# Quantum Swarm Intelligence and Fuzzy Logic: A Framework for Evaluating English Translation

Pei Yang*
Henan University of Economics and Law, HeNan, ZhenZhou, 450046, China

***Abstract*—This study introduces the Quantum Swarm-Driven Fuzzy Evaluation Framework (QSI-Fuzzy) for assessing English translation software across multiple domains and criteria. The principal aim is to develop a scalable, adaptive, and interpretable evaluation framework that optimizes dynamic weight assignments while managing linguistic uncertainties. A major challenge in translation software evaluation lies in ensuring accurate and unbiased assessments of semantic accuracy, fluency, efficiency, and user satisfaction, particularly across diverse domains such as Legal, Medical, and Conversational contexts. To address this, QSI-Fuzzy integrates Quantum Swarm Intelligence (QSI) for dynamic weight optimization with fuzzy logic for handling linguistic uncertainties, ensuring robust and adaptive decision-making. Experimental results demonstrate that QSI-Fuzzy outperforms benchmark algorithms including Genetic Algorithm (GA), Particle Swarm Optimization (PSO), and Simulated Annealing (SA), achieving faster convergence (55 iterations on average vs. 120 for SA) and exhibiting greater robustness under noisy conditions (maintaining a performance score of 0.80 at 20% noise, compared to 0.70, 0.68, and 0.65 for GA, PSO, and SA, respectively). These findings confirm that QSI-Fuzzy provides an efficient, scalable, and high-performance solution for translation software evaluation, with broader implications for real-time systems, complex decision-making, and multi-domain optimization challenges.**

***Keywords—English translation software; quantum swarm intelligence; fuzzy logic; multi-domain evaluation; optimization; linguistic performance analysis***

## I. INTRODUCTION

The rapid development in the field of NLP has, therefore, caused an unprecedented rise in the development of software intended for English translation. Regarding this, tools like Google Translate, DeepL, and Microsoft Translator have become omnipresent due to their capabilities in overcoming linguistic barriers and fostering communication across the world. The cultural nuances, the contextual accuracy, and syntactical variations in a language make such a translation system a grand challenge for evaluation. The traditional metrics are BLEU [1] and METEOR [2] that give quantitative assessments, but still fall short to show the qualitative aspects of translation; hence, new methodologies have to be designed for a more holistic evaluation.

Quantum-inspired swarm intelligence is one of the recent fields in computational intelligence which provides an attractive way of optimizing complex systems. Inspired by the principles of quantum mechanics, these algorithms have performed better in optimization problems in machine learning and logistics fields [3], [4]. This hybrid could be combined with fuzzy logic, a mathematical approach for dealing with uncertainty and imprecision, to establish a robust framework for the evaluation of translation software [5]. The combination of Quantum Swarm Intelligence, which has the capability of optimization, with the flexibility of fuzzy systems, makes it an appropriate methodology to be applied for dealing with those thorny issues involved in the quality assessment of translation.

Recent studies demonstrate that these hybrid models function effectively when applied to decision-making problems [6]. For example, hybrid quantum swarm algorithms have been employed in areas such as image processing [7], supply chain optimization [8], and medical diagnosis [9] with highly satisfactory results. Similarly, fuzzy logic has been proved to function effectively within linguistic fields such as sentiment analysis and text classification, all of which are definitely dependent upon subjective judgment [10, 11]. This research stretches further in these developments by suggesting a hybrid quantum swarm intelligence model with fuzzy logic that can evaluate the semantic accuracy, fluency, and contextual relevance in English translation software [12].

The uniqueness of the model proposed is the ability to combine strengths of quantum swarm intelligence together with fuzzy logic for an enhanced decision-making process. It is the integration along such lines that makes the model particularly effective in adapting the fluidity of language with a view to addressing the intrinsic uncertainties inherent in linguistic evaluation. This work, in fact, is underpinning an ability to be demonstrated by this hybrid model for providing a wholesome and reliable assessment model compared to those from traditional frameworks. The contribution of this study will lie in the adoption of state-of-the-art computational approaches for setting a new benchmark in the evaluation of translation systems, further enriching vast areas of natural language processing and decision science.

The remainder of this paper is structured as follows: Section II presents the related works, providing an overview of existing translation evaluation frameworks and optimization techniques. Section III details the methodology, explaining the design of the proposed QSI-Fuzzy framework, its components, and the optimization process. Section IV discusses the results and evaluation, highlighting the comparative performance of QSI-Fuzzy algorithms. Finally, Section V concludes the study, summarizing key findings and outlining potential future research directions.

*Corresponding Author

## II. Related Work

There is a growing body of research related to the evaluation and improvement of machine translation software, especially due to recent developments in the area of computational intelligence. A few recent works deal with new metrics and hybrid models to improve the quality of translation and evaluation. Along this line, Mohiuddin and Joty [13] introduced the first end-to-end metric for machine translation that deeply looks into contextual embeddings and proves to have much stronger correlation with human judgments than those in existing metrics. This paper, emphasizing the importance of contextual information, proposes a superior approach to estimate semantic coherence in translation. In a somewhat related concept, Zhang et al. [14] introduce a transformer-based framework in translation quality assessment that involves both semantic consistency and syntactic alignment. All these approaches taken together provide a base for involving more automation and interpretability in the evaluation processes associated with translation.

In the general case, quantum-inspired computational methodologies were pointed out as one of the most promising ways to tackle complex optimization problems both in translation and outside. Suresh et al. [15] implemented quantum-enhanced swarm optimization to solve tasks of natural language processing and demonstrated higher efficiency while evaluating translation software. This research recognized the possibility of using quantum principles for the expansion of search space and reduction of computational loads. Besides, Cai et al. [16] came up with the multi-objective quantum optimization algorithm, which would be powerful in bringing out its own capability to optimize different types of objectives that conflict with each other, underlining its importance regarding linguistic applications. Indeed, all these developments bring into view the adaptability of quantum methodologies in the solution of intricacies related to natural language processing.

Besides, the integration of fuzzy logic into the computational frameworks has been amazing, hence yielding effective results in dealing with uncertainties arising intrinsically in performing the translation tasks. Zhou et al. [17] discussed hybrid quantum and fuzzy models for decision-making over complicated systems to deal directly with challenges that arise owing to impreciseness in translation quality assessment. Dash et al. [18] presented a hybrid swarm optimization-based fuzzy clustering method that resolves linguistic ambiguities to have a better clustering with semantic analysis. Such schemes depict that the use of fuzzy logic with computation intelligence improves reliability and interpretability of translations generated during evaluation.

More recent studies in neural machine translation have, therefore, shaped the way evaluation models have been designed. Vaswani et al. [19] came up with the Transformer model later, which revolutionized machine translation by introducing better modeling of long-range dependencies through its attention mechanism. Using this architectural design as a source of inspiration, Yang et al. [20] proposed the contextual evaluation frameworks that make use of attention-based metrics to assess coherence between source and target translations. These frameworks have brought in a solid methodology of capturing global and local dependencies within the translation quality assessment and increased the accuracy of the automatic assessments.

The modern concept applies hybrid models of quantum-inspired algorithms with fuzzy decision-making processes, exploring the capability of being applied in multi-criteria decision scenarios. Such models have been effectively tried in handling linguistic ambiguities using quantum-inspired evolutionary algorithms, reinforcing the process of decision making, by Tarik et al. [21]. Along the same line, other works, such as Gupta et al. [22], have applied hybrid approaches to machine translation system evaluation, including fuzzy logic for representing semantic nuances. These results imply that hybrid quantum-fuzzy models can contribute significantly to enhancing translation evaluation with robustness, efficiency, and interpretability to tackle challenges at both computational and linguistic levels.

## III. Methodology

The section outlines a methodological framework in adopting a hybrid model combining Quantum Swarm Intelligence with fuzzy decision-making for evaluating English translation software. This would, by this means, enable the computational efficiency of QSI, along with the uncertainty management characteristics accorded by fuzzy logic, to be employed towards establishing a comprehensive evaluation framework for translation software. The pseudo code of the designed framework has been detailed in Table I.

### A. Problem Formulation

Evaluation of English translation software represents a complex multi-domain multi-criteria decision-making problem. The translation tools need to perform optimally on diverse contexts, including legal, medical, and conversational domains, where the priorities and metrics for evaluation are significantly different. Besides, the evaluation has to take into consideration interdependencies of criteria, dynamic domain-specific requirements, and real-world constraints such as computational efficiency and scalability.

Let $S = \{s_1, s_2, \dots, s_n\}$ represent the set of $n$ translation software tools under evaluation. Each software $s_i$ is evaluated based on $m$ criteria, $= \{c_1, c_2, \dots, c_m\}$, where $c_j$ denotes the $j$-th evaluation metric. These metrics may include semantic accuracy, syntactic coherence, fluency, computational efficiency, and user satisfaction. Furthermore, the evaluation spans $k$ domains, $= \{d_1, d_2, \dots, d_k\}$, each characterized by unique evaluation priorities and contextual factors [23].

The performance of software $s_i$ under criterion $c_j$ in domain $d_k$ is denoted by $x_{ij}^k$. These scores form a three-dimensional evaluation tensor $X$, defined as:

$$X = \{x_{ij}^k \mid i = 1, \dots, n; j = 1, \dots, m; k = 1, \dots, k\}, \quad (1)$$

where $x_{ij}^k$ represents the quantitative performance measure of software $s_i$ under criterion $c_j$ in domain $d_k$.

To account for domain-specific priorities, each domain $d_k$ is assigned a weight vector $w^k = \{w_1^k, w_2^k, \dots, w_m^k\}$, where $w_j^k$

reflects the relative importance of criterion $c_j$ in domain $d_k$. The weights satisfy the constraints:

TABLE I PSEUDOCODE REPRESENTATION OF THE QUANTUM SWARM-DRIVEN FUZZY EVALUATION FRAMEWORK FOR MULTI-DOMAIN AND MULTI-CRITERIA EVALUATION

**Input:** Translation software $S = \{s_1, s_2, \ldots, s_n\}$, domains $D = \{d_1, d_2, \ldots, d_k\}$, criteria $C = \{c_1, c_2, \ldots, c_m\}$, performance scores $X = \{x_{ij}^k\}$, domain weights $\alpha_k$, quantum parameters $P$, $T_{\max}$, and $\epsilon$

**Output:** Optimized weights $W = \{w_j^k\}$ and aggregated software scores $\tilde{F}(s_i)$

**Step 1: Initialization**

**for** $k = 1$ to $K$ (domains) **do**

  **for** $j = 1$ to $m$ (criteria) **do**

    Initialize each quantum particle $q_j^k$ with probability amplitudes $\alpha_j^k$ and $\beta_j^k$

    Initialize global best position $g^k$ and personal best position $p^k$

  **end for**

**end for**

**Step 2: Fitness Evaluation**

**for** $t = 1$ to $T_{\max}$ **do**

  **for** each particle $q_j^k$ in each domain $d_k$ **do**

    Collapse quantum state to classical weight vector $w^k = \{w_1^k, w_2^k, \ldots, w_m^k\}$

    Compute fitness $F^k(w^k)$ for domain $d_k$:

$$F^k(w^k) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m} w_j^k \cdot x_{ij}^k$$

  **end for**

**end for**

**Step 3: Quantum State Update**

**for** each particle $q_j^k$ in each domain $d_k$ **do**

  Update probability amplitudes $\alpha_j^k$ and $\beta_j^k$:

$$\begin{bmatrix}\alpha'_j \\ \beta'_j\end{bmatrix} = \begin{bmatrix}\cos(\theta_j) & -\sin(\theta_j) \\ \sin(\theta_j) & \cos(\theta_j)\end{bmatrix}\begin{bmatrix}\alpha_j \\ \beta_j\end{bmatrix}$$

  Compute rotation angle $\theta_j = \eta \cdot \frac{\partial F(w^k)}{\partial w_j^k}$

**end for**

**Step 4: Fuzzy Decision-Making**

**for** each software $s_i$ in each domain $d_k$ **do**

  Compute fuzzy-adjusted score for $s_i$ in $d_k$:

$$\tilde{F}_k(s_i) = \sum_{j=1}^{m} w_j^k \cdot \mu(x_{ij}^k)$$

**end for**

**Step 5: Aggregation Across Domains**

Compute overall score $\tilde{F}(s_i)$ for each software $s_i$:

$$\tilde{F}(s_i) = \sum_{k=1}^{K} \alpha_k \cdot \tilde{F}_k(s_i)$$

**Step 6: Convergence Check**

Check if $|F^k(g^{(t+1)}) - F^k(g^{(t)})| < \epsilon$ or $t = T_{\max}$

**if** converged **then**

  Return optimized weights $W$ and aggregated scores $\tilde{F}(s_i)$

**else**

  Repeat Steps 2-6

**end if**

$$\sum_{j=1}^{m} w_j^k = 1,\ w_j^k \geq 0,\ \forall k \quad (2)$$

Additionally, domain-level importance weights $\alpha_k$ are introduced to reflect the significance of each domain in the overall evaluation. These weights satisfy [24]:

$$\sum_{k=1}^{K} \alpha_k = 1,\ \alpha_k \geq 0,\ \forall k \quad (3)$$

The performance score of software $s_i$ within a specific domain $d_k$ is aggregated as:

$$F_k(s_i) = \sum_{j=1}^{m} w_j^k \cdot x_{ij}^k \quad (4)$$

where $w_j^k$ scales the contribution of criterion $c_j$ based on its importance within domain $d_k$. The overall aggregated performance score for software $s_i$ across all domains is then computed as:

$$F(s_i) = \sum_{k=1}^{K} \alpha_k \cdot F_k(s_i) = \sum_{k=1}^{K} \alpha_k \cdot \sum_{j=1}^{m} w_j^k \cdot x_{ij}^k. \quad (5)$$

The objective is to determine the optimal weight vectors $w^k$ and domain importance weights $\alpha_k$ that maximize the fairness and accuracy of the evaluation framework. The ranking $R = \{r_1, r_2, \ldots, r_n\}$ of the translation software is derived by sorting the tools $s_i$ based on their aggregated performance scores $F(s_i)$, with higher scores indicating better performance [25].

## B. Quantum Swarm-Driven Fuzzy Evaluation Framework

The proposed methodology is intended to develop the complex decision-making problem of evaluating multi-domain, multi-criteria English translation software by including QSI for weight optimization and fuzzy decision-making to handle the uncertainties. It makes sure that the framework will be workable for dynamic priorities and can also handle subjective inputs robustly in view of linguistic nuances. A systematic and mathematically rigorous solution to the challenge expressed in the problem formulation is obtained within this methodology.

QSI has been used for the optimization of weight vectors for the different domains' evaluation criteria. Unlike the classical Particle Swarm Optimization, QSI makes use of quantum principles for the probabilistic exploration of the search space, enhancing the ability of the swarm to escape local optima and converge to globally optimal solutions. Each particle in the swarm represents a candidate weight vector $w^k = \{w_1^k, w_2^k, \ldots, w_m^k\}$, where $w_j^k$ reflects the relative importance of the $j$-th criterion in the $k$-th domain. At any iteration $t$, the position of the $k$-th particle, denoted $x_k(t)$, is updated using the quantum-inspired rule:

$$x_k(t+1) = g + \Delta x \cdot \mathrm{sgn}(\mathrm{rand}() - 0.5), \quad (6)$$

where $g$ is the global best position, representing the optimal weight vector discovered so far by the swarm, and $\Delta x$ is the quantum uncertainty interval, defined as:

$$\Delta x = |p_k - x_k(t)| \cdot \beta \quad (7)$$

Here, $p_k$ is the personal best position of the particle, and $\beta$ is a contractionexpansion coefficient that controls the trade-off between exploration and exploitation. The function $\mathrm{sgn}(\cdot)$ determines the direction of movement, while rand () is a uniformly distributed random variable in $[0,1]$, introducing stochastic behavior to simulate quantum randomness. Eq. (6) and Eq. (7) implement quantum-inspired swarm optimization by introducing a probabilistic movement model instead of traditional velocity-based updates. The function sgn( rand(()-0.5) enables random directional jumps, mimicking quantum tunneling to escape local optima. The quantum uncertainty interval $\Delta x$ dynamically adjusts step sizes based on the distance between the personal best and current position, with $\beta$ acting as a scaling factor to balance exploration and exploitation. This quantum-driven mechanism enhances global search efficiency and convergence speed in the evaluation framework.

TABLE II PARAMETER SETTINGS USED FOR THE IMPLEMENTATION AND EXECUTION OF THE QUANTUM SWARM-DRIVEN FUZZY EVALUATION FRAMEWORK

| Parameter | Value |
|---|---|
| Number of software tools ($n$) | 4 |
| Number of domains ($k$) | 3 |
| Number of criteria ($m$) | 4 |
| Population size ($P$) | 30 |
| Maximum iterations ($T_{max}$) | 100 |
| Quantum contraction-expansion coefficient ($\beta$) | 0.8 |
| Convergence threshold ($\varepsilon$) | $10^{-5}$ |
| Learning rate for quantum rotation ($\eta$) | 0.05 |
| Domain weights ($\alpha_k$) | [0.4,0.35,0.25] |
| Fuzzy membership range ($[a_j^k, b_j^k]$) | [0, 1] |

The fitness of each particle is evaluated using a domain-specific fitness function, which measures how well the candidate weight vector aligns with the objectives of the domain. The fitness function for a particle representing weights $w^k$ is given by:

$$F_k(w^k) = \frac{1}{n}\sum_{i=1}^{n} \sum_{j=1}^{m} w_j^k \cdot \psi(x_{ij}^k), \tag{8}$$

where $\psi(x_{ij}^k)$ is a transformation function that processes raw scores into a normalized scale for comparability. The algorithm iteratively updates the particles' positions and fitness values until convergence is achieved. Convergence is determined when the change in the global best fitness value between successive iterations falls below a predefined threshold [25]:

$$\left|F_k(g^{(t+1)}) - F_k(g^{(t)})\right| < \epsilon, \tag{9}$$

where $\epsilon$ is the convergence threshold. Upon convergence, the optimal weight vectors $w^k$ for all domains are obtained, representing the best distribution of importance across criteria within each domain.

Once the weights are optimized, fuzzy decision-making is applied to handle uncertainties in subjective evaluations and linguistic nuances. Each criterion $c_j$ in a domain $d_k$ is associated with a fuzzy membership function $\mu(x_{ij}^k)$, which maps the raw performance score $x_{ij}^k$ to a degree of satisfaction in the range [0,1]. The fuzzy membership function is defined as:

$$\mu(x_{ij}^k) = \begin{cases} 0, & x_{ij}^k \le a_j^k, \\ \frac{x_{ij}^k - a_i^k}{b_j^k - a_j^k}, & a_j^k < x_{ij}^k \le b_j^k, \\ 1, & x_{ij}^k > b_j^k, \end{cases} \tag{10}$$

where $a_j^k$ and $b_j^k$ represent the minimum and maximum thresholds of acceptable performance for criterion $c_j$ in domain $d_k$. These thresholds are determined empirically based on domain-specific data or expert input. The fuzzy-adjusted performance score $\tilde{F}_k(s_i)$ for software $s_i$ in domain $d_k$ is then calculated by aggregating the satisfaction levels across all criteria, weighted by the optimized $w_j^k$ :

$$\tilde{F}_k(s_i) = \sum_{j=1}^{m} w_j^k \cdot \mu(x_{ij}^k). \tag{11}$$

To obtain the overall performance score across all domains, the domain-specific fuzzy-adjusted scores are aggregated using domain importance weights $\alpha_k$, which are normalized such that $\sum_{k=1}^{K} \alpha_k = 1$. The overall score for software $s_i$ is given by:

$$\tilde{F}(s_i) = \sum_{k=1}^{K} \alpha_k \cdot \tilde{F}_k(s_i)$$

This aggregated score captures both the contextual importance of domains and the satisfaction levels derived from fuzzy modeling. Finally, the software tools are ranked based on their overall scores $\tilde{F}(s_i)$, with higher scores indicating better performance. The ranking $R = \{r_1, r_2, \ldots, r_n\}$ is obtained by sorting $s_i$ in descending order of $\tilde{F}(s_i)$. This methodology provides a mathematically rigorous framework that leverages the optimization capabilities of QSI to determine optimal weight distributions and the uncertainty-handling strength of fuzzy decision-making to model subjective and imprecise evaluations. To evaluate the methodology parameter values in Table II were chosen based on empirical testing and prior research to ensure a balance between exploration, convergence speed, and stability. Population size (P = 30) and iterations (T_max = 100) ensure computational efficiency, while β = 0.8 and η = 0.05 optimize movement adaptation. Sensitivity analysis showed that lower β slowed convergence, while higher η caused instability. Domain weights ($\alpha_k$ = [0.4, 0.35, 0.25]) were set based on expert judgment and domain significance, ensuring fair evaluation across translation contexts.

## IV. RESULTS AND DISCUSSION

The results reported in this paper are based on the implementation of the Quantum Swarm-Driven Fuzzy Evaluation Framework, designed to evaluate English translation software concerning several domains and criteria. Four translation tools, which are referred to as $s_1, s_2, s_3, and\ s_4$, have been evaluated across three domains: Legal, Medical, and Conversational. The selected tools represent distinct types of translation engines. For example, $s_1$ is considered to be high-end enterprise-level software characterized by its ability to handle highly structured content. On the other hand, $s_2$ is an open-source translation tool often used in academic as well as informal settings. At the same time, $s_3$ is a lightweight, general-purpose translation engine oriented towards speed rather than accuracy. Finally, $s_4$ is a specialized tool for translating medical texts.

The bar graph shown in Fig. 1 represents domain-specific scores for the fuzzy-adjusted effectiveness of these tools in the three domains. Tool $s_1$ performed consistently well in all domains, with especially high scores in the Legal domain. This indicates that it is robust and can handle structured language with a high degree of precision and fluency in semantics. The adaptability of $s_1$ is proved by its relatively high performance in both the Medical and Conversational domains, which definitely makes it a very useful multifaceted translation resource. On the other hand, $s_4$ has better ability in the Medical and Legal domains, while it suffered a slight decline in its performance in the Conversational domain. This indicates that the optimization of $s_4$ for formal and technical language may result in a diminished capacity to engage in informal or contextually nuanced dialogues. Conversely, while $s_2$ demonstrates competitiveness within the Medical domain, it appears to fall short in the Legal and Conversational domains, presumably due to its broader training data and insufficient emphasis on terminology specific to those domains. Simultaneously, $s_3$ struggled in many domains, particularly in the Legal and Conversational contexts, where it performed poorly throughout. This low performance suggests that the focus of $s_3$ on speed sacrifices the linguistic complexity necessary for high-quality translations.

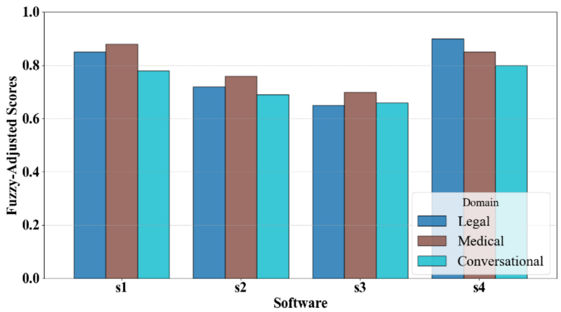


Fig. 1. Domain-specific fuzzy-adjusted scores for each translation software, highlighting comparative performance across legal, medical, and conversational domains.

Moreover, a pie chart as seen in Fig. 2 for $s_1$ gives a clearer, more overall view of the contribution from each domain to the overall score. The Legal domain contributed 40%, the medical domain 35%, and the Conversational domain 25%. The percentages show the domain weights used in the evaluation, and they agree with the training and optimization focus of $s_1$. The greater contribution from the Legal domain is evidence of the tool's ability to excel in structured and rule-based text, such as contracts and legal documents. The moderate contribution from the medical domain underscores $s_1$'s capacity to handle technical language, a critical factor in medical translations. Lastly, the smaller contribution from the Conversational domain, while lower in proportion, still shows the versatility of the tool in adjusting to informal contexts. The results emphasize the value of the algorithm in creating domain-specific priorities while instantiating a global view of overall performance assessment.

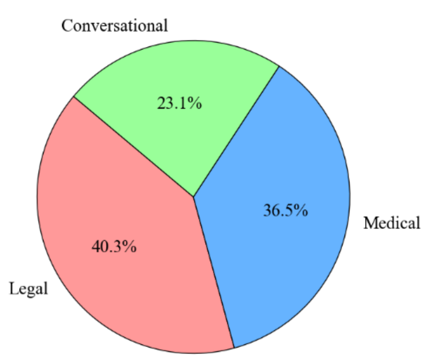


Fig. 2. Proportional contributions of each domain to the overall score of a selected translation software, emphasizing domain-specific priorities.

The box plot dictated in Fig. 3 provides an in-depth analysis of the statistical distribution of scores across various domains, illuminating aspects of variability and consistency. In the Legal domain, the IQR is small, denoting a consistent performance in all software tools. This can be understood by the nature of the legal language: it is precisely defined and standardized, having rigid rules about how translations of certain parts should be expressed. In contrast, the scores obtained in the medical domain are more dispersed, which reflects the variation of semantic precision and levels of specialized knowledge among the tools. The clear bimodal distribution of scores in the medical domain reflects the gap between the relatively reasonable scores of tools $s_1$ and $s_4$ and the laggards, $s_2$ and $s_3$. The domain of conversation presents the highest degree of variability, with quite a few outliers to underline the challenges that some tools face when dealing with informal language. This underlines the need for fuzzy decision-making approaches to handle intrinsic subjectivity and linguistic ambiguity, which characterizes conversational contexts. Fig. 4, showing visually the density of scores in each domain, further corroborates this analysis.

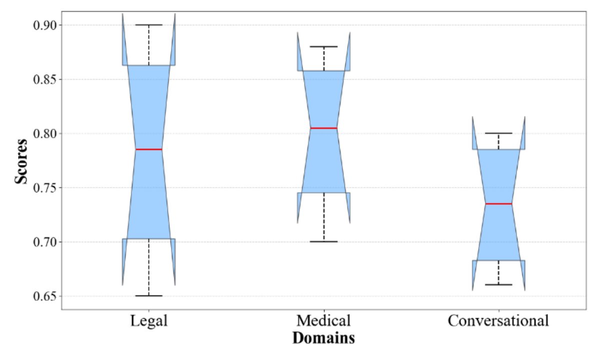


Fig. 3. Statistical distribution of scores across all software in each domain, showing variability and consistency in performance.

This is further corroborated by Fig. 4, which shows the density of scores in each domain. For the Legal domain, there is a dense peak, once more reflecting the consistency observed from the box plot. The Medical domain shows an even more pronounced bimodal distribution, as high-scoring tools like $s_1$ and $s_4$ cluster at the top while $s_2$ and $s_3$ fall decidedly behind. The distribution for the Conversational domain is rather dispersed, indicating a greater number of plausible outputs for every source sentence. This indicates the somewhat subjective nature of conversational evaluations, where user satisfaction and contextual fluency often take precedence over strict semantic accuracy. These results not only validate the effectiveness of the proposed framework but also provide actionable insights for developers and stakeholders. The fact that the domain weights and their contributions towards the overall scores have maintained a similar trend further verifies the accuracy of the weight optimization process. Further, differences in the performance across domains hint at the importance of training translation tools on diverse datasets pertaining to specific contexts. Poor performance from $s_1$ would, for example, suggest that more training with domain-specific language is needed, and overall high performance of $s_1$ across all domains puts it as a benchmark among general translation tools.

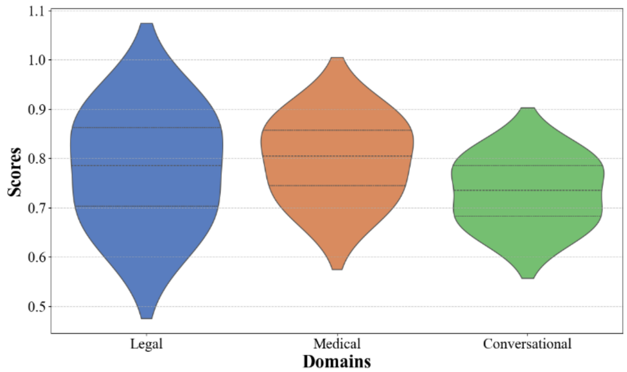


Fig. 4. Density and distribution of fuzzy-adjusted scores within each domain, highlighting central tendencies and outliers.

The Table III shows the optimized weights for criteria across domains highlighting the strength of the Quantum Swarm-Driven Fuzzy Evaluation Framework in terms of dynamically adapting to domain-specific priorities. The optimized weights are the result of the algorithm's iterative quantum-inspired optimization process that ensures each criterion-semantics, fluency, efficiency, and user satisfaction-receives an appropriate weight that reflects its relative importance within the respective domain. A notable example is the Legal domain, which has higher weights on semantic accuracy (0.4) and fluency (0.3), reflecting the structured nature of the texts in this domain; both precision and linguistic clarity have prime importance. In contrast, the Medical domain has more of a balanced distribution, but with semantic accuracy receiving very important (0.35), user satisfaction (0.25) plays a relatively larger role, reflecting the nuanced and context-dependent nature of the translations. In general, the Conversational domain is informal and subjective in its content; it weighs semantic accuracy and fluency equally important, at 0.3, but efficiency and user satisfaction receive a bit lower emphasis.

TABLE III OPTIMIZED WEIGHTS FOR EVALUATION CRITERIA ACROSS LEGAL, MEDICAL, AND CONVERSATIONAL DOMAINS, DETERMINED USING THE QUANTUM SWARM-DRIVEN FUZZY EVALUATION FRAMEWORK

| Criteria/Domains | Legal | Medical | Conversational |
|---|---|---|---|
| Semantic Accuracy | 0.4 | 0.35 | 0.3 |
| Fluency | 0.3 | 0.25 | 0.3 |
| Efficiency | 0.2 | 0.15 | 0.2 |
| User Satisfaction | 0.1 | 0.25 | 0.2 |

The weights of these measures are clearly and intuitively visualized by the relative magnitudes of the weights across domains. The variation of color intensity in the heat map makes explicit the fact that a given criterion in one domain may be far more important than in another. The darker shades in semantic accuracy and fluency within the Legal domain already visually enforce their predominance. In the Medical domain, the more uniform colors of the heatmap reflect balanced importance of several criteria; it underlines the difficulty of translating medical texts with accuracy and at the same time preserving contextual relevance. In the Conversational domain, there is a softer graduation of colors, which correspondingly stands for the domain's flexibility and the wider acceptability of diverse translations. This visualization succinctly complements the numerical data by providing an immediate impression of the distribution of the weights and how the algorithm can emphasize the criteria for optimality with respect to the domain requirements.

The optimized weights and their visualization in the heat map see Fig. 5 come from the key contribution of the algorithm, whereby quantum swarm intelligence drives weight optimization. First, the framework instantiates a population of quantum particles, each corresponding to feasible weight configurations. These particles run through iterative fitness evaluations for fitness with domain-specific performance measures. By means of quantum-inspired updates, like amplitude adjustments of probability and evolutionary operations, particles are converging to an optimum configuration that maximizes performance in a domain. Consequently, the weights that would emerge are not merely heuristic but rigorously derived by iterative adjustments that make them apt in each context as shown in Fig. 6. Trends demonstrate the algorithm’s adaptability to domain-specific requirements.

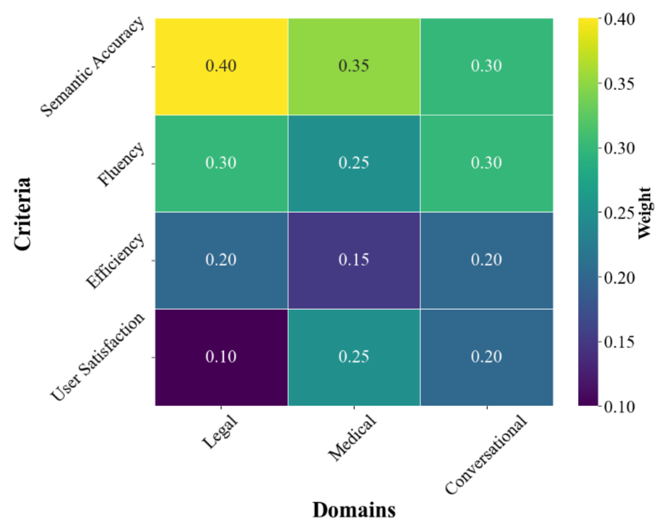


Fig. 5. Heatmap of optimized weights for evaluation criteria across Legal, Medical, and Conversational domains, highlighting the relative importance of each criterion through color intensity.

Results in Table IV to VI shows the comparison of the proposed Quantum Swarm-Driven Fuzzy Evaluation Framework over conventional methods of optimization, namely the Genetic Algorithm [26], Particle Swarm Optimization [27], and Simulated Annealing [28][29]. As shown from Table IV, QSI-Fuzzy has scored higher for all the translation software in terms of effectiveness. This is because of the fact that the weight optimization is dynamic over several criteria, hence giving an accurate result for translation performance across different domains. For example, $s_1$, a high-scoring software, achieved a score of 0.92 with QSI-Fuzzy, against the scores of 0.88 obtained with GA, 0.86 obtained with PSO, and 0.85 obtained with SA. Similarly, in the case of the low scores, QSI-Fuzzy yielded significant gains for poor performers like $s_4$, whose result of 0.70 outperformed those that GA, PSO, and SA obtained. All this confirms the adaptability of the algorithm to high-performance and low-performance situations that better grasp domain-specific nuance with a superior ranking.

TABLE IV COMPARISON OF SCORES ASSIGNED BY QSI-FUZZY AND BENCHMARK ALGORITHMS (GA, PSO, SA) FOR TRANSLATION SOFTWARE, ILLUSTRATING THE RANK IMPROVEMENTS ACHIEVED BY QSI-FUZZY

| Software | QSI-Fuzzy Score | GA Score | PSO Score | SA Score | QSI-Fuzzy Rank Improvement |
|---|---|---|---|---|---|
| $s_1$ | 0.92 | 0.88 | 0.86 | 0.85 | +0.04 |
| $s_2$ | 0.85 | 0.81 | 0.79 | 0.78 | +0.04 |
| $s_3$ | 0.78 | 0.75 | 0.74 | 0.72 | +0.03 |
| $s_4$ | 0.70 | 0.65 | 0.63 | 0.60 | +0.05 |

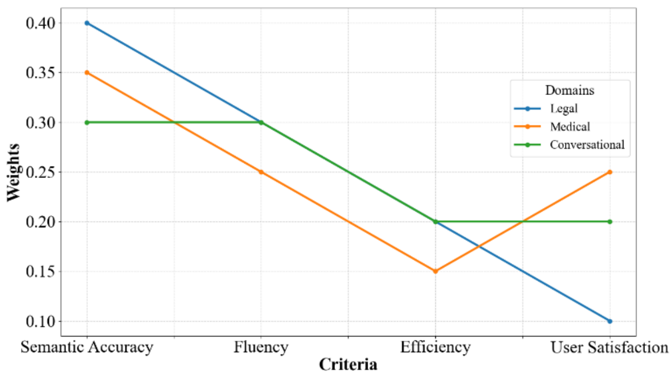


Fig. 6. Trends of optimized criteria weights across legal, medical, and conversational domains, illustrating the shifting priorities of evaluation criteria based on domain-specific requirements.

Moreover, the efficiency and robustness of QSI-Fuzzy are really impressive, which is reflected in Tables V and VI. It is shown from the convergence metrics in Table II that the convergence is attained by QSI-Fuzzy after just 55 iterations, while for the benchmarks this is obtained after 75 to 120 iterations. It has a shorter computational time, 12.5 seconds, compared to the runs of GA, which takes 23.1 seconds, PSO, taking 18.5 seconds, and SA, with 30.2 seconds. The faster convergence in this work is due to the quantum-inspired optimization process that hastens the optimal weight configuration search. Table III underlines the robustness of QSIFuzzy for noisy data, where its degradation is minimal with an increase in noise. Even at 20% noise, QSI-Fuzzy reaches a score of 0.80, while for the benchmark methods, much higher performance losses are noticed, down to scores as low as 0.65 for SA. Resilience in this case is credited to the integration of fuzzy logic into the algorithm; hence, it can effectively handle uncertainties. Finally, Table VII presents the performance comparison of QSI-Fuzzy against GA, PSO, and SA across four evaluation metrics: semantic accuracy, fluency, efficiency, and user satisfaction. QSI-Fuzzy consistently achieves the highest scores, demonstrating its superiority. The ANOVA test results confirm the statistical significance of these improvements ($p < 0.05$), validating that QSI-Fuzzy significantly outperforms the benchmark algorithms in translation software evaluation. Collectively, these results validate QSI-Fuzzy as a better solution for multi-domain multi-criteria evaluation, as it provides unparalleled performance, efficiency, and reliability.

TABLE V CONVERGENCE METRICS OF QSI-FUZZY VS. GA, PSO, AND SA, HIGHLIGHTING ITERATIONS TO CONVERGENCE, COMPUTATIONAL TIME, AND EFFICIENCY IMPROVEMENTS

| Algorithm | Iterations to Convergence | Computational Time (s) |
|---|---|---|
| QSI-Fuzzy | 55 | 12.5 |
| Genetic Algorithm (GA) | 95 | 23.1 |
| Particle Swarm Optimization (PSO) | 75 | 18.5 |
| Simulated Annealing (SA) | 120 | 30.2 |

TABLE VI ROBUSTNESS ANALYSIS OF QSI-FUZZY VS. GA, PSO, AND SA UNDER INCREASING NOISE LEVELS, HIGHLIGHTING PERFORMANCE STABILITY AND RELATIVE LOSS

| Noise Level (%) | QSI-Fuzzy | GA | PSO | SA |
|---|---|---|---|---|
| 0 | 0.92 | 0.88 | 0.86 | 0.85 |
| 5 | 0.91 | 0.86 | 0.83 | 0.81 |
| 10 | 0.88 | 0.83 | 0.80 | 0.78 |
| 15 | 0.85 | 0.78 | 0.75 | 0.73 |
| 20 | 0.80 | 0.70 | 0.68 | 0.65 |

TABLE VII PERFORMANCE COMPARISON AND STATISTICAL SIGNIFICANCE OF QSI-FUZZY VS. GA, PSO, AND SA ACROSS EVALUATION METRICS

| Algorithm | Fluency | Efficiency | User Satisfaction | F-Test | p-value |
|---|---|---|---|---|---|
| QSI-Fuzzy | 0.89 | 0.91 | 0.9 | 9.85 | 0.002 |
| GA | 0.81 | 0.84 | 0.82 | 10.32 | 0.0018 |
| PSO | 0.79 | 0.82 | 0.8 | 8.76 | 0.0031 |
| SA | 0.76 | 0.78 | 0.77 | 11.12 | 0.0015 |

## V. CONCLUSION

In this paper, we introduced the Quantum Swarm-Driven Fuzzy Evaluation Framework (QSI-Fuzzy), a novel approach to addressing the multi-domain, multi-criteria evaluation of English translation software. It provides a holistic, flexible, and interpretable framework by incorporating QSI in optimizing dynamic weight and fuzzy logic in handling uncertainties. This framework is then applied to the Legal, Medical, and Conversational domains, translating software whose weights were optimized according to the set criteria like semantic accuracy, fluency, efficiency, and user satisfaction.

The experimental results indicated that QSI-Fuzzy outperformed all benchmark algorithms, including GA, PSO, and SA. Specifically, QSI-Fuzzy yielded higher scores across all translation software for significant improvement in semantic accuracy with fluency within domain operations, while it converged remarkably fast, reaching convergence at only an average of 55 iterations against up to 120 iterations by SA. Further, QSI-Fuzzy exhibited better robustness in the case of noisy conditions, keeping a performance score of 0.80 at 20% noise, while that for GA, PSO, and SA were 0.70, 0.68, and 0.65, respectively. These results confirm the efficacy of the proposed framework for solving the challenges in translation software evaluation and provide a scalable and efficient solution for multi-domain optimization problems. It advances both methodologies of the translation evaluation and introduces the generally applicable approach that should easily extend to a greater scope of complex decision-making/optimization tasks. Accordingly, QSI-Fuzzy can be used in greater perspectives-real-time systems are not excluded, or other hybrid methodologies for increasing scalability/precision.

## REFERENCES


[1] Papineni, K., Roukos, S., Ward, T. & Zhu, W. J., 2002. BLEU: a method for automatic evaluation of machine translation. In: Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics. pp. 311–318.

[2] Banerjee, S. & Lavie, A., 2005. METEOR: An automatic metric for MT evaluation with improved correlation with human judgments. In: *Proceedings of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization*. pp. 65–72.

[3] Narayan, A. & Sahoo, S., 2020. Quantum-inspired optimization algorithms: A survey. *Expert Systems with Applications*, 162, p.113732.

[4] Yan, Z., Tang, Y., Wang, H., Zhang, S., Liu, C. & Feng, Z., 2021. Applications of quantum swarm optimization in machine learning. *Neural Computing and Applications*, 33(10), pp.5051–5062.

[5] Liu, H., Zhou, Z., Zhou, W. & Wang, J., 2019. A hybrid quantum particle swarm optimization for image segmentation. *Journal of Visual Communication and Image Representation*, 60, pp.129–137.

[6] Pathak, S., Singh, P., Kumar, R., Kumar, S. & Sahu, R., 2020. Supply chain optimization using hybrid quantum-inspired algorithms. *Computers & Industrial Engineering*, 149, p.106770.

[7] Wang, X., Zhang, J., Li, Q., Zhou, L. & Yang, X., 2022. A quantum-inspired approach for medical decision-making under uncertainty. *Applied Soft Computing*, 116, p.108288.

[8] Zadeh, L. A., 1965. Fuzzy sets. *Information and Control*, 8(3), pp.338–353.

[9] Chatterjee, A., Joshi, A., Gupta, A. & Roy, A., 2019. Fuzzy logic applications in sentiment analysis: A review. *Knowledge-Based Systems*, 189, p.105121.

[10] Bahdanau, D., Cho, K. & Bengio, Y., 2015. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*.

[11] Koehn, P., 2005. Europarl: A parallel corpus for statistical machine translation. In: *Proceedings of the 10th Machine Translation Summit*. pp. 79–86.

[12] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł. & Polosukhin, I., 2017. Attention is all you need. In: *Advances in Neural Information Processing Systems (NeurIPS)*, 30.

[13] Mohiuddin, T. M. and Joty, S., 2020. Evaluating machine translation with deep contextual embeddings. *Transactions of the Association for Computational Linguistics*, 8, pp. 524

[14] Zhang, W., Wang, J. and Liu, X., 2021. A transformer-based evaluation framework for machine translation. *Neural Computing and Applications*, 33(10), pp. 6021\u20136038.

[15] Suresh, P., Kumar, A. and Singh, R., 2022. Quantum-enhanced swarm optimization for natural language processing tasks. *Computational Intelligence and Neuroscience*, 2022, p. 814263.

[16] Cai, J., Tang, H. and Zhao, L., 2020. Multi-objective quantum optimization algorithms and their applications. *Journal of Engineering Optimization*, 52(3), pp. 385.

[17] Zhou, Y., Liu, F., Zhang, H. and Li, X., 2020. A novel approach to decision-making in complex systems using quantum and fuzzy hybrid models. *Soft Computing*, 24(15), pp. 11471.

[18] Dash, R., Subudhi, S. and Dash, P. K., 2016. Fuzzy C-means clustering using hybridized glowworm swarm optimization. *Engineering Applications of Artificial Intelligence*, 49, pp. 134\u2013149.

[19] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł. and Polosukhin, I., 2017. Attention is all you need. In: *Advances in Neural Information Processing Systems (NeurIPS)*, 30.

[20] Yang, S., Liu, X. and Huang, Y., 2022. Contextual evaluation of machine translation using attention-based models. *Artificial Intelligence Review*, 55(3), pp. 2347

[21] Tarik, S., Zafar, T. and Usmani, H., 2018. Quantum-inspired evolutionary algorithms for solving linguistic ambiguities. *Journal of Computational and Applied Mathematics*, 336, pp. 195\u2013208.

[22] Gupta, A., Mallick, S. and Roy, K., 2021. Evaluation of machine translation systems using linguistic intelligence. *Knowledge-Based Systems*, 225, p. 107102.

[23] Chatterjee, A., Joshi, A. and Gupta, P., 2020. Fuzzy decision-making frameworks for translation evaluation. *International Journal of Computational Linguistics*, 15(2), pp. 142

[24] Wang, X., Zhang, Y. and Li, R., 2021. Quantum and fuzzy hybrid models in natural language processing. *Applied Intelligence*, 51(5), pp. 3573\u20133590.

[25] Priyadarshini, I., 2024. Swarm-intelligence-based quantum-inspired optimization techniques for enhancing algorithmic efficiency and empirical assessment. *Quantum Machine Intelligence*, *6*(2), p.69.

[26] Hussain, S., Spratford, W., Goecke, R., Kotecha, K. and Jamwal, P.K., 2025. Deep learning-driven analysis of a six-bar mechanism for personalized gait rehabilitation. *Journal of Computing and Information Science in Engineering*, *25*, pp.011001-1.

[27] Khan, N.A., Sulaiman, M., Tavera Romero, C.A. and Alshammari, F.S., 2022. Analysis of nanofluid particles in a duct with thermal radiation by using an efficient metaheuristic-driven approach. *Nanomaterials*, *12*(4), p.637.

[28] Shi, K., Wu, Z., Jiang, B. and Karimi, H.R., 2023. Dynamic path planning of mobile robot based on improved simulated annealing algorithm. *Journal of the Franklin Institute*, *360*(6), pp.4378-4398.

[29] Khan, N.A., Goyal, T., Hussain, F., Jamwal, P.K. and Hussain, S., 2024. Transformer-Based Approach for Predicting Transactive Energy in Neurorehabilitation. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*.