**Patent Mining of Quantum Computing Innovations using NLP**

**1. Introduction & Objective**

Quantum Computing is emerging as one of the most disruptive technologies, with firms such as Google, IBM, Intel, Cisco, Amazon, IonQ, Bull SAS, Classiq, and Fujitsu leading innovation. Patent mining provides a way to systematically analyze intellectual property to identify research hotspots, trends, and competitive strategies.

**Objective:**

To analyze Quantum Computing patents (since 1 January 2025, from Lens.org) using NLP techniques (TF-IDF, Word2Vec, Topic Modeling) in order to uncover innovation landscapes, thematic clusters, and semantic relationships.

**2. Methodology**

**2.1 Data Source**

Dataset: Quantum-Computing.csv

Field analyzed: Patent Titles

**2.2 Preprocessing**

Converted all text to lowercase.

Removed numbers, punctuation, and non-alphabetic characters.

Tokenized into words.

Removed stopwords using NLTK + custom patent stopwords (system, method, apparatus, etc.).

**2.3 TF-IDF Analysis**

Used TfidfVectorizer (scikit-learn).

Extracted top 10 keywords per patent title.

**2.4 Word2Vec Embeddings**

Trained Word2Vec (Gensim) on tokenized documents.

Parameters: vector\_size = 100, window = 5, min\_count = 1, sg = 1.

Checked similarity for keywords “quantum” and “qubit”.

Applied PCA for visualization of embeddings in 2D.

**2.5 Topic Modeling (LDA)**

Constructed Bag-of-Words corpus.

Trained LDA model with 3 topics and 15 passes.

Used pyLDAvis for visualization.

Assigned dominant topic per patent.

**3. Results & Observations**

**3.1 TF-IDF Findings**

Most frequent/high-weight terms included: quantum, computing, circuit, qubits, state, error, low.

📌 Suggests heavy focus on error correction, quantum-classical methods, and circuit-level optimization.

**3.2 Word2Vec Similarity**

**Query Word Top Similar Words**

Quantum [('computing', 0.8667), ('circuit', 0.8513), ('qubits', 0.8442), ('error', 0.8223), ('state', 0.8133)]

Qubit [('quantum', 0.7552), ('qubits', 0.7159), ('low', 0.7139), ('error', 0.7059), ('computing', 0.6963)]

📌 Observation: Both quantum and qubit cluster strongly with computing, error, and circuit, showing the centrality of fault-tolerant architectures.

**3.3 PCA Visualization (Word Embeddings)**



**3.4 Topic Modeling (LDA – 3 Topics)**

Topic 1:

0.120\*"quantum" + 0.061\*"computing" + 0.024\*"system" + 0.021\*"device"

+ 0.011\*"classical" + 0.010\*"methods" + 0.010\*"apparatus"

+ 0.009\*"circuit" + 0.007\*"hybrid" + 0.007\*"learning"

Topic 2:

0.079\*"quantum" + 0.023\*"gates" + 0.019\*"computing" + 0.017\*"devices"

+ 0.014\*"circuit" + 0.013\*"ion" + 0.013\*"qubit" + 0.010\*"dot"

+ 0.010\*"device" + 0.009\*"computers"

Topic 3:

0.086\*"quantum" + 0.032\*"error" + 0.017\*"computing" + 0.017\*"correction"

+ 0.015\*"codes" + 0.014\*"processing" + 0.012\*"information"

+ 0.011\*"qubit" + 0.010\*"circuits" + 0.009\*"qubits"

**📌 Interpretation:**

Topic 1 → Hybrid quantum-classical architectures, ML integration.

Topic 2 → Hardware focus (ion-trap qubits, gates, semiconductor devices).

Topic 3 → Error correction, coding theory, and fault tolerance.



**4. Discussion**

TF-IDF confirmed frequent innovation terms like quantum, computing, circuit, error, hybrid.

Word2Vec revealed semantic proximity between quantum, qubit, error, computing, circuit.

LDA Topics uncovered three distinct innovation streams:

Hybrid systems with ML.

Hardware/gate-level innovations.

Error correction & coding.

**5. Conclusion**

This patent mining study demonstrates the effectiveness of NLP (TF-IDF, Word2Vec, LDA) for extracting insights from quantum computing patents.

Key Innovation Areas Identified:

Fault-tolerant architectures via error correction.

Hybrid quantum-classical integration for optimization.

Hardware innovations, especially ion-trap and circuit-level design.

Strategic Implication:

Leading firms (Google, IBM, IonQ, etc.) are converging on error resilience and hybrid quantum architectures, signaling the next phase of scalable, industrial quantum computing.

Final Output Dataset:

📂 Quantum-Computing\_with\_topics.csv