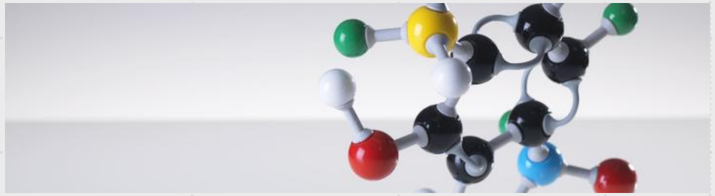
Combating Diabetes Misinformation in Africa: A Transformer Model Approach

Presented by Linda O. Okpanachi



Outline

- Academic and Research Profile
- Introduction
- Problem Statement
- Research Plan
- Research Questions
- Transformer Model Architecture
- Highlights of work done so far
- Publication
- Summary

Academic and Research Profile

- From Nigeria

- Bachelor Degree in Computer Science from Federal University Lokoja, Nigeria.

- Graduate Student Researcher at the Persuasive Technologies and Social Computing Lab.
- Masters Student at the University of British Columbia, Okanagan.

- **Research Interest include:**
- Machine Learning and Deep Learning
- Natural Language Processing
- Social Network Analysis

- **Email**
- lindaa15@student.ubc.ca

- **Linkedin**
- https://www.linkedin.com/in/linda-okpanachi-12327710b/

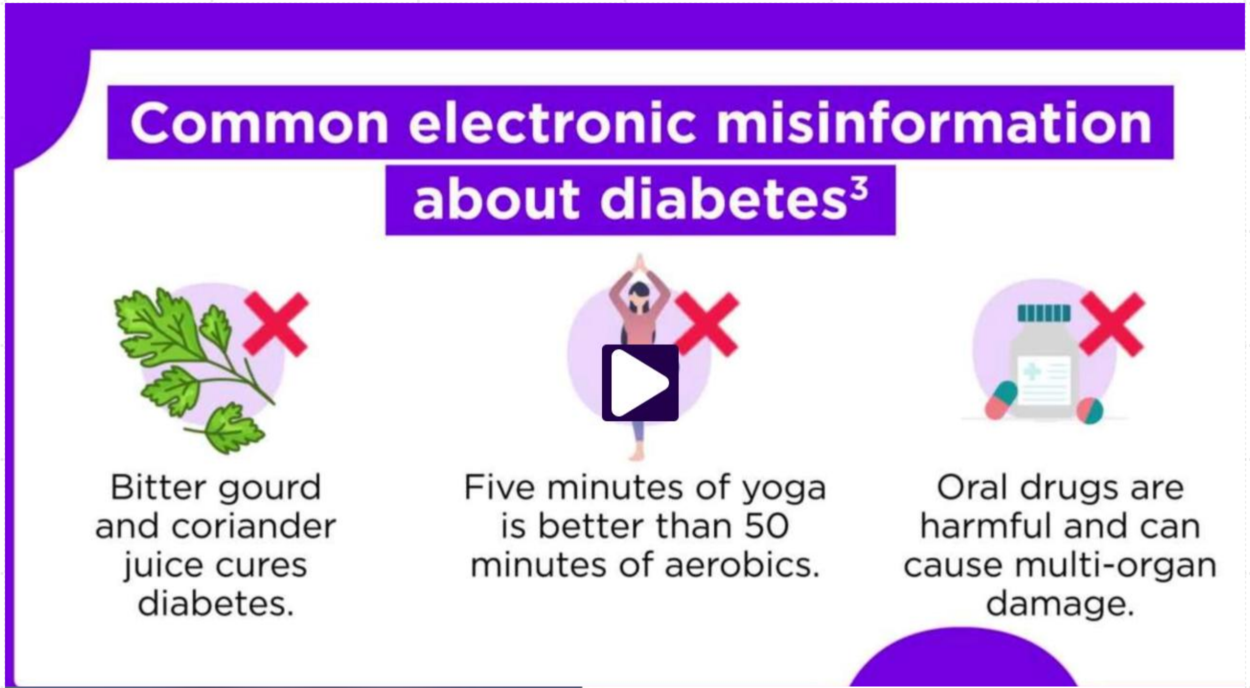
Introduction

- Misinformation refers to information that is incorrect or misleading, regardless of whether it is intentionally deceptive or not (Aimeur et al., 2023).
- Health misinformation is a growing concern, especially in low-resource settings.
- Misinformation around diabetes can lead to poor health outcomes due to improper care and self-management.
- In Africa, the spread of misinformation is amplified due to limited access to reliable health information (Ahinkorah et al., 2020).

Introduction

- Misinformation around diabetes, especially on social media, creates false perceptions about treatments, leading to harmful practices.
- Examples: False cures and myths about diabetes management circulating online.

Diabetes Myths and Misinformation



Images adapted from Sanofi

Problem Statement

- Prevalence of Health Misinformation
- Impact on Public Health
- Challenges in Detection and Verification
- Need for Automated Detection Systems

Research Plan

- **Objective**: Develop a browser extension that uses a transformer-based model (e.g., BERT) to detect and classify diabetes-related misinformation in text highlighted by users.

- **Scope**: The extension will categorize misinformation into four categories: exaggerated, misconstrued, false, and real.

- Data Collection and Preparation
- Model Selection and Training
- Prototype Development
- Testing and Evaluation
- Deployment and Maintenance
- Usability Study

Research Question (Misinformation Detection System)

- RQ1: How effective is the transformer-based model (e.g., BERT) in classifying diabetes-related misinformation into categories such as exaggerated, misconstrued, false, and real?
- RQ2: What are the strengths and limitations of the model when detecting misinformation specifically related to diabetes, compared to other health topics?
- RQ3: What are the challenges in training and fine-tuning the transformer-based model for misinformation detection, and how can these challenges be mitigated?

Research Questions (Usability Study)

- RQ1: How do users perceive the effectiveness of the browser extension in detecting and categorizing diabetes-related misinformation?
- RQ2: What are the user experience challenges associated with the browser extension, and how do they impact the overall usability and satisfaction?
- RQ3: How does the confidence scoring system influence user trust in the misinformation detection results provided by the extension?

Transformer Model Architecture

- The transformer model architecture has become a cornerstone of modern natural language processing (NLP) due to its ability to handle long-range dependencies and parallelize computations efficiently (Kumar et al., 2024, Jwa et al., 2019).

Encoder Components:

- Self-Attention Mechanism:
- Scaled Dot-Product Attention
- Multi-Head Attention:
- Feed-Forward Neural Network:
- Add & Norm

- **Decoder Components**
- Masked Self-Attention:
- Encoder-Decoder Attention
- Feed-Forward Neural Network
- Add & Norm

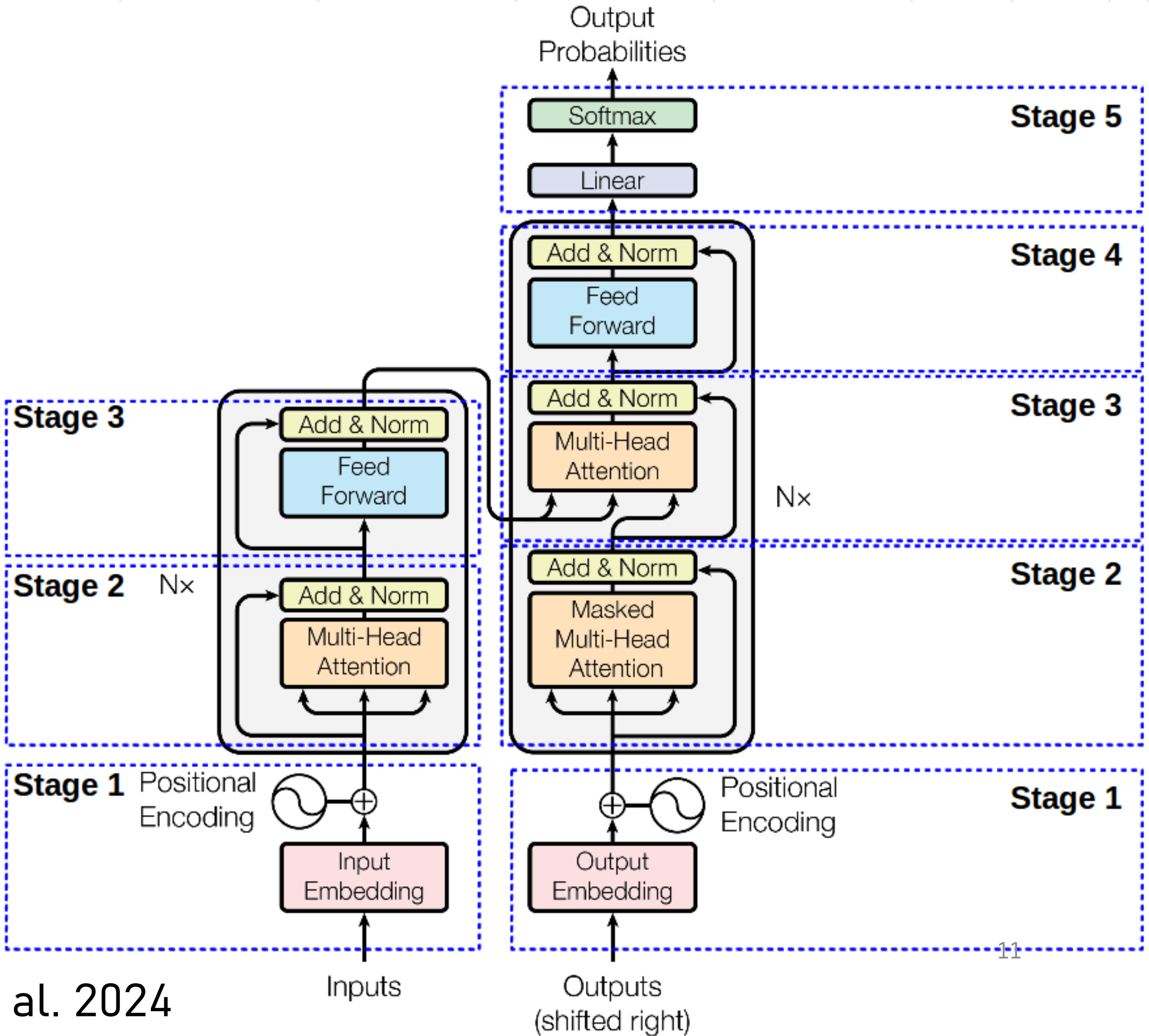


Image adapted from Kumar et al. 2024

Phase 1: Replication Study

Original Paper: Ghafourianaghahasanpour, M., Liu, Y., Yang, Z., and Zhang, E. H. COVID19 Fake News Detection using Transformer Based Model. In 2023 15th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC) (8 2023), IEEE, pp. 49–52.

- **Methodology:** Two-step approach
- **Language Model**: BERT-based with L=12, H=768, A=12 (12 layers, 768 hidden units, 12 attention heads).
- **Text Classifier**: Fine-tuned on COVID-19 misinformation detection.
- **Preprocessing**:
- Lowercase conversion, punctuation removal, stop word removal, number replacement.
- **Tokenization**: Added [CLS] and [SEP] tokens, max sequence length = 128.

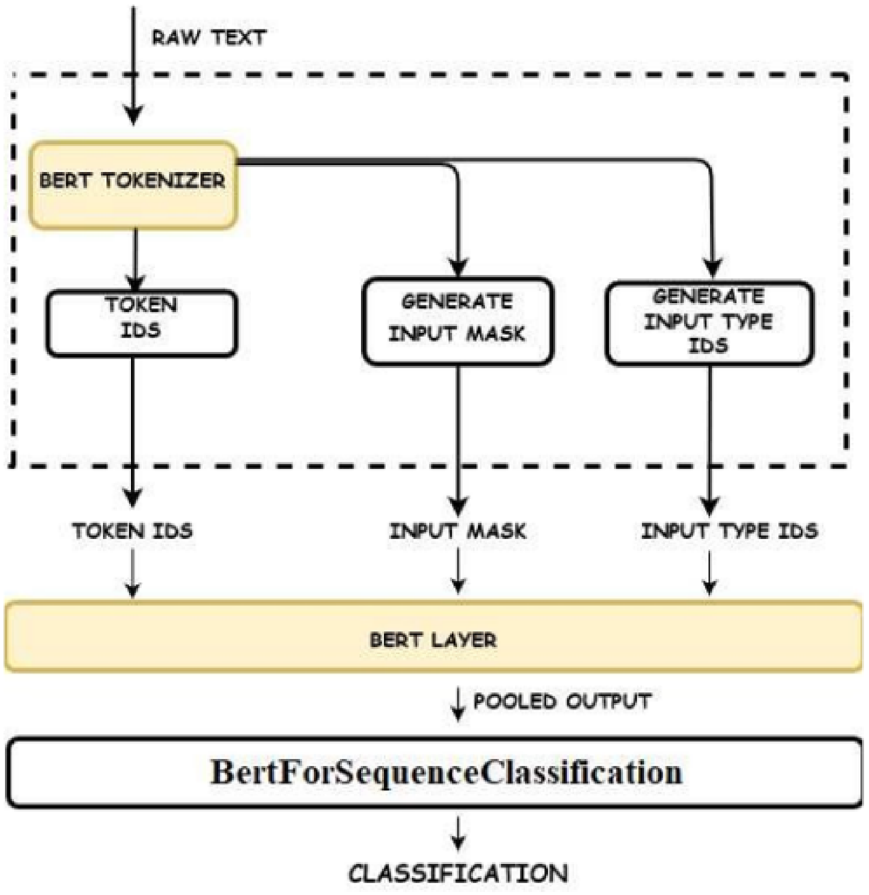


Image Adapted from Ghafourianaghahasanpour et al.(2023)

Pretraining, Fine-tuning & Evaluation

Pretraining:

MLM: Masked tokens prediction.

NSP: Predict sentence order.

Fine-tuning:

Dataset: COVID-19 news articles.

Parameters: 10 epochs, learning rate 2×10^{-5}, batch size 32.

Performance:

Accuracy: 95%, Precision: 0.91, Recall: 0.92, F1 score: 0.92.

Concerns: Data imbalance (98,215 real vs. 21,418 fake news points) affects reliability.

No link to the dataset

Replicated Study

- **Goal**: Validate and reproduce findings from the original study.
- **Frameworks** used:
- **TensorFlow**: As per the original paper, with Keras API.
- **PyTorch**: Added for experimentation and comparison.
- **ISOT NEWS DATASET**

News	Size	Subjects: Article Size
Real News	21,417	World news:10,145 Politics-news:11,272
Fake News	23,481	Government news: 1,570 Middle-east: 778 US news: 783 Left news: 4,459 Politics: 6,841 News: 9,050

TensorFlow Implementation

- **Tokenization & Preprocessing**:
  - Maximum sequence length = 128, using BERT input features (input IDs, masks, token type IDs).
  - Model initialized with **BERT layer** (L=12, H=768, A=12).
- **Model Architecture**:
  - Input layers, BERT layer, **Dropout layer** (0.1), **Sigmoid activation** (binary classification).
  - Optimizer: **Adam** (learning rate: 2e-5), loss function: **Binary Cross-Entropy**.
  - Epochs: 3, with early stopping after loss convergence.

Class Label	Precision	Recall	F1-Score
0 (Fake)	0.98	0.99	0.99
1(Not Fake)	0.99	0.98	0.99
Macro Average	0.98	0.99	0.99
Weighted Average	0.98	0.99	0.99

PyTorch Implementation

- **Model Setup**:
  - Tokenization and batching similar to TensorFlow setup.
  - BERT-based classification architecture with **ReLU activation** for complex pattern learning.
  - **Output layer**: Sigmoid activation (for binary classification).

- **Training**:
  - **Adam optimizer** (learning rate: 2e-5), binary cross-entropy loss.
  - Custom function for **backpropagation** to optimize model performance over 3 epochs.

	Precision	Recall	F1-Score
Class 0	0.95	0.85	0.92
Class 1	0.89	0.95	0.92
Average	0.92	0.92	092

Class Imbalance on Model Performance

Dataset Skew:

▪Replication study used **17.86% fake news** and **82.14% real news** (as in original paper).

Replicated Study (Imbalanced Dataset)

	Precision (%)	Recall (%)	F1 Score(%)
Class 0 (Fake News)	78	89	83
Class 1 (Real News)	97	94	96
Average Metrics	88	91	89

Replicated Study (Balanced Dataset)

	Precision (%)	Recall (%)	F1 Score (%)
Class 0	95	85	92
Class 1	89	95	92
Average Metrics	92	92	92

Findings from the Study

- Both TensorFlow and PyTorch frameworks showed strong performance (high precision and recall).
- **Class Imbalance Impact:** Imbalanced dataset led to lower performance, especially in detecting fake news.
- **Comparison to Original Study:**
- Differences in loss rates, architecture, and hyperparameters affected performance.
- Hyperparameter tuning and class imbalance handling played a significant role.
- The results are consistent with that of the original paper, there is a need to explore methodologies to mitigate the effect of imbalance in data.

Phase 2: Study on post content vs user profile information for misinformation detection

Methodology

- **BiLSTM for Textual Data**: Chosen for its ability to capture past and future context in text, making it highly effective for text classification tasks (Siami-Namini et el., 2019)
- **Machine Learning Models for User Profile Features**:
- Random Forest (RF)
- Gradient Boosting Machine (GBM)
- Logistic Regression (LR)
- Support Vector Machine (SVM)
- **Rationale**: BiLSTM is less suited for non-textual, independent features like user profiles, hence using traditional machine learning models.

Methodology: Stages of Analysis

Stage 1: User Profile Analysis: Focus on user features (e.g., follower count, friends count) to detect fake news spreaders.

Stage 2: Textual Data Analysis: Preprocessing tweets and applying the BiLSTM model to detect linguistic patterns.

Stage 3: Combined Approach: Integration of both user profile and textual data for a more comprehensive fake news detection model.

Dataset and Model Architecture

- **Dataset**: Truthseeker dataset with over 180,000 tweets (51% real, 49% fake), includes both textual and user profile features.
- **BiLSTM Architecture**:
- Combines forward and backward LSTMs for sequence processing.
- Effective for textual features in binary classification.
- **Machine Learning Models**:
- Used for numerical, non-sequential user profile data (e.g., follower and friends count).

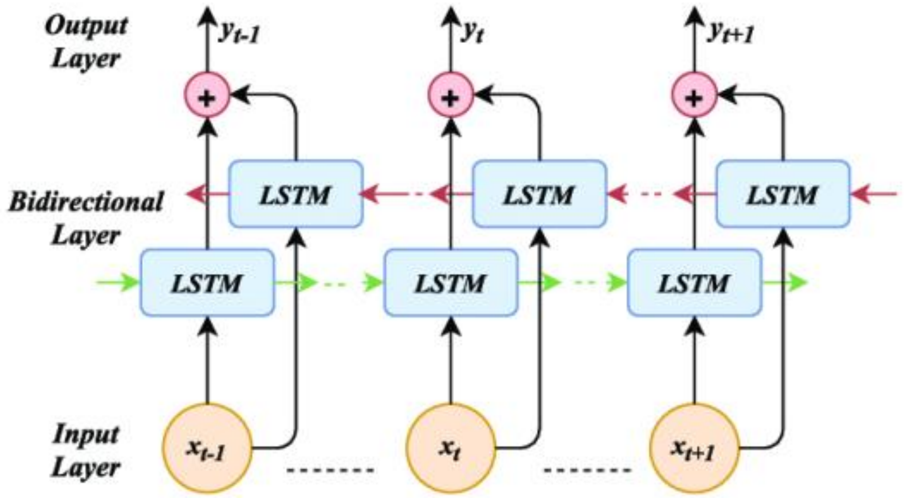


Image adapted from Pytorch Forum

Results from Phase 2

- **Text Features Only: Test Accuracy**: 97.88%

	Precision	Recall	F1- Score
Fake News	0.97	0.98	0.98
Real News	0.99	0.97	0.98

- **User Profile Features**

Model	Accuracy	Precision	Recall	F1-Score
Random Forest (RF)	0.58	0.56	0.57	0.57
Logistic Regression (RF)	0.55	0.54	0.53	0.55
Support Vector Machine (SVM)	0.59	0.59	0.60	0.60
Gradient Boost Machine (GBM)	0.56	0.56	0.55	0.57

Results from Phase 2

Combined Features (Text + User Profile)

- Accuracy: 97.61%
- **Combined Features (Text + User Profile)**: Accuracy slightly dropped to **97.61%**, showing that combining user profile data with text does not significantly enhance performance.

	Precision	Recall	F1- Score
Fake News	0.97	0.98	0.98
Real News	0.98	0.97	0.98
Macro Avg	0,98	0.98	

- **Model Performance Insights**: Textual features alone are highly effective for detecting fake news, while user profile data provides marginal improvements and may introduce noise without proper optimization.

Systematic Literature Review

- Fake News Spreaders Detection using Transformer Model Based Approach: A Systematic Literature Review

Research Questions

- RQ1: What transformer-based architectures are most commonly used for fake news detection in academic research?
- RQ2: How effective are transformer models compared to other deep learning approaches in detecting fake news?
- RQ3: What datasets and benchmarks are most frequently used in transformer-based fake news detection studies?
- RQ4: What evaluation metrics are employed to assess the performance of these models?

Systematic Literature Review

Methodology

- **Searching**: Used comprehensive search strategies with keywords like "transformer models," "fake news detection," and "misinformation" across four academic databases (SpringerLink, PubMed, IEEE Xplore, ACM) to retrieve studies published between 2018 and 2024.
- **Screening & Data Extraction**: Applied a two-step screening process (titles/abstracts and full texts) based on predefined inclusion/exclusion criteria.
- Extracted data on models, datasets, content types (e.g., political or health misinformation), and evaluation metrics (e.g., accuracy, precision).
- **Synthesis**: Synthesized data through comparative analysis, identifying trends and gaps in transformer models' performance for fake news detection.

Results so far from the Systematic Literature Review

Text-Based Transformers	Image and Multimodal Transformers
BERT (Bidirectional Encoder Representations from Transformers)	CLIP (Contrastive Language–Image Pre-training)
RoBERTa (Robustly Optimized BERT Pretraining Approach)	Swin Transformer
DistilBERT	VILT (Vision-and-Language Transformer)
ALBERT (A Lite BERT)	DALL-E
XLNet	CoAtNet
GPT-3 (Generative Pre-trained Transformer 3)	VideoBERT
T5 (Text-to-Text Transfer Transformer)	Perceiver
ELECTRA	ImageBERT
DeBERTa (Decoding-enhanced BERT with disentangled attention)	VTN (Video Transformer Network)
UniLM (Unified Language Model)	TimeSformer

Publication

Submitted Papers

- From Tweets to Truths: Leveraging BERT for Improved Text Classification: 23rd International Conference on Machine Learning and Applications
- Detecting Fake News Spreaders on Social Media Using Posts Content Versus Profile Information: 2024 IEEE International Conference on Big Data

Summary

Replication Study
Key Insight: BERT is effective for misinformation detection.

Text vs Profile Data
Key Insight: Textual data performs better than profile data for misinformation detection.

Model Selection and Training

Prototype Development
Testing and Evaluation

Deployment and Maintenance
Usability Study

References

- Aïmeur, E., Amri, S., and Brassard, G. Fake news, disinformation and misinformation in social media: a review. Social Network Analysis and Mining 13, 1 (12 2023)
- Ahinkorah, B. O., Ameyaw, E. K., Hagan, J. E., Seidu, A., & Schack, T. (2020). Rising Above Misinformation or Fake News in Africa: Another Strategy to Control COVID-19 Spread. *Frontiers in Communication*, *5*, 549910. https://doi.org/10.3389/fcomm.2020.00045
- Kumar, Deepak & Singh, Shoumya. (2024). ADVANCEMENTS IN TRANSFORMER ARCHITECTURES FOR LARGE LANGUAGE MODEL: FROM BERT TO GPT-3 AND BEYOND. International Research Journal of Modernization in Engineering Technology and Science. 06. 2582-5208. 10.56726/IRJMETS55985.
- Ghafourianaghahasanpour, M., Liu, Y., Yang, Z., and Zhang, E. H. COVID19 Fake News Detection using Transformer Based Model. In 2023 15th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC) (8 2023), IEEE, pp. 49–52. [11] Gundapu, S., and Mamidi, R. Transformer based Automatic COVID-19 Fake News Detection System.
- Jwa, H., Oh, D., Park, K., Kang, J. M., and Lim, H. exBAKE: Automatic fake news detection model based on Bidirectional Encoder Representations from Transformers (BERT). Applied Sciences (Switzerland) 9, 19 (10 2019).
- Siami-Namini, N. Tavakoli, and A. S. Namin, “The performance of LSTM and BiLSTM in forecasting time series,” in 2019 IEEE International Conference on Big Data (Big Data), IEEE, 2019, pp. 3285–3292. doi: https://doi.org/10.1109/BigData47090.2019.9005997.

LINDA O. OKPANACHI

PASCL LAB

Department of Computer Science, Mathematics and Statistics, University of British Columbia, Okanagan Campus

Email: lindaa15@student.ubc.ca

LinkedIn: https://www.linkedin.com/in/linda-okpanachi-12327710b/

THANK YOU!!! QUESTIONS??