Quantum-Based Language Models: A Survey of Principles and Advances

TNSA Research

November 6, 2025

1 Quantum vs Classical Computing

Classical computers process information in binary bits (0 or 1), whereas quantum computers use **qubits**, which can exist in superpositions of 0 and 1 simultaneously [1, 2]. A single qubit can be in a state $\alpha|0\rangle + \beta|1\rangle$ (with complex amplitudes α, β) and multiple qubits can become **entangled**, correlating their states in non-classical ways [2]. These quantum features give quantum systems an exponentially large state space: for example, three qubits together can represent 8 classical states at once [3]. In contrast, n classical bits represent only one of 2^n states at a time. Superposition and entanglement thus enable **quantum parallelism** and expressive power beyond classical systems [2].

Quantum circuits implement computations by applying sequences of quantum gates to qubits [2]. Gates like the Hadamard put qubits into equal superpositions, and two-qubit gates like CNOT generate entanglement. A full quantum algorithm is built from such gates on a qubit register; however, current NISQ (noisy intermediate-scale quantum) hardware can only execute relatively shallow circuits due to noise. For example, building a 3-qubit Toffoli gate requires many primitive gates (5 CNOTs and 9 single-qubit gates), so even a few dozen gates can accumulate large error. As TechTarget notes, "in theory, quantum computers will have more compute power and be more scalable . . . because of qubits' unique properties" [1].

2 Fundamental Principles of Quantum Language Models

Quantum language models (QLMs) apply these quantum principles to text processing. In essence, they seek to encode linguistic data as quantum states and manipulate them via quantum circuits, exploiting amplitudes and entanglement to capture relationships between words.

A key idea is that words or phrases can be represented by quantum states or tensors in a high-dimensional Hilbert space [4, 5]. Some quantum-inspired models treat each word as a normalized vector or density operator and apply unitary transformations to model semantic composition. For example, QLM-UT uses a trainable unitary matrix to encode the *dynamic evolution* of sentence meaning, incorporating word-order via real-valued embeddings [4].

Beyond quantum-inspired models, **fully quantum** formulations also exist. In *circuit-based QNLP*, a sentence's grammatical structure is mapped to a quantum circuit [6, 7]. The DisCoCat and DisCoCirc frameworks parse sentences into diagrams that translate into parameterized quantum circuits. In such models, word meanings compose like quantum systems. Open-source tools like `lambeq` automate such conversion of text into circuits [8].

3 Comparison with Classical Language Models

Classical LLMs like GPT or BERT use deep neural networks with billions of parameters. In contrast, QNLP models are smaller (a few qubits) and often *hybrid*, where a quantum layer augments a classical model [9].

Classical transformers use dot-product similarity to model dependencies. Quantum circuits can entangle qubits and encode input vectors, capturing complex correlations. IonQ's hybrid model showed improved accuracy on SST-2 sentiment classification with more qubits [9]. Theory suggests some quantum models have greater *expressive power* than classical ones [10].

However, quantum models currently lack the scalability of classical LMs. They are typically interpretable (e.g., via compositional grammars), but constrained by hardware (qubit count, noise). Thus, classical LLMs dominate, while QNLP offers a different, potentially complementary paradigm.

4 Examples of Quantum and Hybrid Models

A notable example is IonQ's hybrid architecture for fine-tuning classical LLMs. A pre-trained sentence transformer encodes text, followed by a parameterized quantum circuit for classification. The quantum head outperforms classical baselines as qubit count increases [9].

Other hybrid examples include quantum-enhanced attention transformers, which use variational quantum circuits for attention scores [11]. These models show marginal gains (+1.5%) in sentiment analysis on IMDb compared to classical baselines.

References

[1] TechTarget. Quantum machine learning vs classical machine learning. 2022.

[2] Liu, J. (2022). Basics of quantum gates and circuits.

[3] Huang et al. (2022). Demonstration of Quantum Advantage in Learning.

[4] Pan et al. (2023). QLM-UT: A Quantum Language Model with Unitary Transform.

[5] Xu et al. (2023). QPFE-ERNIE: Quantum-inspired Position Feature Encoding.

[6] Coecke, B. et al. (2010). Mathematical foundations for QNLP.

[7] Zhong, W. et al. (2023). DisCoCirc: Quantum Compositional Models.

[8] TQML. (2021). lambeq: A QNLP circuit construction library.

[9] IonQ. (2023). Quantum-enhanced fine-tuning of LLMs.

[10] Biamonte, J. et al. (2017). Quantum machine learning.

[11] Tomal et al. (2025). Quantum-enhanced Transformer for NLP.