ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) | 978-1-7281-6327-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICASSP49357.2023.10095215

PQLM - MULTILINGUAL DECENTRALIZED PORTABLE QUANTUM LANGUAGE MODEL

Shuyue Stella Li[⋆1] *Xiangyu Zhang*[⋆1] *Shu Zhou*[3] *Hongchao Shu*[1]
Ruixing Liang[1] *Hexin Liu*[4] *Leibny Paola Garcia*[1,2]

[1]Center for Language and Speech Processing, Johns Hopkins University
[2]Human Language Technology Center of Excellence, Johns Hopkins University
[3]Department of Physics, Hong Kong University of Science and Technology
[4]School of Electrical and Electronic Engineering, Nanyang Technological University

ABSTRACT

With careful manipulation, malicious agents can reverse engineer private information encoded in pre-trained language models. Security concerns motivate the development of quantum pre-training. In this work, we propose a highly portable quantum language model (PQLM) that can easily transmit information to downstream tasks on classical machines. The framework consists of a cloud PQLM built with random Variational Quantum Classifiers (VQC) and local models for downstream applications. We demonstrate the ad hoc portability of the quantum model by extracting *only* the word embeddings and effectively applying them to downstream tasks on classical machines. Our PQLM exhibits comparable performance to its classical counterpart on both intrinsic evaluation (loss, perplexity) and extrinsic evaluation (multilingual sentiment analysis accuracy) metrics. We also perform ablation studies on the factors affecting PQLM performance to analyze model stability. Our work establishes a theoretical foundation for a portable quantum pre-trained language model that could be trained on private data and made available for public use with privacy protection guarantees.

Index Terms— Quantum Machine Learning, Language Modeling, Federated Learning, Model Portability

1. INTRODUCTION

A competitive language model can be extremely useful for downstream tasks such as machine translation and speech recognition despite the domain mismatch between pre-training and downstream tasks [1, 2]. They become more powerful with increased training data, but there is a trade-off between data privacy and utility[3]. Previous works on ethical AI have shown that pre-trained language models (PLM)s memorizes training data in addition to learning about the language [4, 5], which opens up vulnerabilities for potential adversaries to recover sensitive training data from the model.

While some argue that language models should only be trained on data *explicitly* produced for public use [6], private data are richer in certain domains compared to public corpora, including dialogue systems, code-mixing languages, and medical applications [7, 8]. Therefore, it is essential to develop new methods to mitigate potential data security and privacy problems while being able to take advantage of the rich linguistic information encoded in private data.

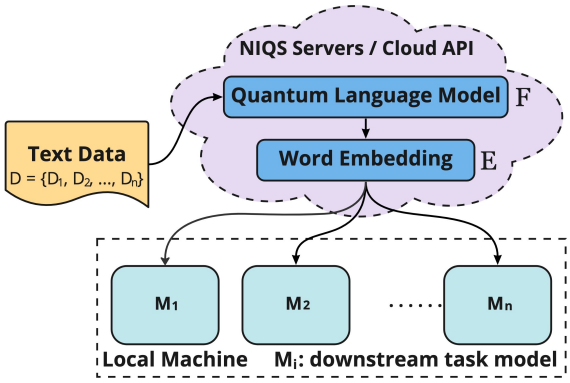


Fig. 1: Decentralized Quantum Language Model Pipeline. Text data is trained on language model on NISQ servers, the word embeddings are transferred to downstream models $\mathcal{M}_i$

Recently, there has been growing interest in leveraging random quantum circuits in neural models to solve data privacy issues [9]. The entanglement of states from the random configuration of gates in the quantum circuits makes it possible to securely encode sensitive information contained in training data [10, 9]. The combination of random quantum circuits and decentralized training ensures privacy [11]. Additionally, quantum computing has become the next logical step in the development of deep learning for its efficiency in manipulating large tensors [12].

The architecture of a large quantum computer is vastly different from the classical computer as physical and environmental constraints cannot be met [13], which means that the model trained on the large quantum computer is difficult to be directly used by others on the classical computer. Ad hoc portability is defined as the model's ability to transmit the most essential information contained in the language model

⋆Equal contribution in alphabetical order

across different machines. In this work, we provide a method to seamlessly transmit the information learned from the quantum training step to the classical machine without requiring any additional model adaptations, making the quantum model highly portable.

As shown in Figure 1, we propose a decentralized PQLM pipeline where private data is fed into a quantum model composed of a Recurrent Neural Network (RNN) with its gates replaced with Variational Quantum Classifiers (VQC)[14]. As we will describe more in detail in Section 2, Noisy intermediate-scale quantum (NISQ) computers on quantum servers are used for such variational quantum algorithm computations [15]. After the PQLM hosted on remote NISQ is trained to convergence, instead of downloading the entire model to a classical system, we directly use the embeddings trained by the PQLM to initialize downstream tasks.
In summary, our contributions include the following:

- We propose a decentralized pipeline for transmitting knowledge learned from a secure, fast quantum pre-trained model to classical machines for downstream tasks.
- We demonstrate the ad hoc portability of our pre-trained language model by showing that extracting the embeddings trained by the PQLM is sufficient for downstream tasks such as setiment analysis (SA).
- We show the stability of the model across different orders of complexity with ablation studies on factors including the number of qubits and training corpus size.

2. RELATED WORK

Differential Privacy and Federated Learning Traditional approaches to protecting sensitive data such as Differential Privacy (DP) introduces random noise to the system to protect individual data change, but usually sacrifices model performance and efficiency [3] and makes strong assumptions on the training data [6]. More recent work towards a safe data pipeline involves decentralized training for federated training, where the training data is strictly kept to remote machines, and the gradients are exported and aggregated on another machine for downstream tasks [16, 17, 18]. This resolves the need for any private data to join a centralized data pool from the root. Even then, it is possible to leak sensitive information [19].

Quantum ML with Variational Quantum Circuits Some recent works attempt to make use of the irrecoverability of quantum circuits in differential privacy algorithms [9] or federated learning architectures [20] involving the use of VQCs on NISQ clusters with reliable optimization [21]. VQCs are quantum circuits with quantum parameters that can absorb the noise inherently contained in quantum computations and can be optimized iteratively with classical gradient descent [22, 23]. As shown in Figure 2a, there are three parts in the VQC architecture: (1) the encoding stage where the input vector is encoded; (2) the quantum circuit stage where entanglement strategies are applied with quantum gates and parameters are stored and trained; and (3) the measurement stage where a hermitian operator projects the quantum states onto its eigenvectors [14].

For decentralized training in which the data and the quantum portion are hosted on a NISQ server, any adversaries will not be able to recover the structure or data without knowing the quantum gate configurations in the random circuit, therefore providing a security guarantee [24, 11, 10, 15]. More secure procedures include the Quantum Circuit Obfuscation method that add dummy CNOT gates to the circuit so the data is protected from both the quantum and local machines, yet at the cost of additional computation [25].

Decentralized Quantum Learning Applications Some previous work has looked into the area of sequential data modeling [22] and natural language processing [26], but with a focus on the speedup of quantum computations. Applications of quantum decentralized privacy protection approach have been used for speech feature extraction [11], image recognition [27], language processing [28], and reinforcement learning [29]. In this paper, we take a single-party delegated training approach that can be easily extended to multi-machine decentralized learning to train a secure PQLM. We focus on providing a portable transfer of information from the quantum server to the classical side.

3. METHODOLOGY

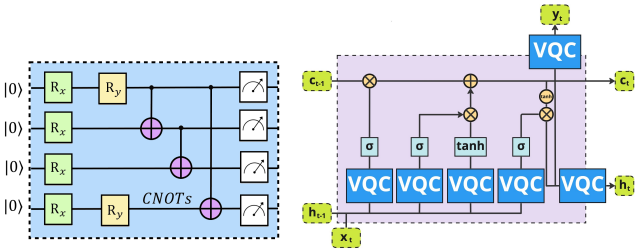
3.1. Quantum-LSTM Language Model

The primary task of language modeling is that given sequence words $x^1, x^2, ..., x^t$, we need to predict the probability $P(x^{t+1} \mid x^1, x^2, ...x^t)$. In previous studies, Long-Short Term Memory (LSTM) has shown to be a good deep learning model for building language models [30]. In our studies, We use Quantum LSTM (Q-LSTM) [22] which is based on LSTM model and random VQC to train our language model. The basic architecture of Q-LSTM is shown in Figure 2b. Q-LSTM replaces some components such as forget gate, input gate, update gate, and output gate in classical LSTM with VQC and uses the mechanism of backpropagation to update parameters of the Q-LSTM model. To meet the coherence time specification, our Q-LSTM language model is built with a shallow circuit of 2 layers and each VQC gate is built with 4 qubits. Both the classical and Q-LSTM have a word embedding size of 64 and vocab size as the output size.

As shown in Figure 2a, we use a random circuit to encode the rotational vectors to protect the parameters against any 3rd party attacks in the entanglement stage.

3.2. Decentralized Training

Quantum computers are in theory exponentially faster than classical computers, making them ideal for training large-scale models. Therefore, decentralized training on a quantum



(a) Quantum Circuit [28] (b) Q-LSTM

Fig. 2: Model Architecture

machine will provide both security and speedup. However, because the quantum model learns in the Hilbert space represented by qubits, it will be difficult to load the entire model into classical machines for downstream tasks the same way we do for current pre-trained language models like BERT. We introduce a decentralized PQLM framework as shown in Figure 1, which simplifies the transmission of information from the quantum side to the classical side using extracted word embeddings. Given decentralized PQLM F, we input a set of text documents $D_1, D_2, D_3..., D_n$ into the NISQ servers. After training, the word embedding E will be extracted from PQLM and using in local downstream task.

$$E = F(D_1, D_2, D_3..., D_n) \quad (1)$$

In Equation 1, the PQLM F can be any quantum deep learning model that can be used to train the language model. In this paper, we present Q-LSTM as an example to train our language model. The output word embeddings E is a set of vectors that can be further processed by other classical or quantum routines. Finally, we make our work open source for future explorations [1].

4. EXPERIMENTS

4.1. Dataset

We use two distinct datasets to train two language models. The first one is a multilingual Twitter dataset [2] consisting of 69491 training documents and 998 test documents with 4 different labels (negative, positive, neutral and irrelevant). The second one is an English-Hindi code-mixed Twitter dataset from SemEval-2020 Task 9 [31][3], which consists of 15000 training and validation documents and 3,789 test documents. These two datasets are selected because they are linguistically complex in nature, and they have gold labels for SA, which we later use in our evaluation step to train a classification model without introducing further supervision or noise.

4.2. Preprocessing

A coarse filtering of the training datasets is performed before they are fed into the language model. First, empty strings, hash symbols, and URLs are removed from the text. All emojis and emoticons are replaced by their English descriptions using the emoji library[4]. Sentiment labels are ignored during language model training.

4.3. Q-LSTM LM vs. Classical LSTM LM

In order to fairly assess the performance of the PQLM, we create a classical language model with the same model architecture and size. With a hidden size of 5 in the classical LSTM, the number of parameters is of the same magnitude as the Q-LSTM with 4 qubits [22].

The models are trained until the loss converges. The negative log-likelihood loss during training and the model perplexity are recorded to evaluate the model. Figure 3a shows the training loss over 15 epochs of the Q-LSTM and its classical counterpart. The PQLM converges much faster than the classical language model of the same size.

Model	LSTM	Q-LSTM-4q	Q-LSTM-6q
perplexity	1152.78	1153.67	972.44

Table 1: Model Perplexity on Multilingual Twitter Corpus LSTM: classical LSTM; Q-LSTM-4q: Q-LSTM built with 4-qubit VQC; Q-LSTM-6q: Q-LSTM built with 6-qubit VQC.

Despite the quantum model converging faster than the classical language model, the model quality as measured by model perplexity for both models is extremely close as shown in Table 1. Given the computing constraints, we only used 4 qubits to train the LSTM model, resulting in an extremely limited number of parameters (around 200) for both models. Although the model perplexity is not ideal compared to conventional PLMs, it still provides a valuable basis for comparing the classical language model and the PQLM. Another factor that contributes to the high perplexity of our language model is the multilinguality of the training data, which contains richer linguistic information, making it difficult for the small model to learn.

4.4. Model Evaluation - Sentiment Analysis

We use a downstream SA task to compare the quality of the PQLM with its classical counterpart. Given a sequence of tweets $D_1, D_2...D_n$, we need to predict the sentiment class (positive, neutral, negative, or irrelevant). We train a local transformer-based four-way classifier using the pre-trained word embeddings from the Q-LSTM language model. We use 4 transformer blocks and each transformer model has 4 attention heads. In order to avoid introducing more noise to our model pipeline, we use a random subset of the training data for the language model. Finally, the sentiment of an unknown document can be predicted as

$$y^i = softmax(W^i f(x) + b^i) \quad (2)$$

Where y^i is a predicted sentiment label for a test document, f is a transformer function to encode the input text x. We report both the accuracy and weighted f1 scores since the dataset is distributed unevenly across the four labels,

[1] git@github.com:stellali7/quantunLM.git

[2] https://www.kaggle.com/datasets/jp797498e/twitter-entity-sentiment-analysis

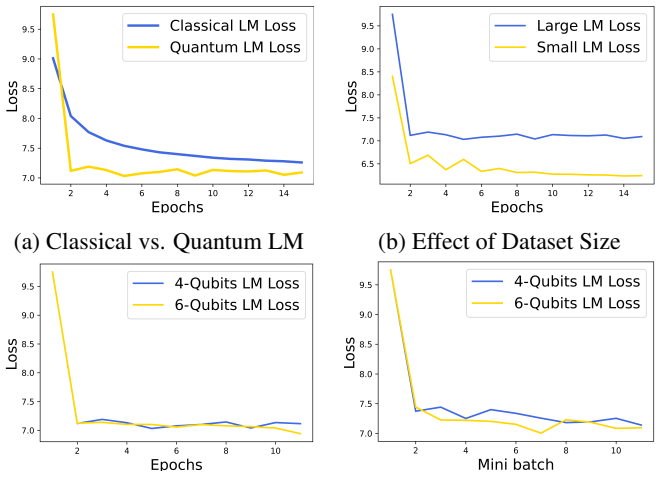
[3] https://ritual-uh.github.io/sentimix2020/

[4] https://pypi.org/project/emoji/

PLM	LSTM	Q-LSTM (4q)
accuracy	0.928	0.934
weighted f1	0.93	0.93

Table 2: SA Performance on Multilingual Twitter Dataset

Table 2 summarizes the four-way sentiment classification performance of the local transformer-based classifier initialized with embeddings trained by the classical LSTM and the quantum LSTM, while keeping all other hyperparameters the same. The embedding trained by Q-LSTM achieves slightly higher accuracy than the classical LSTM, and has the same weighted f1 score. Our results confirm that the decentralized PQLM training and the portable information transmission do not sacrifice performance. With the random circuits in its VQC gates, the quantum model has higher expressibility - a circuit's ability to generate states in the Hilbert space [32]. High expressibility allows the model to better search the solution space given the training data compared to the classical model, despite having the same number of parameters. The high expressibility of the quantum model makes it a "better learner," contributing to the slightly higher accuracy.



(a) Classical vs. Quantum LM

(b) Effect of Dataset Size

(c) Effect of # qubits (12 epoch)

(d) Effect of # qubits (1 epoch)

Fig. 3: Training Loss for Different Experiments

4.5. Ablation Studies

4.5.1. Effect of Number of Qubits

A single qubit can be described by a two-dimensional Hilbert space $\mathcal{H}$. Then, a system of n qubits is the tensor product of n such Hilbert spaces

$$\underbrace{\mathcal{H} \otimes \mathcal{H} \otimes \cdots \otimes \mathcal{H}}_{n}. \tag{3}$$

We explore the effect of having more qubits in the VQC gates to examine the embeddings trained on the higher dimensional Hilbert space. Both the 4-Qubit model and 6-Qubit model converge rapidly as shown in Figure 3c, so we provide a fine-grained comparison of the training loss for each batch in the first epoch in Figure 3d. The 6-Qubit language model has a much lower perplexity (Table 1) and a faster convergence within the first epoch (Figure 3d). This indicates that the 6-Qubit model better captures the non-linear relationship in a higher dimensional embedding space to better fit the data. As the number of qubits increases, so does the noise introduced to the system. However, the smooth loss curve demonstrates model stability as we extend the model to more qubits, indicating the model's robustness against quantum noise.

4.5.2. Effect of Training Data Size

Dataset	Multilingual Twitter	Code-Mixing
Data size	69491	15000
Vocab size	17000	5000
perplexity	1153.672	368.031

Table 3: Model Perplexity for Different Data Sizes

We train the Q-LSTM on a smaller corpus to confirm that the high perplexity on both the classical and the quantum model is due to the small model size trained on a large dataset. The small dataset has a much lower model perplexity (Table 3) and a lower training loss (Figure 3b). Due to computation constraints imposed by the NISQ machines, our model capacity is limited, so the results indicate parameter saturation but also justify the large perplexity. Despite the difference in the corpus size, there is no obvious difference in the convergence speed of the two models. Furthermore, although the code-mixing corpus is more linguistically complex and thus harder to predict, the linguistic complexity still does not overpower the effect of dataset size on the model loss and perplexity. Therefore, this experiment demonstrates model stability and justifies the high perplexity due to limited model size.

5. CONCLUSIONS

This work provides a solution to integrate secure and fast quantum language modeling and flexible classical downstream applications. We proposed a decentralized training framework for PQLMs that provides both parameter protection and ad hoc portability. We show that the embedding extraction is sufficient for downstream tasks, which enables the ad hoc transfer of the model from the quantum server to the classical machine. Our PQLM achieves competitive if not better results compared to its classical counterpart on multilingual Twitter SA tasks. We also studied the effect of the number of qubits in the VQC and the training corpus size to demonstrate model stability.

Our work provides a promising direction and a theoretical foundation for future NLP research that have privacy protection demands. Some of the future work include extending this decentralized quantum training procedure into a wider range of language model architectures. As the required quantum computing hardware becomes available to train large-scale PQLMs, our approach can be used to easily transmit the quantum information to downstream classical tasks. The ad hoc portability method can also be further studied to enable downstream fine-tuning and model compression.

6. REFERENCES

[1] Rishi Bommasani, Drew A Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S Bernstein, Jeannette Bohg, Antoine Bosselut, Emma Brunskill, et al., "On the opportunities and risks of foundation models," *CoRR*, vol. abs/2108.07258, 2021.

[2] Shaohua Wu, Xudong Zhao, Tong Yu, Rongguo Zhang, Chong Shen, Hongli Liu, Feng Li, Hong Zhu, Jiangang Luo, Liang Xu, and Xuanwei Zhang, "Yuan 1.0: Large-scale pre-trained language model in zero-shot and few-shot learning," *CoRR*, vol. abs/2110.04725, 2021.

[3] Weiyan Shi, Aiqi Cui, Evan Li, Ruoxi Jia, and Zhou Yu, "Selective differential privacy for language modeling," in *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2022, pp. 2848–2859.

[4] Nicholas Carlini, Chang Liu, Úlfar Erlingsson, Jernej Kos, and Dawn Song, "The secret sharer: Evaluating and testing unintended memorization in neural networks," in *28th USENIX Security Symposium (USENIX Security 19)*, Santa Clara, CA, Aug. 2019, pp. 267–284, USENIX Association.

[5] Nicholas Carlini, Florian Tramèr, Eric Wallace, Matthew Jagielski, Ariel Herbert-Voss, Katherine Lee, Adam Roberts, Tom Brown, Dawn Song, Úlfar Erlingsson, Alina Oprea, and Colin Raffel, "Extracting training data from large language models," in *30th USENIX Security Symposium (USENIX Security 21)*. Aug. 2021, pp. 2633–2650, USENIX Association.

[6] Hannah Brown, Katherine Lee, Fatemehsadat Mireshghallah, Reza Shokri, and Florian Tramèr, "What does it mean for a language model to preserve privacy?," *2022 ACM Conference on Fairness, Accountability, and Transparency*, pp. 2280–2292, 2022.

[7] Munazza Zaib, Quan Z. Sheng, and Wei Emma Zhang, "A short survey of pre-trained language models for conversational ai-a new age in nlp," in *Proceedings of the Australasian Computer Science Week Multiconference*, New York, NY, USA, 2020, ACSW '20, Association for Computing Machinery.

[8] Thomas Reitmaier, Electra Wallington, Dani Kalarikalayil Raju, Ondrej Klejch, Jennifer Pearson, Matt Jones, Peter Bell, and Simon Robinson, "Opportunities and challenges of automatic speech recognition systems for low-resource language speakers," in *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, New York, NY, USA, 2022, CHI '22, Association for Computing Machinery.

[9] William M Watkins, Samuel Yen-Chi Chen, and Shinjae Yoo, "Quantum machine learning with differential privacy," *Scientific Reports*, vol. 13, no. 1, pp. 2453, 2023.

[10] Daniel Shaffer, Claudio Chamon, Alioscia Hamma, and Eduardo R Mucciolo, "Irreversibility and entanglement spectrum statistics in quantum circuits," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2014, no. 12, pp. P12007, dec 2014.

[11] Chao-Han Huck Yang, Jun Qi, Samuel Yen-Chi Chen, Pin-Yu Chen, Sabato Marco Siniscalchi, Xiaoli Ma, and Chin-Hui Lee, "Decentralizing feature extraction with quantum convolutional neural network for automatic speech recognition," in *Proc. ICASSP 2021*. IEEE, 2021, pp. 6523–6527.

[12] Seth Lloyd, Masoud Mohseni, and Patrick Rebentrost, "Quantum algorithms for supervised and unsupervised machine learning," *arXiv preprint arXiv:1307.0411*, 2013.

[13] John M Martinis, "Quantum supremacy in a superconducting quantum processor," in *Photonics for Quantum 2021*. SPIE, 2021, vol. 11844, p. 118440D.

[14] Israel Griol-Barres, Sergio Milla, Antonio Cebrián, Yashar Mansoori, and José Millet, "Variational quantum circuits for machine learning. an application for the detection of weak signals," *Applied Sciences*, vol. 11, no. 14, 2021.

[15] John Preskill, "Quantum Computing in the NISQ era and beyond," *Quantum*, vol. 2, pp. 79, Aug. 2018.

[16] Priyanka Mary Mammen, "Federated learning: Opportunities and challenges," *CoRR*, vol. abs/2101.05428, 2021.

[17] Tian Li, Anit Kumar Sahu, Ameet Talwalkar, and Virginia Smith, "Federated learning: Challenges, methods, and future directions," *IEEE Signal Processing Magazine*, vol. 37, no. 3, pp. 50–60, 2020.

[18] Maoguo Gong, Yu Xie, Ke Pan, Kaiyuan Feng, and A.K. Qin, "A survey on differentially private machine learning [review article]," *IEEE Computational Intelligence Magazine*, vol. 15, no. 2, pp. 49–64, 2020.

[19] Keith Bonawitz, Vladimir Ivanov, Ben Kreuter, Antonio Marcedone, H. Brendan McMahan, Sarvar Patel, Daniel Ramage, Aaron Segal, and Karn Seth, "Practical secure aggregation for privacy-preserving machine learning," in *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, New York, NY, USA, 2017, CCS '17, p. 1175–1191, Association for Computing Machinery.

[20] Samuel Yen-Chi Chen and Shinjae Yoo, "Federated quantum machine learning," *Entropy*, vol. 23, no. 4, 2021.

[21] Marco Cerezo, Andrew Arrasmith, Ryan Babbush, Simon C Benjamin, Suguru Endo, Keisuke Fujii, Jarrod R McClean, Kosuke Mitarai, Xiao Yuan, Lukasz Cincio, et al., "Variational quantum algorithms," *Nature Reviews Physics*, vol. 3, no. 9, pp. 625–644, 2021.

[22] Samuel Yen-Chi Chen, Shinjae Yoo, and Yao-Lung L. Fang, "Quantum long short-term memory," in *Proc. ICASSP 2022*, 2022, pp. 8622–8626.

[23] Leo Zhou, Sheng-Tao Wang, Soonwon Choi, Hannes Pichler, and Mikhail D. Lukin, "Quantum approximate optimization algorithm: Performance, mechanism, and implementation on near-term devices," *Phys. Rev. X*, vol. 10, pp. 021067, Jun 2020.

[24] Chuntang Li, Yinsong Xu, Jiahao Tang, and Wenjie Liu, "Quantum blockchain: a decentralized, encrypted and distributed database based on quantum mechanics," *Journal of Quantum Computing*, vol. 1, no. 2, pp. 49, 2019.

[25] Aakarshitha Suresh, Abdullah Ash Saki, Mahabubul Alam, Dr Ghosh, et al., "A quantum circuit obfuscation methodology for security and privacy," *arXiv preprint arXiv:2104.05943*, 2021.

[26] Ivano Basile and Fabio Tamburini, "Towards quantum language models," in *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, Copenhagen, Denmark, Sept. 2017, pp. 1840–1849, Association for Computational Linguistics.

[27] Jun Qi, "Federated quantum natural gradient descent for quantum federated learning," *arXiv preprint arXiv:2209.00564*, 2022.

[28] Chao-Han Huck Yang, Jun Qi, Samuel Yen-Chi Chen, Yu Tsao, and Pin-Yu Chen, "When bert meets quantum temporal convolution learning for text classification in heterogeneous computing," in *Proc. ICASSP 2022*, 2022, pp. 8602–8606.

[29] Samuel Yen-Chi Chen, Chao-Han Huck Yang, Jun Qi, Pin-Yu Chen, Xiaoli Ma, and Hsi-Sheng Goan, "Variational quantum circuits for deep reinforcement learning," *IEEE Access*, vol. 8, pp. 141007–141024, 2020.

[30] Wojciech Zaremba, Ilya Sutskever, and Oriol Vinyals, "Recurrent neural network regularization," *CoRR*, vol. abs/1409.2329, 2014.

[31] Parth Patwa, Gustavo Aguilar, Sudipta Kar, Suraj Pandey, Srinivas PYKL, Björn Gambäck, Tanmoy Chakraborty, Thamar Solorio, and Amitava Das, "SemEval-2020 task 9: Overview of sentiment analysis of code-mixed tweets," in *Proceedings of the Fourteenth Workshop on Semantic Evaluation*, Barcelona (online), Dec. 2020, pp. 774–790, International Committee for Computational Linguistics.

[32] Sukin Sim, Peter D. Johnson, and Alán Aspuru-Guzik, "Expressibility and entangling capability of parameterized quantum circuits for hybrid quantum-classical algorithms," *Advanced Quantum Technologies*, vol. 2, no. 12, pp. 1900070, 2019.