1. $q\_semantic\_similarity$ – Cosine similarity between subject & body

How it was added: We take the **email subject** (text) and the **body content** (`web_content`). Use **TF-IDF vectorization** to convert text into numeric vectors. Compute **cosine similarity** between subject vector and body vector:

$$\text{cosine similarity} = \frac{A \cdot B}{\|A\|\|B\|}$$

Result is a value between 0 (completely different) and 1 (very similar).

Why it helps detect phishing: Phishing emails often have subjects that **don't match the email body** ("urgent payment request" vs unrelated content). A low similarity can indicate suspicious or deceptive emails.

Quantum connection: In **quantum machine learning**, this similarity can be **encoded as a rotation angle** on a qubit: Feature vector $\rightarrow R_y(\theta)$ rotation. Qubits can represent **superpositions of semantic states**, enabling a QML model to detect patterns that classical linear correlations might miss.

2. q_domain_trust – Combines TLD, HTTPS, and domain age

How it was added: Check **top-level domain (TLD)** (`.gov`, `.edu`, `.org`) → usually trustworthy. Check **HTTPS presence** → 1 if secure, 0 if not. Check **domain age** → older domains are generally more legitimate. Combine as a **weighted score**:

$$\text{trust} = 0.4 \cdot \text{TLD} + 0.4 \cdot \text{HTTPS} + 0.2 \cdot \text{age (normalized)}$$

Why it helps detect phishing: Phishing emails often use **new, obscure, or insecure domains**. Low domain trust score → higher chance of phishing.

Quantum connection: In QML, this could be encoded as **controlled entanglement**: Example: `domain_trust` controls rotation on another qubit representing `email_url_len`. Captures **feature interactions** that classical linear models may not represent well.

3. $q\_sentiment\_interference$ – Contradictory tone in content

How it was added: Use **TextBlob** to calculate:

- `polarity` (positive ↔ negative)
- `subjectivity` (objective ↔ subjective)

Compute the **absolute difference**:

$$\text{sentiment interference} = |\text{polarity} - \text{subjectivity}|$$

Larger values → tone is contradictory or emotionally manipulative.

Why it helps detect phishing: Phishing often mixes **positive/polite wording with urgent requests or threats**. Contradictory tone is a strong phishing signal.

Quantum connection: Analogy to **quantum interference**: Conflicting signals create interference patterns. QML models can naturally exploit interference to detect non-linear patterns across features.

4. *q_url_entropy* – Shannon entropy of URL string

How it was added: Compute **character-level Shannon entropy** of the main URL:

$$H = -\sum_i p_i \log_2 p_i$$

Normalize by dividing by 8 (max entropy per byte). High entropy → unusual/random URLs, possibly obfuscated.

Why it helps detect phishing: Phishing often uses **long, random-looking URLs** to bypass detection or hide real domains. Lower entropy → simpler, more legitimate URLs.

Quantum connection: Can be encoded as **amplitude of a qubit** in QML: High entropy → high amplitude. Qubits can capture **superposition of URL patterns**, helping the model find suspicious structures beyond simple thresholds.

Summary Table of Features

Feature	**How Added**	**Phishing Signal**	**Quantum Analogy**
q_semantic_similarity	Cosine similarity between subject & body TF-IDF vectors	Mismatch between subject and body	Angle encoding in qubit rotations
q_domain_trust	Weighted score of TLD, HTTPS, domain age	Low trust → suspicious domain	Controlled rotation / entanglement
q_sentiment_interference	|polarity - subjectivity|	Contradictory tone → manipulative content	Quantum interference analogue
q_url_entropy	Shannon entropy of URL string	Random/obfuscated URL	Encoded as qubit amplitude

Table 1: Quantum-inspired features for phishing detection

Key Point

These features are **not inherently quantum themselves** — they're classical features inspired by signals that **QML models can naturally exploit using qubit representations, superposition, and entanglement**.

In other words, they form a **quantum feature space** when encoded into qubits, which may help a QML model detect **complex patterns that classical ML could miss**.