# A Survey of Quantum Natural Language Processing: From Compositional Models to NISQ-Era Empiricism

**Arpan Phukan and Asif Ekbal**
Department of Computer Science and Engineering, IIT Patna
arpan_2121cs33@iitp.ac.in, asif@iitp.ac.in

## Abstract

Quantum Natural Language Processing (QNLP) has emerged as a novel paradigm that leverages the principles of quantum mechanics to address fundamental challenges in language modeling, particularly in capturing compositional meaning. This survey charts the evolution of QNLP, from its theoretical foundations in the Distributional Compositional Categorical (DisCoCat) framework to its modern implementation on Noisy Intermediate-Scale Quantum (NISQ) hardware. We review the primary architectural approaches, including variational quantum circuits and tensor networks, and summarize the growing body of empirical work in tasks such as text classification, sentence similarity, and question answering. A recurring finding is the potential for QNLP models to achieve competitive performance with significantly fewer parameters than their classical counterparts. However, the field is critically constrained by the limitations of NISQ-era hardware. We conclude by discussing these challenges and outlining the future trajectory towards achieving a demonstrable quantum advantage and building more interpretable, efficient language models.

## 1 Introduction

Quantum Natural Language Processing (QNLP) is an integrative and rapidly developing field that applies the principles of quantum computing to the challenges of natural language processing (Pallavi and Prasanna Kumar, 2025). It is motivated by a foundational hypothesis that extends beyond the simple pursuit of computational speedup: the idea that language is "quantum native" (Widdows et al., 2024). This proposition suggests that the mathematical formalism of quantum mechanics, particularly the compositional structure of Hilbert spaces, provides a natural and perhaps ideal framework for modeling the compositional nature of linguistic meaning. By grounding language in a physical computational model, QNLP seeks a paradigm shift from the purely statistical and often opaque methods of classical NLP to a more structured and interpretable approach (Phukan et al., 2024).

This pursuit is driven by the persistent limitations of classical models. Even state-of-the-art Large Language Models (LLMs) struggle to robustly handle the principle of compositionality, the process by which the meanings of individual words combine according to grammatical rules to form the meaning of a sentence (Song et al., 2025). Many classical architectures effectively treat sentences as a "bag of words" or a flat sequence of tokens, failing to capture the deep, hierarchical relationships encoded in syntax (Chen et al., 2024b). Furthermore, natural language is inherently ambiguous. A phrase such as "The bank was crowded" presents a challenge that classical models resolve through statistical inference (Wu et al., 2021). Quantum mechanics, with its principles of superposition and entanglement, offers a potentially more efficient solution, allowing for the simultaneous representation and processing of multiple meanings within a single quantum state (Schuld and Killoran, 2019; Phukan et al., 2025). This quantum representation can then "collapse" to a definite meaning as more context becomes available, a process that arguably mirrors human cognitive processing of ambiguity (Phukan and Ekbal, 2023).

Finally, the exponential growth in the parameter counts and energy consumption of classical LLMs has created an urgent need for more efficient and scalable learning paradigms (Ji and Jiang, 2026). QNLP presents a potential path toward models that are not only more powerful but also more resource-efficient (Phukan et al., 2024; Phukan and Ekbal, 2023).

This survey provides a comprehensive overview of the QNLP landscape. Section 2 details the foundational Distributional Compositional Categorical



*Proceedings of the QuantumNLP: Integrating Quantum Computing with Natural Language Processing*, pages 65–75
November 24, 2025 

(DisCoCat) (Coecke et al., 2010) framework. Section 3 reviews the primary architectures used for implementing QNLP models on near-term quantum devices. Section 4 summarizes the empirical progress across key NLP tasks. Section 5 offers a critical discussion of the field's current challenges and future outlook.

## 2 Foundational Framework: Compositionality via DisCoCat

The theoretical cornerstone of modern QNLP is the Distributional Compositional Categorical (DisCoCat) framework (Coecke et al., 2010), which provides a mathematically rigorous unification of two central pillars of linguistic theory: the distributional hypothesis (a word's meaning is defined by its context) and the principle of compositionality (the meaning of a whole is a function of the meaning of its parts and how they are combined).

The DisCoCat model operates through a formal mapping between grammar and meaning. On the grammatical side, it employs a categorial grammar, typically a pregroup grammar, where words are assigned abstract grammatical types. For instance, a noun might be assigned type $n$, while a transitive verb that takes a noun as its object and a noun as its subject to form a sentence would have the type $n^r s n^l$, where $s$ is the type for a sentence and the superscripts $r$ and $l$ denote right and left adjoints, respectively (Refer Figure 1). A sequence of words is considered grammatical if its sequence of types can be reduced to the sentence type $s$ through a series of predefined rules (Yeung and Kartsaklis, 2021).

On the semantic side, word meanings are represented as vectors (or more generally, tensors) in a high-dimensional Hilbert space, following standard distributional semantics. The central innovation of DisCoCat is the use of category theory to define a structure-preserving function that maps the category of grammar to the category of vector spaces (i.e., semantics). This ensures that the reduction of grammatical types corresponds directly to a specific mathematical operation on the meaning vectors, namely, tensor contraction (Sadrzadeh et al., 2018). This entire compositional process can be visualized and reasoned about using string diagrams, an intuitive graphical calculus where boxes represent word meanings (tensors) and wires represent their grammatical types (tensor indices).

The profound insight that catalyzed the field of QNLP was the observation that the mathematical structure of pregroup grammars and the structure of quantum processes both form a rigid monoidal category. This shared structure allows for a direct and systematic translation from a sentence's grammatical string diagram to a quantum circuit. In this mapping, words become quantum states or operations, and the grammatical rules dictating their composition become entangling gates or measurements (Correia et al., 2022). This correspondence makes quantum computers the "native environment" for executing DisCoCat models. The framework can thus be conceptualized as a "compiler" for language: it takes a high-level linguistic structure (a sentence) as input, parses it according to a formal grammar, and outputs a low-level, executable representation (a quantum circuit) (Peral-García et al., 2024; Laakkonen et al.). Toolkits such as `lambeq`[1] and `DisCoPy`[2] have been developed to automate this compilation pipeline, providing a principled method for generating quantum algorithms for NLP tasks.

## 3 Architectures for Quantum Language Models

The implementation of QNLP models on present-day hardware has led to a variety of architectural approaches (Refer Table 1). These can be seen as existing on a spectrum, reflecting a fundamental trade-off between adherence to the pure, linguistically-grounded theory of DisCoCat and the pragmatic need to achieve robust performance on noisy, resource-constrained quantum devices.

### 3.1 Variational Quantum Circuits (VQCs)

Variational Quantum Circuits[3] (VQC) (Qi et al., 2023) are the dominant paradigm for executing machine learning tasks on Noisy Intermediate-Scale Quantum (NISQ) hardware (Phukan et al., 2024). A VQC is a quantum circuit that includes gates with adjustable parameters (e.g., rotation angles). It operates within a hybrid quantum-classical loop: the circuit is executed on a quantum processing unit (QPU), the output is measured to compute a classical loss function, and a classical optimizer updates the circuit parameters to minimize this loss, analogous to training a classical neural network(Bashiri and Naderi, 2024; Hong and Lopez,

[1] https://github.com/CQCL/lambeq
[2] https://discopy.org/
[3] https://pennylane.ai/qml/glossary/variational_circuit

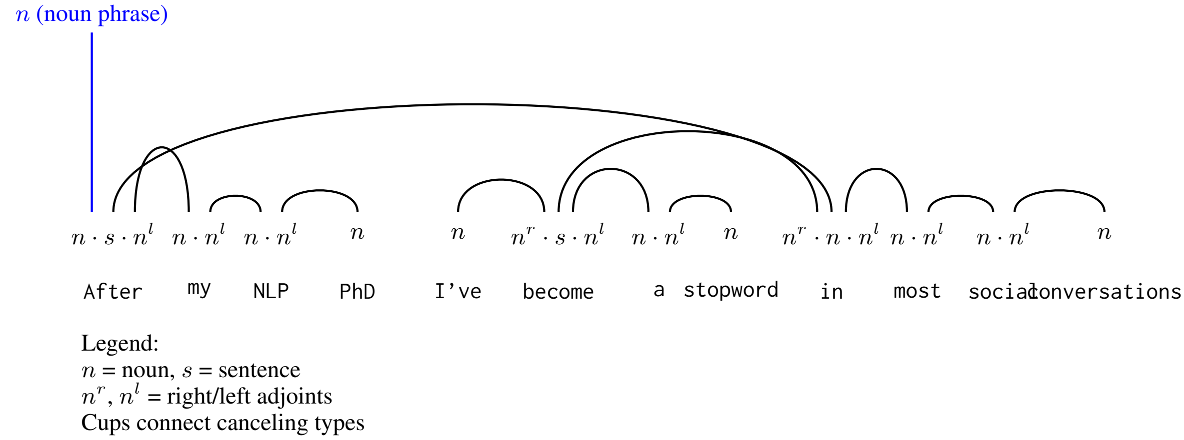


Figure 1: DisCoCat compositional diagram for "After my NLP PhD, I've become a stopword in most social conversations."

2025). Instead of loading pre-computed word vectors, the meanings of words are learned directly as the parameters of the quantum circuits (Zeng and Coecke, 2016). The DisCoCat framework provides the "quantum circuit skeleton" based on a sentence's grammar, and the free parameters within this structure are then optimized end-to-end for a specific downstream task, such as text classification.

### 3.2 Tensor Network (TN) Representations

Tensor Networks are a set of techniques originating from many-body quantum physics for efficiently representing and manipulating large, high-dimensional tensors (Christandl et al., 2024). This framework is deeply connected to QNLP, as quantum circuits themselves can be formally described as a specific class of tensor network (Rieser et al., 2023; Zhang et al., 2019). TNs, particularly one-dimensional structures like Matrix Product States (MPS), are naturally suited for modeling sequential data like language, as they are designed to efficiently capture local correlations (Berezutskii et al., 2025; Zhang et al., 2019; Teixeira et al.). They serve a dual role in QNLP: as a powerful tool for classically simulating quantum language models and as a class of machine learning models in their own right, offering a structured approach that lies between classical recurrent models and full quantum implementations (Berezutskii et al., 2025).

### 3.3 Hybrid Quantum-Classical Models

Representing the most pragmatic end of the architectural spectrum, hybrid models seek to enhance proven classical architectures by replacing specific components with quantum counterparts (Pandey et al., 2022, 2023; Phukan and Ekbal, 2023; Pandey and Pakray, 2023; Phukan et al., 2025). This approach aims to leverage quantum effects for computationally challenging subroutines while retaining the overall power and stability of classical frameworks.

A prominent example is the Quantum Transformer, which replaces classical modules like the self-attention mechanism with a VQC-based implementation (Concepcion, 2025; Kerenidis et al., 2024). The goal is to harness quantum properties like entanglement to capture complex contextual relationships between tokens more efficiently than is possible classically (Chen et al., 2024a). A specific instantiation of this idea is the Quantum Self-Attention Neural Network (QSANN) (Li et al., 2024a), which introduces a quantum version of self-attention designed to be scalable and implementable on NISQ devices. Notably, this model bypasses the need for rigid syntactic pre-processing required by pure DisCoCat models, making it more readily applicable to larger, real-world dataset.

These distinct architectural philosophies highlight the field's dynamic search for an optimal balance between theoretical elegance and empirical viability in the NISQ era.

## 4 Empirical Progress in QNLP Tasks

Despite the constraints of NISQ hardware, a growing body of empirical work has begun to explore the capabilities of QNLP models on a range of standard NLP tasks (Refer Table 2). A consistent theme emerging from these experiments is not necessarily a quantum speedup in terms of wall-clock time, but a significant advantage in terms of parameter and

| Approach | Core Principle | Strengths | Limitations/Challenges |
|---|---|---|---|
| Compositional (DisCoCat) | Maps grammatical structure directly to quantum processes via category theory. | Theoretically grounded, highly interpretable, "quantum-native" foundation. | Relies on rigid grammatical parsing, can be brittle; may be inefficient without variational training. |
| Variational QNLP (VQC-based) | Uses grammar-informed circuit skeletons with parameters learned via a hybrid quantum-classical loop. | Enables training on NISQ hardware, learns word meanings from data, avoids need for QRAM. | Susceptible to barren plateaus, optimization is challenging, performance is limited by hardware noise. |
| Hybrid (e.g., Quantum Transformer) | Replaces specific components of classical architectures (e.g., attention) with quantum circuits. | Leverages power of established classical models, targets specific computational bottlenecks. | Less theoretically pure, potential for quantum advantage is localized to a single component. |

Table 1: Comparison of Various Approaches

data efficiency.

### 4.1 Text Classification

Text classification is a foundational NLP task that involves assigning predefined labels to text data, with applications ranging from spam detection to topic analysis (Taha et al., 2024; Sun et al., 2023; Li et al., 2024b). It is the most widely explored task in experimental QNLP, serving as a primary benchmark for new models (Peral-García et al., 2024).

#### 4.1.1 Sentiment Classification

Sentiment classification, a task focused on identifying the emotional tone (e.g., positive, negative, neutral) of a text, has been a key testbed for QNLP (Jim et al., 2024). Early proof-of-concept experiments on IBM's quantum computers successfully demonstrated that VQC-based models derived from DisCoCat could be trained to solve simple classification problems (Ganguly et al., 2022). These studies also showed that on carefully constructed datasets, syntax-aware models provided a performance advantage over syntax-agnostic "bag-of-words" baselines, validating the core premise of the compositional approach (Kando et al., 2022).

More recent work has scaled these experiments to real-world data. The QSANN model (Li et al., 2024a), for example, was evaluated on sentiment classification tasks using subsets of the Yelp [4], IMDb (Maas et al., 2011), and Amazon datasets (McAuley and Leskovec, 2013). It was shown to outperform a comparable classical self-attention baseline while using dramatically fewer trainable parameters, for instance, 49 quantum parameters versus 785 classical parameters on the Yelp dataset. Other architectures, such as Quantum Recurrent Neural Networks (QRNNs) (Pandey et al., 2024), have also been proposed for classification, though current implementations remain limited to smaller datasets. While comparative studies acknowledge that classical methods generally still achieve higher accuracy on standard benchmarks, the potential for QNLP models to learn effectively with greater model efficiency is a key finding (Zhang et al., 2024).

#### 4.1.2 Aspect-Based Sentiment Analysis (ABSA)

Aspect-Based Sentiment Analysis (ABSA)[5] is a fine-grained sentiment analysis task that aims to identify the sentiment expressed towards specific aspects or features within a text (Hua et al., 2024). For example, in the sentence "The phone has a great camera but a poor battery life," ABSA would identify a positive sentiment towards the "camera" (aspect) and a negative sentiment towards the "battery life" (aspect) (Zhang et al., 2022). This level of detail is crucial for applications like customer feedback analysis. While ABSA is a significant area of research in classical NLP, its exploration using dedicated QNLP models is still an emerging field (Kayal et al., 2025).

### 4.2 Sentence Similarity

The task of measuring the semantic similarity between two sentences classically involves encoding them into vector embeddings and calculating a distance metric, such as cosine similarity (Gao et al., 2025). Quantum approaches have explored this task using several methods, such as semantic matching frameworks that leverage the density matrix formalism in Hilbert space (Zhang et al., 2025).

#### 4.2.1 Semantic Textual Similarity

One prominent technique for semantic textual similarity is Quantum Kernel Methods (Schuld et al., 2015; Schuld and Killoran, 2019), where sentences are mapped to quantum states via a quantum feature map. The similarity between two sentences is then calculated as the inner product (or fidelity)

[4] http://www.yelp.com/dataset_challenge

[5] https://www.sciencedirect.com/topics/computer-science/aspect-based-sentiment-analysis

of their corresponding quantum states, $f(x_i, x_j) = | < \psi(x_i)|\psi(x_j) > |^2$. This value serves as a kernel that can be fed into a classical support vector machine for classification or similarity tasks (Egginger et al., 2024).

Another innovative, quantum-inspired approach involves using a quantum circuit-based architecture (simulated on classical hardware) as a projection head to compress high-dimensional sentence embeddings from a classical model like BERT (Kankeu et al., 2025). The similarity between the resulting compressed representations is measured using a metric based on the fidelity of quantum states. This method achieved performance competitive with classical techniques but with 32 times fewer parameters, and it demonstrated particularly strong performance in low-data settings.

#### 4.2.2 Paraphrase Identification

Paraphrase identification is a specific binary classification task within sentence similarity that determines whether two sentences convey the same meaning, even if they use different wording (Alvi et al., 2021). This task is critical for applications such as plagiarism detection, text summarization, and improving question answering systems (Palivela, 2021). While a core challenge in classical NLP, the development of specialized QNLP models for paraphrase identification is an active area of research.

#### 4.2.3 Information Retrieval (IR)

Information Retrieval is the task of finding relevant documents or information from a large collection in response to a user's query. The connection to sentence similarity is foundational; vector space models, which represent both queries and documents as vectors, measure relevance based on the similarity between these vectors (Sordoni and Nie, 2013). This vector-based paradigm is well-suited for quantum implementation. Mathematical models for language explicitly motivated by quantum theory have been successfully applied to IR (Fan et al., 2024), suggesting that quantum computers could offer a natural and efficient environment for these tasks. This link is formally demonstrated by frameworks that generalize the probability framework of quantum physics for interactive retrieval tasks (Piwowarski and Lalmas, 2009).

### 4.3 Question Generation (QG)

Question Generation is the task of automatically generating natural language questions from a given input, such as a text passage or a knowledge base (Mulla and Gharpure, 2023). QG has important applications in creating educational materials, augmenting datasets for QA systems, and enhancing conversational agents (Duan et al., 2017; Dey and Rudrapal, 2024). While QNLP has been successfully applied to the dual task of question answering, direct applications to the generative task of QG are still in the early stages of exploration, with current research still focused on addressing core challenges in classical settings (Dey and Rudrapal, 2024), and thus represent a key direction for future research.

### 4.4 Question Answering (QA)

Question answering remains a nascent but promising application area for QNLP, with research exploring how quantum algorithms can provide advantages like potential speedups (Correia et al., 2022). While classical QA systems involve complex pipelines for information retrieval and answer generation, quantum-inspired techniques are beginning to emerge. One such innovation is the Quantum Fusion Module (QFM) (Duan et al., 2024), designed for open-domain QA. This module applies principles from quantum theory to fuse the token embeddings from a question and a candidate passage into a single representation analogous to a "quantum mixed state." This fused representation allows a classifier to more accurately predict whether the passage contains the answer, thereby improving the performance of a larger, classical T5-based model (Raffel et al., 2020).

### 4.5 Processing Multimodal Data

A significant new field for QNLP is the processing of multimodal data, which involves integrating information from different formats such as text, images/video, and audio (Li et al., 2021; Phukan and Ekbal, 2023; Phukan et al., 2024, 2025). The MultiQ-NLP framework (Hawashin and Sadrzadeh, 2024) has been developed specifically for this purpose, extending the compositional, structure-aware models of QNLP to handle both text and images. In this framework, both linguistic components and image features are translated into quantum circuits. This allows the model to create a unified representation that captures the interactions between modalities. When tested on an image classification task that requires understanding the relationship between a subject, verb, and object (SVO-Probes), the structure-aware quantum model performed on par with state-of-the-art classical models, demonstrating the viability of QNLP for complex multimodal reasoning. Phukan and Ekbal (2023) introduced QeMMA, a framework for multimodal sentiment analysis. On the CMU-MOSEI dataset (Zadeh et al., 2018), their model, implemented with Qiskit[6] and executed on quantum hardware, outperformed classical baselines by 3.52% in accuracy and 10.14% in F1-score, providing concrete evidence of the empirical advantage of quantum approaches.

Further advancing this line of work, Phukan et al. (2024) proposed a hybrid quantum-classical architecture designed to jointly tackle the interrelated tasks of sarcasm, emotion, and sentiment detection. Their model integrates a Variational Quantum Circuit (VQC) with a classical deep neural network, hypothesizing that quantum entanglement and state space properties can more effectively model the nuanced cross-modal interactions and task correlations present in such complex inference problems.

#### 4.5.1 Emotion Recognition

Emotion recognition is a more granular task than sentiment analysis, aiming to identify specific emotions like joy, anger, or guilt from data (Rasool et al., 2025). While much of the work in this area uses classical NLP models on text, QNLP research has begun to explore this task, often by processing multimodal data. Quantum-enhanced models have been applied to recognize emotions from physiological signals using quantum-enhanced Support Vector Machines (SVMs) (Bayro and Jeong, 2025), and from facial expressions using Quantum Neural Networks (QNNs) (Alsubai et al., 2024). Furthermore, hybrid quantum-classical models have been developed for multimodal data, including speech and text (Li et al., 2025). Such frameworks are increasingly designed to analyze sentiment, emotion, and other related states simultaneously, leveraging quantum properties to model the complex correlations between them (Li et al., 2025).

#### 4.5.2 Sarcasm Detection

Sarcasm detection is a particularly challenging NLP task due to the inherent incongruity between the literal and intended meaning of an utterance (Xi et al., 2025; Lu et al., 2025). Quantum computing has been applied to this problem to better model such complex linguistic phenomena, for instance, through Quantum Neural Networks (QNNs) for analyzing sentiment, emotion, and sarcasm (Singh et al., 2025; Phukan et al., 2024). Other quantum-based models designed directly for this task include the Quantum Fuzzy Neural Network (QFNN) for multimodal sarcasm detection (analyzing text, audio, and visuals) (Tiwari et al., 2024) and frameworks that use the mathematics of quantum probability to jointly model sarcasm, sentiment, and emotion (Singh et al., 2025; Phukan et al., 2025).

## 5 Challenges and Future Outlook

The trajectory of QNLP is shaped by a dynamic interplay between the ambitious theoretical promises of quantum computation and the stark realities of current hardware. Progress in the field can be understood as advancing on two parallel but interdependent fronts: a hardware-aware effort to extract value from today's imperfect machines, and a theory-forward exploration of what will be possible with the fault-tolerant computers of the future.

### 5.1 The NISQ Bottleneck

The primary constraint on QNLP research is the Noisy Intermediate-Scale Quantum (NISQ) era of hardware (Rai et al., 2022). Today's quantum processors are limited by several fundamental factors:

- Low Qubit Counts: Systems with 50-100 qubits are typical, which severely restricts the size and complexity of the language models that can be implemented (Riel, 2021; Pan et al., 2025).
- High Error Rates: Qubits are extremely sensitive to their environment, suffering from decoherence that introduces noise and leads to high error rates (typically 0.1%-1%) in gate operations and measurements (Rai et al., 2022). This noise accumulates rapidly, corrupting the results of all but the simplest computations (Khan et al., 2024).
- Limited Circuit Depth: Because noise accumulates with every operation, there is a strict limit on the number of gates (circuit depth, typically 20-100 layers) that can be applied before the quantum signal is overwhelmed by noise (Pan et al., 2025).

[6] https://www.ibm.com/quantum/qiskit

| Task | Model/Approach | Dataset(s) | Key Finding / Reported Advantage |
|---|---|---|---|
| **Text Classification** | DisCoCat + VQC (Ganguly et al., 2022) | Synthetic (Ganguly et al., 2022) | Syntax-aware model outperforms bag-of-words. |
| **Text Classification** | QSANN (Li et al., 2024a) | Yelp[4], IMDb (Maas et al., 2011), and Amazon datasets (McAuley and Leskovec, 2013) | Outperforms classical baseline with >90% fewer parameters. |
| **Sentence Similarity** | Quantum-Inspired Projection Head (Kankeu et al., 2025) | TREC 2019 DL (Voorhees et al., 2020) and TREC 2020 DL (Craswell et al., 2021) | Competitive performance with 32x fewer parameters; excels in low-data regimes. |
| **Question Answering** | Quantum-Inspired Fusion-in-Decoder (QFiD) (Duan et al., 2024) | Natural Questions (Kwiatkowski et al., 2019) and TriviaQA (Joshi et al., 2017) | Quantum-inspired fusion improves relationship prediction over baseline. |
| **Multimodal Classification** | MultiQ-NLP (Hawashin and Sadrzadeh, 2024) | SVO-Probes (Hendricks and Nematzadeh, 2021) | Performs on par with classical models while offering better interpretability. |
| **Multimodal Classification** | HQNN (VQC+NN) (Phukan et al., 2024) | Extended MUStARD(Chauhan et al., 2020) | Outperforms classical baselines for multitask sarcasm, sentiment, and emotion detection. |

Table 2: QNLP Tasks

These hardware limitations collectively mean that current QNLP experiments are necessarily confined to small-scale datasets and simplified “toy” problems. Scaling these models to handle the complexity and volume of data processed by modern LLMs is, for now, impossible.

### 5.2 Algorithmic Hurdles

Beyond hardware, significant algorithmic challenges must be overcome. A primary obstacle in training VQCs is the phenomenon of barren plateaus, where the gradient of the loss function vanishes exponentially as the number of qubits increases, rendering the optimization process intractable for larger models (Pande, 2023). Furthermore, the task of data encoding, efficiently translating classical text data into quantum states, is a non-trivial problem that can itself become a computational bottleneck (Ranga et al., 2024). Active research on the hardware-aware front is focused on developing mitigation strategies, including noise-aware training protocols (Rahman and Zhuang, 2025), quantum error mitigation techniques (Atban et al., 2025), and the design of specialized loss functions and optimizers (Pande, 2023).

## 6 Conclusion

The long-term vision for QNLP is to achieve a demonstrable quantum advantage, where a quantum computer solves a practical NLP problem more efficiently, more accurately, or with fewer resources than is possible with any classical machine. While theoretical results have shown that quantum language models are, in principle, more expressive than their classical counterparts (i.e., the problem is BQP-complete), a practical advantage has not yet been realized.

Perhaps one of the most compelling promises of the theory-forward front is the potential for QNLP to deliver more interpretable and trustworthy AI. By building models on the transparent, compositional structure of DisCoCat, QNLP offers a path away from the “black box” nature of many LLMs, toward systems whose reasoning can be traced and understood.

Ultimately, the future of QNLP depends on the co-evolution of hardware and theory. As quantum technology matures beyond the NISQ era and toward fault-tolerance, the full potential of these linguistically-grounded, quantum-native models can be explored. The goal is not merely to replicate classical NLP on a different substrate, but to fundamentally reimagine language modeling by leveraging the unique computational capabilities of the quantum world, potentially leading to a new generation of AI that is more powerful, efficient, and reliable.

## 7 Limitations

This survey provides a broad overview of the QNLP landscape, but it is subject to several inherent limitations reflecting the nascent and rapidly evolving nature of the field.

First, the scope of this review is primarily centered on the theoretical lineage from the DisCoCat framework to its empirical implementation on NISQ-era hardware. While we touch upon quantum-inspired classical models and tensor networks, an exhaustive analysis of all classical algorithms that leverage principles from quantum theory is beyond the scope of this paper.

Second, the field of QNLP is advancing at an exceptional pace. This survey represents a snapshot of the research landscape at the time of writing. New hardware developments, algorithmic breakthroughs, and empirical findings are published frequently, and some emerging work may not be captured here.

## References


Shtwai Alsubai, Abdullah Alqahtani, Abed Alanazi, Mohemmed Sha, and Abdu Gumaei. 2024. Facial emotion recognition using deep quantum and advanced transfer learning mechanism. *Frontiers in Computational Neuroscience*, 18:1435956.

Faisal Alvi, Mark Stevenson, and Paul Clough. 2021. Paraphrase type identification for plagiarism detection using contexts and word embeddings. *International Journal of Educational Technology in Higher Education*, 18(1):42.

Furkan Atban, Muhammed Yusuf Küçükkara, and Cüneyt Bayılmış. 2025. Enhancing variational quantum classifier performance with meta-heuristic feature selection for credit card fraud detection. *The European Physical Journal Special Topics*, pages 1–14.

Hadis Bashiri and Hassan Naderi. 2024. Comprehensive review and comparative analysis of transformer models in sentiment analysis. *Knowledge and Information Systems*, 66(12):7305–7361.

Allison Bayro and Heejin Jeong. 2025. Enhancing emotional response detection in virtual reality with quantum support vector machine learning. *Computers & Graphics*, 128:104196.

Aleksandr Berezutskii, Minzhao Liu, Atithi Acharya, Roman Ellerbrock, Johnnie Gray, Reza Haghshenas, Zichang He, Abid Khan, Viacheslav Kuzmin, Dmitry Lyakh, and 1 others. 2025. Tensor networks for quantum computing. *Nature Reviews Physics*, pages 1–13.

Dushyant Singh Chauhan, SR Dhanush, Asif Ekbal, and Pushpak Bhattacharyya. 2020. Sentiment and emotion help sarcasm? a multi-task learning framework for multi-modal sarcasm, sentiment and emotion analysis. In *Proceedings of the 58th annual meeting of the association for computational linguistics*, pages 4351–4360.

I-Chi Chen, Harshdeep Singh, VL Anukruti, Brian Quanz, and Kavitha Yogaraj. 2024a. A survey of classical and quantum sequence models. In *2024 16th International Conference on COMmunication Systems & NETworkS (COMSNETS)*, pages 1006–1011. IEEE.

Jiaao Chen, Xiaoman Pan, Dian Yu, Kaiqiang Song, Xiaoyang Wang, Dong Yu, and Jianshu Chen. 2024b. Skills-in-context: Unlocking compositionality in large language models. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 13838–13890.

Matthias Christandl, Vladimir Lysikov, Vincent Steffan, Albert H Werner, and Freek Witteveen. 2024. The resource theory of tensor networks. *Quantum*, 8:1560.

Bob Coecke, Mehrnoosh Sadrzadeh, and Stephen Clark. 2010. Mathematical foundations for a compositional distributional model of meaning. *arXiv preprint arXiv:1003.4394*.

Wesley Concepcion. 2025. Hybrid quantum transformer for natural language processing.

Adriana D Correia, Michael Moortgat, and Henk TC Stoof. 2022. Quantum computations for disambiguation and question answering. *Quantum Information Processing*, 21(4):126.

Nick Craswell, Bhaskar Mitra, Emine Yilmaz, and Daniel Campos. 2021. Overview of the TREC 2020 deep learning track. *CoRR*, abs/2102.07662.

Debopam Dey and Dwijen Rudrapal. 2024. An analysis of automatic question generation research progress and challenges. In *International Conference on Data Science and Network Engineering*, pages 247–258. Springer.

Nan Duan, Duyu Tang, Peng Chen, and Ming Zhou. 2017. Question generation for question answering. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 866–874, Copenhagen, Denmark. Association for Computational Linguistics.

Ruixue Duan, Xin Liu, Zhigang Ding, and Yangsen Zhang. 2024. Quantum-inspired fusion for open-domain question answering. *Electronics*, 13(20):4135.

Sebastian Egginger, Alona Sakhnenko, and Jeanette Miriam Lorenz. 2024. A hyperparameter study for quantum kernel methods. *Quantum Machine Intelligence*, 6(2):44.

Zipeng Fan, Jing Zhang, Peng Zhang, Qianxi Lin, Yizhe Li, and Yuhua Qian. 2024. Quantum-inspired language models based on unitary transformation. *Information Processing & Management*, 61(4):103741.

Srinjoy Ganguly, Sai Nandan Morapakula, and Luis Miguel Pozo Coronado. 2022. Quantum natural language processing based sentiment analysis using lambeq toolkit. In *2022 Second International Conference on Power, Control and Computing Technologies (ICPC2T)*, pages 1–6. IEEE.

Hui Gao, Peng Zhang, Jing Zhang, and Chang Yang. 2025. Qsim: a quantum-inspired hierarchical semantic interaction model for text classification. *Neurocomputing*, 611:128658.

Hala Hawashin and Mehrnoosh Sadrzadeh. 2024. Multimodal structure-aware quantum data processing. *arXiv preprint arXiv:2411.04242*.

Lisa Anne Hendricks and Aida Nematzadeh. 2021. Probing image-language transformers for verb understanding. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 3635–3644.

Ying-Yi Hong and Dylan Josh Domingo Lopez. 2025. A review on quantum machine learning in applied systems and engineering. *IEEE Access*.

Yan Cathy Hua, Paul Denny, Jörg Wicker, and Katerina Taskova. 2024. A systematic review of aspect-based sentiment analysis: domains, methods, and trends. *Artificial Intelligence Review*, 57(11):296.

Zhenya Ji and Ming Jiang. 2026. A systematic review of electricity demand for large language models: evaluations, challenges, and solutions. *Renewable and Sustainable Energy Reviews*, 225:116159.

Jamin Rahman Jim, Md Apon Riaz Talukder, Partha Malakar, Md Mohsin Kabir, Kamruddin Nur, and Mohammed Firoz Mridha. 2024. Recent advancements and challenges of nlp-based sentiment analysis: A state-of-the-art review. *Natural Language Processing Journal*, 6:100059.

Mandar Joshi, Eunsol Choi, Daniel Weld, and Luke Zettlemoyer. 2017. TriviaQA: A large scale distantly supervised challenge dataset for reading comprehension. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1601–1611, Vancouver, Canada. Association for Computational Linguistics.

Shunsuke Kando, Hiroshi Noji, and Yusuke Miyao. 2022. Multilingual syntax-aware language modeling through dependency tree conversion. In *Proceedings of the Sixth Workshop on Structured Prediction for NLP*, pages 1–10, Dublin, Ireland. Association for Computational Linguistics.

Ivan Kankeu, Stefan Gerd Fritsch, Gunnar Schönhoff, Elie Mounzer, Paul Lukowicz, and Maximilian Kiefer-Emmanouilidis. 2025. Quantum-inspired embeddings projection and similarity metrics for representation learning. *arXiv preprint arXiv:2501.04591*.

Mrinmoy Kayal, Mohinikanta Sahoo, Jayadeep Pati, and Ranjan Kumar Behera. 2025. Quantum-inspired aspect-based sentiment analysis using natural language processing. In *Advances in Quantum Inspired Artificial Intelligence: Techniques and Applications*, pages 151–169. Springer.

Iordanis Kerenidis, Natansh Mathur, Jonas Landman, Martin Strahm, Yun Yvonna Li, and 1 others. 2024. Quantum vision transformers. *Quantum*, 8:1265.

Misha Urooj Khan, Muhammad Ahmad Kamran, Wajiha Rahim Khan, Malik Muhammad Ibrahim, Muhammad Umair Ali, and Seung Won Lee. 2024. Error mitigation in the nisq era: Applying measurement error mitigation techniques to enhance quantum circuit performance. *Mathematics*, 12(14):2235.

Tom Kwiatkowski, Jennimaria Palomaki, Olivia Redfield, Michael Collins, Ankur Parikh, Chris Alberti, Danielle Epstein, Illia Polosukhin, Jacob Devlin, Kenton Lee, and 1 others. 2019. Natural questions: a benchmark for question answering research. *Transactions of the Association for Computational Linguistics*, 7:453–466.

Tuomas Laakkonen, Konstantinos Meichanetzidis, and Bob Coecke. Quantum algorithms for compositional text processing. *EPTCS 406*, page 162.

Guangxi Li, Xuanqiang Zhao, and Xin Wang. 2024a. Quantum self-attention neural networks for text classification. *Science China Information Sciences*, 67(4):142501.

Qiuchi Li, Dimitris Gkoumas, Christina Lioma, and Massimo Melucci. 2021. Quantum-inspired multimodal fusion for video sentiment analysis. *Information Fusion*, 65:58–71.

Xintong Li, Jinya Jiang, Ria Dharmani, Jayanth Srinivasa, Gaowen Liu, and Jingbo Shang. 2024b. Open-world multi-label text classification with extremely weak supervision. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 15084–15096, Miami, Florida, USA. Association for Computational Linguistics.

YaoChong Li, Yi Qu, Ri-Gui Zhou, and Jing Zhang. 2025. Qmlsc: A quantum multimodal learning model for sentiment classification. *Information Fusion*, 120:103049.

Ming Lu, Zhiqiang Dong, Ziming Guo, Xiaoming Zhang, Xinxi Lu, Tianbo Wang, and Litian Zhang. 2025. A multi-modal sarcasm detection model based on cue learning. *Scientific Reports*, 15(1):10261.

Andrew L. Maas, Raymond E. Daly, Peter T. Pham, Dan Huang, Andrew Y. Ng, and Christopher Potts. 2011. Learning word vectors for sentiment analysis. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 142–150, Portland, Oregon, USA. Association for Computational Linguistics.

Julian McAuley and Jure Leskovec. 2013. Hidden factors and hidden topics: understanding rating dimensions with review text. In *Proceedings of the 7th ACM conference on Recommender systems*, pages 165–172.

Nikahat Mulla and Prachi Gharpure. 2023. Automatic question generation: a review of methodologies, datasets, evaluation metrics, and applications. *Progress in Artificial Intelligence*, 12(1):1–32.

Hemant Palivela. 2021. Optimization of paraphrase generation and identification using language models in natural language processing. *International Journal of Information Management Data Insights*, 1(2):100025.

Gundala Pallavi and Rangarajan Prasanna Kumar. 2025. Quantum natural language processing and its applications in bioinformatics: a comprehensive review of methodologies, concepts, and future directions. *Frontiers in Computer Science*, 7:1464122.

Yi Pan, Hanqi Jiang, Junhao Chen, Yiwei Li, Huaqin Zhao, Lin Zhao, Yohannes Abate, Yingfeng Wang, and Tianming Liu. 2025. Bridging classical and quantum computing for next-generation language models. *arXiv preprint arXiv:2508.07026*.

Mandaar B Pande. 2023. Review of optimization techniques and barren plateaus in training of quantum machine learning problems. In *International Conference on Computing and Network Communications*, pages 3–15. Springer.

Shyambabu Pandey, Nihar Jyoti Basisth, Tushar Sachan, Neha Kumari, and Partha Pakray. 2023. Quantum machine learning for natural language processing application. *Physica A: Statistical Mechanics and its Applications*, 627:129123.

Shyambabu Pandey, Pankaj Dadure, Morrel VL Nunsanga, and Partha Pakray. 2022. Parts of speech tagging towards classical to quantum computing. In *2022 IEEE Silchar Subsection Conference (SILCON)*, pages 1–6. IEEE.

Shyambabu Pandey and Partha Pakray. 2023. Bi-quantum long short-term memory for part-of-speech tagging. In *Proceedings of the 20th International Conference on Natural Language Processing (ICON)*, pages 301–307.

Shyambabu Pandey, Partha Pakray, and Riyanka Manna. 2024. Quantum classifier for natural language processing applications. *Computación y Sistemas*, 28(2):695–700.

David Peral-García, Juan Cruz-Benito, and Francisco José García-Peñalvo. 2024. Comparing natural language processing and quantum natural processing approaches in text classification tasks. *Expert Systems with Applications*, 254:124427.

Arpan Phukan and Asif Ekbal. 2023. Qemma: Quantum-enhanced multi-modal sentiment analysis. In *Proceedings of the 20th International Conference on Natural Language Processing (ICON)*, pages 815–821.

Arpan Phukan, Anas Anwarul Haq Khan, and Asif Ekbal. 2025. Qumin: quantum multi-modal data fusion for humor detection. *Multimedia Tools and Applications*, 84(18):18855–18872.

Arpan Phukan, Santanu Pal, and Asif Ekbal. 2024. Hybrid quantum-classical neural network for multimodal multitask sarcasm, emotion, and sentiment analysis. *IEEE Transactions on Computational Social Systems*, 11(5):5740–5750.

Benjamin Piwowarski and Mounia Lalmas. 2009. A quantum-based model for interactive information retrieval. In *Conference on the Theory of Information Retrieval*, pages 224–231. Springer.

Jun Qi, Chao-Han Yang, and Pin-Yu Chen. 2023. Qtn-vqc: An end-to-end learning framework for quantum neural networks. *Physica Scripta*, 99(1):015111.

Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of Machine Learning Research*, 21(140):1–67.

Maqsudur Rahman and Jun Zhuang. 2025. Nqnn: Noise-aware quantum neural networks for medical image classification. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 433–442. Springer.

Anshuman Rai, Amey Talekar, Javed Khan, and Ahlam Ansari. 2022. A review article on quantum natural language processing. *Int J Res Appl Sci Eng Technol*, 10(1).

Deepak Ranga, Aryan Rana, Sunil Prajapat, Pankaj Kumar, Kranti Kumar, and Athanasios V Vasilakos. 2024. Quantum machine learning: Exploring the role of data encoding techniques, challenges, and future directions. *Mathematics*, 12(21):3318.

Abdur Rasool, Saba Aslam, Naeem Hussain, Sharjeel Imtiaz, and Waqar Riaz. 2025. nbert: Harnessing nlp for emotion recognition in psychotherapy to transform mental health care. *Information*, 16(4):301.

Heike Riel. 2021. Quantum computing technology. In *2021 IEEE international Electron devices meeting (IEDM)*, pages 1–3. IEEE.

Hans-Martin Rieser, Frank Köster, and Arne Peter Raulf. 2023. Tensor networks for quantum machine learning. *Proceedings of the Royal Society A*, 479(2275):20230218.

Mehrnoosh Sadrzadeh, Dimitri Kartsaklis, and Esma Balkır. 2018. Sentence entailment in compositional distributional semantics. *Annals of Mathematics and Artificial Intelligence*, 82(4):189–218.

Maria Schuld and Nathan Killoran. 2019. Quantum machine learning in feature hilbert spaces. *Physical review letters*, 122(4):040504.

Maria Schuld, Ilya Sinayskiy, and Francesco Petruccione. 2015. An introduction to quantum machine learning. *Contemporary Physics*, 56(2):172–185.

Jaiteg Singh, Kamalpreet Singh Bhangu, Abdulrhman Alkhanifer, Ahmad Ali Alzubi, and Farman Ali. 2025. Quantum neural networks for multimodal sentiment, emotion, and sarcasm analysis. *Alexandria Engineering Journal*, 124:170–187.

Jiajun Song, Zhuoyan Xu, and Yiqiao Zhong. 2025. Out-of-distribution generalization via composition: a lens through induction heads in transformers. *Proceedings of the National Academy of Sciences*, 122(6):e2417182122.

Alessandro Sordoni and Jian-Yun Nie. 2013. Looking at vector space and language models for ir using density matrices. In *International Symposium on Quantum Interaction*, pages 147–159. Springer.

Xiaofei Sun, Xiaoya Li, Jiwei Li, Fei Wu, Shangwei Guo, Tianwei Zhang, and Guoyin Wang. 2023. Text classification via large language models. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 8990–9005, Singapore. Association for Computational Linguistics.

Kamal Taha, Paul D Yoo, Chan Yeun, Dirar Homouz, and Aya Taha. 2024. A comprehensive survey of text classification techniques and their research applications: Observational and experimental insights. *Computer Science Review*, 54:100664.

Erico Souza Teixeira, Yara Rodrigues Inácio, and Pamela TL Bezerra. Applying quantum tensor networks in machine learning: A systematic.

Prayag Tiwari, Lailei Zhang, Zhiguo Qu, and Ghulam Muhammad. 2024. Quantum fuzzy neural network for multimodal sentiment and sarcasm detection. *Information Fusion*, 103:102085.

Ellen M Voorhees, Nick Craswell, Bhaskar Mitra, Daniel Campos, and Emine Yilmaz. 2020. Overview of the trec 2019 deep learning track.

Dominic Widdows, Willie Aboumrad, Dohun Kim, Sayonee Ray, and Jonathan Mei. 2024. Quantum natural language processing. *KI-Künstliche Intelligenz*, 38(4):293–310.

Sixuan Wu, Jian Li, Peng Zhang, and Yue Zhang. 2021. Natural language processing meets quantum physics: A survey and categorization. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 3172–3182.

Zhonghao Xi, Bengong Yu, and Haoyu Wang. 2025. Multimodal sarcasm detection based on sentiment-clue inconsistency global detection fusion network. *Expert Systems with Applications*, 275:127020.

Richie Yeung and Dimitri Kartsaklis. 2021. A CCG-based version of the DisCoCat framework. In *Proceedings of the 2021 Workshop on Semantic Spaces at the Intersection of NLP, Physics, and Cognitive Science (SemSpace)*, pages 20–31, Groningen, The Netherlands. Association for Computational Linguistics.

AmirAli Bagher Zadeh, Paul Pu Liang, Soujanya Poria, Erik Cambria, and Louis-Philippe Morency. 2018. Multimodal language analysis in the wild: Cmu-mosei dataset and interpretable dynamic fusion graph. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2236–2246.

William Zeng and Bob Coecke. 2016. Quantum algorithms for compositional natural language processing. *EPTCS*, 221(arXiv: 1608.01406):67–75.

Chenchen Zhang, Qiuchi Li, Dawei Song, and Prayag Tiwari. 2025. Quantum-inspired semantic matching based on neural networks with the duality of density matrices. *Engineering Applications of Artificial Intelligence*, 140:109667.

Hui Zhang, Qinglin Zhao, and Chuangtao Chen. 2024. A light-weight quantum self-attention model for classical data classification. *Applied Intelligence*, 54(4):3077–3091.

Lipeng Zhang, Peng Zhang, Xindian Ma, Shuqin Gu, Zhan Su, and Dawei Song. 2019. A generalized language model in tensor space. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 7450–7458.

Wenxuan Zhang, Xin Li, Yang Deng, Lidong Bing, and Wai Lam. 2022. A survey on aspect-based sentiment analysis: Tasks, methods, and challenges. *IEEE Transactions on Knowledge and Data Engineering*, 35(11):11019–11038.