss, Katherine Lee, Adam Roberts

et al. "Extracting Training Data from Large Language Models." In USENIX Security Symposium, vol. 6. 2021.

[14] Shen, Y., Heacock, L., Elias, J., Hentel, K. D., Reig, B., Shih, G., & Moy, L. (2023). ChatGPT and other large language models are double-edged swords. *Radiology*, 230163.

[15] Yang, D., Zhou, Y., Zhang, Z., Li, T. J. J., & LC, R. (2022, March). AI as an Active Writer: Interaction strategies with generated text in human-AI collaborative fiction writing. In *Joint Proceedings of the ACM IUI Workshops* (Vol. 10).

[16] Gayed, J. M., Carlon, M. K. J., Oriola, A. M., & Cross, J. S. (2022). Exploring an AI-based writing assistant's impact on English language learners. *Computers and Education: Artificial Intelligence*, *3*, 100055.

[17] Hossain, T., Mauni, H. Z., & Rab, R. (2022). Reducing the effect of imbalance in text classification using SVD and GloVe with ensemble and deep learning. *Computing and Informatics*, *41*(1), 98-115.

[18] Dharma, E. M., Gaol, F. L., Warnars, H. L. H. S., & Soewito, B. E. N. F. A. N. O. (2022). The accuracy comparison among Word2vec, Glove, and Fasttext towards convolution neural network (CNN) text classification. *Journal of Theoretical and Applied Information Technology*, *100*(2), 31.

[19] Al-Haddad, R., Sahwan, F., Aboalmakarem, A., Latif, G., & Alufaisan, Y. M. (2020, September). Email text analysis for fraud detection through machine learning techniques. In *3rd Smart Cities Symposium (SCS 2020)* (Vol. 2020, pp. 613-616). IET.

[20] Feinauer, D. M., Latif, G., Alenazy, A. M., Tayem, N., Alghazo, J., & Alzubaidi, L. (2022, December). Oil Spill Identification using Deep Convolutional Neural Networks. In *2022 14th International Conference on Computational Intelligence and Communication Networks (CICN)* (pp. 240-245). IEEE.

[21] Latif, G. (2022). DeepTumor: Framework for Brain MR Image Classification, Segmentation and Tumor Detection. *Diagnostics*, *12*(11), 2888.

[22] Ren, X., Gu, H., & Wei, W. (2021). Tree-RNN: Tree structural recurrent neural network for network traffic classification. *Expert Systems with Applications*, *167*, 114363.

[23] Latif, G., Alghazo, R., Pilotti, M. A., & Brahim, G. B. (2021). Identifying" At-Risk" Students: An AI-based Prediction Approach. *International Journal of Computing and Digital System*.