A Fuzzy Attribution Based LoRA Fine-Tuning for Vision-Language Models within Telecom Inspections

Mazen Ahmed, Hani Hagras
The Artificial Intelligence Group
School of Computer Science and Electronic Engineering
University of Essex
Colchester, UK
mm24829@essex.ac.uk

Hugo Leon-Garza, Anasol Pena-Rios
Business Modelling and Operational Transformation Practice
British Telecom
Ipswich, UK
anasol.penarios@bt.com

Abstract— **Automating the inspection of telecommunication Customer Service Point (CSP) boxes remains a challenge due to small-scale components, clutter and the subtle nature of installation errors. This paper introduces an explainable hybrid framework that leverages parameter-efficient fine-tuning of a Vision-Language Model (VLM) and a fuzzy attribution layer for interpretable audit decision-making. The proposed system adapts LLaMA 3.2 Vision-Instruct (11 B) model using Low-Rank Adaptation (LoRA) with 4-bit quantization to achieve domain-specific auditing of completeness and correctness across 37 controlled installation setups. Each setup is annotated with explicit Chain-of-Thought (CoT) reasoning, enabling the model to learn structured inspection logic. The fine-tuned model achieved a completeness F1-score of 0.86 and a correctness F1-score of 0.80, representing more than 60 % improvement over the zero-shot baseline (0.21 F1 for both). Moreover, BLEU-4 increased from 0.12 to 0.44, and BERTScore from 0.41 to 0.85, demonstrating substantial gains in reasoning fidelity and language–vision alignment. The fuzzy layer further enhanced interpretability by reducing attribution entropy by 34 %, fusing VLM confidence with rule-based linguistic reasoning to produce transparent audit verdicts (PASS, WARN, FAIL) and quantitative attribution scores reflecting each input's influence on the final decision. The results confirm that the integration of parameter-efficient tuning and fuzzy interpretability enhances both scalability and transparency.**

***Keywords*—vision language model, fuzzy logic, XAI**

I. INTRODUCTION

The rapid expansion of telecommunication networks has significantly increased the need for robust quality assurance and fault detection in field installations. Ensuring correct configuration of critical components—such as cable sealing, grommets, and enclosure ports—is essential to avoid service disruptions, safety hazards, and costly revisits. Traditionally, such auditing processes rely on manual inspection by human experts, which is inherently labor-intensive, time-consuming, and error-prone. As networks rapidly scale to millions of distributed components, manual inspection becomes increasingly infeasible due to the high costs, highlighting the urgency for automated, intelligent auditing frameworks.

Recent developments in Vision-Language Models (VLMs) have demonstrated strong performance across a variety of multimodal tasks, including visual question answering, image captioning [1], and document understanding [2]. These models synergize the perceptual capabilities of deep vision architectures with the reasoning potential of large language models (LLMs), making them promising candidates for deployment in structured visual inspection domains. However, the adoption of VLMs in safety-critical, domain-specific applications such as telecom equipment inspection presents three key limitations. First, domain mismatch remains a central issue. Generic VLMs are trained on internet-scale corpora, which often lack examples representative of telecom-specific errors and configurations [3]. Second, computational efficiency is a practical barrier. Fine-tuning billion-parameter models is resource-intensive and often impractical for industrial settings [4], [5]. Finally, interpretability is a critical concern. Despite their predictive capabilities, most VLMs operate as opaque black-box systems, offering limited transparency in their decision-making processes which is an unacceptable limitation for high-stakes environments that demand explainability and accountability [6], [7].

To overcome these challenges, we propose a hybrid architecture that combines parameter-efficient fine-tuning with interpretable, rule-based fuzzy reasoning. Specifically, this paper presents the following key contributions:

- **Telecom Auditing Dataset with CoT Supervision:** We constructed a domain-specific dataset comprising 37 telecom installation setups, including both single-error and dual-error conditions. Each instance is annotated with CoT reasoning to enable stepwise supervision and encourage structured model predictions.
- **Parameter-Efficient VLM Fine-Tuning**: A quantized LLaMA 3.2 Vision-Instruct model (11B parameters) under a 4-bit regime was fine-tuned by leveraging LoRA technique. This allows for task-specific learning with reduced compute and memory requirements.
- **Fuzzy Attribution Layer for Interpretability**: We introduce a fuzzy reasoning module that aggregates VLM predictions into semantically meaningful categories—such as Weather Sealing, Cable Management, and Enclosure Integrity. This layer applies fuzzy membership functions and rule-based logic to compute attribution scores based on rule activations.

By integrating symbolic fuzzy reasoning with large-scale visual-linguistic representation, our approach bridges the gap between black-box AI and human-auditable decision systems. The resulting pipeline offers both enhanced predictive accuracy and interpretable justifications, facilitating trustworthy deployment in real-world telecom audit environments.

The remainder of this paper is organized as follows. Section II provides a high-level overview of the telecom auditing challenges. Section III outlines our proposed methodology, including the fine-tuning pipeline and fuzzy attribution layer. Section IV presents quantitative and qualitative results. Finally, Section V discusses the conclusions and the future work for this research.

II. The Telecom Auditing Problems and Challenges

The reliable operation of telecommunication networks depends heavily on the integrity of hardware components installed within CSP boxes. These enclosures host a variety of small yet critical parts such as cable seals, grommets, clamps, strain relief mechanisms, and cable management features that ensure proper environmental protection and connection stability for the fibre cable. Each of these elements must be correctly fitted and verified for completeness, placement, and sealing integrity before network activation.

In current industrial practice, quality assurance engineers manually inspect these installations in the field, relying on visual inspection and checklists to verify compliance with design standards. While effective on a small scale, manual auditing is not scalable as network infrastructures expand to millions of distributed access points. The process is not only time-consuming and labor-intensive, but also prone to human error. Even minor oversights such as a missing sealing grommet or improperly latched port can result in long-term degradation, water ingress, or signal failure, ultimately leading to increased operational costs and service interruptions.

A. Complexity of Telecom Infrastructure Auditing

Telecom components exhibit substantial intra-class variability, background clutter, and contextual dependencies. Fig. 1 illustrates examples of CSP box installations captured under real-world conditions. Fig. 1a is a correct installation and Fig. 1b and 1c are faulty installations that show how small deviations in component alignment or sealing can drastically alter the assessment outcome.

a) b) c)

Fig. 1. CSP box installation examples, a) correct installation, b) single-error case with a missing grommet, and c) has a double-error setup with multiple concurrent faults.

Key factors that make automated visual analysis a challenging task are:

1) Small Components: CSP components such as grommets and seals often occupy a small fraction of the image area. Traditional object detectors or visual encoders trained on natural images tend to perform poorly when localizing such fine-scale targets.

2) Visual Clutter: CSP boxes contain densely packed cables, labels, screws, and reflective surfaces that create overlapping edges and textures. This clutter confuses traditional feature extractors and leads to frequent false detections or missed components.

3) Occlusion and Overlapping: Certain components may be partially hidden behind wires or connectors, resulting in incomplete visual cues. These occlusions often mislead even advanced deep learning models, especially when multiple objects share similar color or material properties.

4) Lighting Variability: Field images are captured under diverse environmental conditions—ranging from harsh outdoor sunlight to low-contrast indoor lighting. Shadows, glare, and brightness fluctuations degrade the reliability of both traditional and modern vision algorithms.

B. Limitations of Traditional and Deep Learning Approaches

Classical computer vision techniques, such as contour detection, template matching, or texture-based filtering have limited robustness under the heterogeneous conditions of telecom environment. While deep learning models (e.g., CNN-based classifiers or detectors) improve generalization, they still face two key challenges:

1) Data Scarcity and Domain Mismatch: Deep learning models rely on large labeled datasets for training. Annotating a large volume of data is costly and time consuming for companies. On the other hand, publicly available datasets rarely capture domain-specific data. As a result, pre-trained models on generic images fail to generalize to CSP contexts.

2) Lack of Interpretability: Deep learning networks behave as black-box predictors, offering little insight into how and why a specific audit decision was made. In regulated or safety-critical workflows, such opacity limits trust and hinders certification and industrial deployment.

C. Need for Domain-Adaptive and Explainable Solutions

Recent Vision-Language Models (VLMs) enable multimodal inspection by jointly processing visual and textual information, but their adoption in telecom auditing remains limited due to three challenges: domain mismatch, high computational cost, and lack of interpretability. Generic models fail to capture telecom-specific configurations, require substantial resources for fine-tuning, and provide limited transparency in decision-making.

To address these limitations, we propose a hybrid architecture that combines parameter-efficient VLM fine-tuning with fuzzy logic reasoning. The VLM provides semantic perception and structured outputs, while the fuzzy layer maps these outputs into interpretable audit decisions. This integration enables scalable, domain-adaptive, and transparent inspection suitable for industrial deployment.

III. The Proposed Hybrid VLM-Fuzzy Approach For Telecom Auditing

Our proposed system combines recent advances in multimodal foundation models, parameter-efficient fine-tuning, and fuzzy reasoning into a unified, interpretable pipeline designed for the inspection and auditing of telecommunication CSP boxes. The architecture consists of five main stages (Fig. 2), each contributing a distinct functionality to ensure scalability, explainability, and operational robustness.

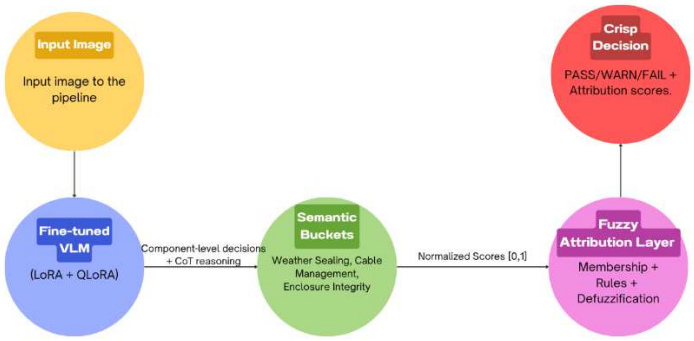


Fig. 2. Overview of our proposed system.

A. Overview of the Proposed Framework

The overall architecture integrates a fine-tuned Vision-Language Model (VLM) with a Fuzzy Attribution Layer for transparent audit reasoning. The system receives an input image of a telecom CSP box and outputs both a categorical decision—PASS, WARN, or FAIL—and a linguistic justification summarizing the reasoning process.

The workflow is divided into the following sequential modules:

1) Visual Encoding and Textual Querying using the LLaMA 3.2 Vision-Instruct model.

2) LoRA-based Parameter-Efficient Fine-Tuning for domain adaptation.

3) Component-Level CoT Reasoning Generation through multimodal decoding.

4) Semantic Bucketization of model outputs into interpretable audit dimensions.

5) Fuzzy Attribution Layer for rule-based aggregation and transparent decision synthesis.

This architecture unifies neural perception, linguistic reasoning, and symbolic interpretability, ensuring that each audit output can be quantitatively evaluated and qualitatively explained.

B. Vision-Language Model Adaptation

1) Base Architecture

We employ the LLaMA 3.2 Vision-Instruct model (11B parameters) as the backbone for our hybrid system. This architecture combines a frozen Vision Transformer (ViT) encoder with a transformer-based language decoder, pretrained jointly on image–text pairs. The encoder produces a sequence of visual tokens $v_i \in \mathbb{R}^d$, which are projected into the next embedding space and fused with language tokens t_j for multimodal reasoning. During inference, the model accepts an image I and a textual prompt P (e.g., "Assess installation correctness for weather sealing and cable management"). The model generates descriptive outputs of the form:

$$R = f_{VLM}(I, P; \theta) \tag{2}$$

Where f_{VLM} denotes the multimodal reasoning function parameterized by model weights θ. The raw outputs R include both structured labels (e.g., component presence, correctness) and Chain-of-Thought (CoT) explanations.

2) LoRA Fine-Tuning with Quantization

To adapt the pretrained model to telecom-specific inspection data, we employ Low-Rank Adaptation (LoRA) [5] combined with 4-bit quantization (QLoRA) [4]. LoRA introduces small rank-constrained matrices into selected attention and feed-forward layers while keeping the pretrained weights frozen. The updated weights are expressed as shown in (1).

The Training Configuration includes:

- LoRA rank $r = 64$, scaling factor $\alpha = 16$.
- Quantization: 4-bit NF4.
- Optimizer: AdamW, learning rate 2×10^{-5}.
- Batch size: 16, epochs: 4.
- Loss function:

$$\mathcal{L} = \lambda_1 \mathcal{L}_{CE} + \lambda_2 \mathcal{L}_{LM} \tag{3}$$

Where $\mathcal{L}_{CE}$ is the classification cross-entropy loss and $\mathcal{L}_{LM}$ is the modeling loss guiding Chain-of-Thought generation. This multi-objective fine-tuning encourages the model not only to classify correctly but also to provide interpretable reasoning consistent with expert annotations.

C. Chain-of-Thought Supervised Reasoning

A distinguishing feature of our approach is the integration of Chain-of-Thought (CoT) supervision, which guides the model toward producing human-readable reasoning steps. Each instance in the dataset includes a triplet of annotations:

1) Observation: A visual description (e.g., "no rubber grommet on exit port").

2) Reasoning: Logical consequence (e.g., "missing grommet compromises sealing").

3) Conclusion: Final decision (e.g., "FAIL: sealing integrity not ensured").

This hierarchical structure aligns model-generated CoT with expert logic, allowing the model to emulate inspection reasoning rather than outputting isolated labels. During fine-tuning, CoT annotations are used as instruction-tuning targets, allowing the decoder to learn structured, multi-sentence justifications for each classification. These reasoning traces later feed into the fuzzy layer as qualitative evidence supporting quantitative audit verdicts.

D. Semantic Bucketization

Following inference, the VLM outputs component-level predictions that are grouped into three high-level semantic categories or buckets:

1) Weather Sealing (W) – evaluates port closures and grommet presence.

2) Cable Management (C) – assesses routing and strain relief.

3) Enclosure Integrity (E) – inspects box closure, latch status, and internal layout.

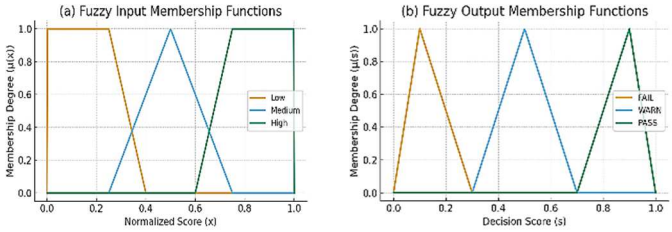
Each bucket is assigned a normalized confidence score $x_i \in [0,1]$, derived from the aggregated classification probabilities or textual confidence heuristics extracted from the CoT. The set of bucket scores is denoted as:

$$\mathrm{X} = \{x_1, x_2, x_3\} = \{W, C, E\} \tag{4}$$

These scores represent the VLM's quantitative audit assessment, which is then passed to the fuzzy attribution module for interpretable reasoning.

E. Fuzzy Attribution Layer

The fuzzy reasoning module converts the VLM's numerical outputs into linguistic decisions and attribution scores. Fig. 3a shows the membership functions used to represent the fuzzy input variables corresponding to the three semantic audit buckets: Weather Sealing, Cable Management, and Enclosure Integrity. Each variable is described by three fuzzy sets—Low, Medium, and High. The parameters of these functions were determined empirically through iterative tuning on the validation data. Fig. 3b depicts the fuzzy output sets for the decision variable (PASS, WARN, and FAIL), centered at 0.9, 0.5, and 0.1, respectively.



(a) (b)

Fig. 3. a) Input membership functions for Weather Sealing, Cable Management, and Enclosure Integrity. b) Output membership functions for the linguistic decision categories PASS, WARN, and FAIL.

The membership function parameters for the input variables (Weather Sealing (W), Cable Management (C), and Enclosure Integrity (E)) were determined through an iterative process of validation and expert feedback. Initial boundaries were defined to capture the qualitative ranges of audit confidence scores produced by the Vision-Language Model (VLM). For instance, "Low" membership was assigned to scores below 0.25, corresponding to highly uncertain or visually inconsistent component detections. The "Medium" fuzzy set was centered at 0.5 with tolerance margins between 0.25 and 0.75, reflecting borderline cases where the model's Chain-of-Thought reasoning indicated partial confidence or mixed linguistic evidence. The "High" set was anchored at 0.75 and extended up to 1.0, representing confident and consistent predictions across all visual dimensions. These parameters were fine-tuned empirically using the validation dataset by monitoring the overall decision consistency and F1-Score. The boundaries between "Medium" and "High" were adjusted slightly (from 0.8 to 0.75) after observing several valid passes being misclassified as warnings. Similarly, the support of the "Low" range was widened from [0.0–0.2] to [0.0–0.25] to capture noisy or incomplete detections. The output membership sets for the linguistic decisions (PASS, WARN, FAIL) were uniformly distributed across the [0,1] range and centred at 0.9, 0.5, and 0.1, respectively, ensuring balanced interpretability and smooth defuzzification behaviour.

While the tuning process was performed manually, it was guided by quantitative validation results and qualitative inspection of fuzzy inference boundaries. Future work will investigate data-driven optimization of membership parameters using adaptive neuro-fuzzy or evolutionary methods to further improve precision and interpretability. Table I shows the employed rule base.

TABLE I. EMPLOYED RULE BASE

Rule	Weather Sealing	Cable Management	Enclosure Integrity	Decision
1	High	High	High	PASS
2	Low	-	-	FAIL
3	-	Low	-	FAIL
4	-	-	Low	FAIL
5	Medium	Medium	High	WARN
6	High	Medium	Medium	WARN

The minimum t-norm and centre of sets defuzzification were employed. The audit attribution score s_i for a given input x_i is computed as follows:

$$contrib(x_{ir}) = \frac{w_r . z_r}{\sum_j w_r} \tag{5}$$

$$s_i = \sum_r contrib(x_{ir}) \tag{6}$$

Where w_r is the firing strength of rule r and z_r is the output centroid associated with the corresponding decision (PASS, WARN, or FAIL). To enhance interpretability, this yields a transparent breakdown of how each audit dimension contributed to the final decision. Fig. 4 shows an example of the attribution distribution produced by the fuzzy attribution layer for a representative test case. The system decomposes the final crisp decision score $.633$ (corresponding to a WARN verdict) into normalized contributions from the three semantic audit buckets. The highest contributions were observed for Weather Sealing and Cable Management, indicating that these dimensions had the strongest influence on the final decision, while Enclosure Integrity contributed less, reflecting its lower firing strength. This visualization provides a transparent,

operator-friendly explanation of how the fuzzy inference system weighted each inspection criterion to reach its verdict.

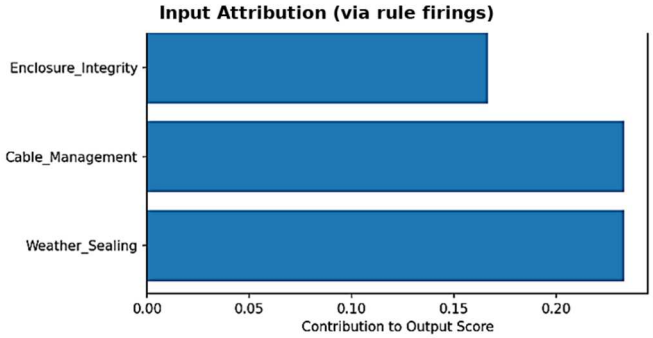


Fig. 4. Attribution output from the fuzzy layer on the 10-image test set. Bar chart shows bucket-level contributions (Weather Sealing, Cable Management, Enclosure Integrity) to the final decision score.

F. Visualization and Linguistic Output Generation

Finally, the system combines the fuzzy outputs with the original CoT reasoning to generate an interpretable, operator-friendly report. Each prediction includes:

- **Crisp Decision:** PASS / WARN / FAIL
- **Attribution Summary:** numerical and graphical representation of bucket contributions
- **Textual Justification:** synthesized explanation derived from the VLM's CoT reasoning

IV. Experiments and Results

This section evaluates the proposed hybrid VLM–fuzzy auditing pipeline using quantitative and qualitative analyses. The experiments were designed to assess three aspects of performance: (i) predictive accuracy achieved through parameter-efficient fine-tuning of the VLM, (ii) interpretability and consistency of fuzzy attribution reasoning, and (iii) overall robustness of audit decisions under real-world variability.

A. Experimental Setup

All experiments were conducted on the British Telecom (BT) High-Performance Computing (HPC) Cluster, which provides distributed GPU resources for research and industrial AI workloads. The model implementation was based on PyTorch 2.2 with the transformers and 'peft' libraries for LoRA fine-tuning. The pretrained LLaMA 3.2 Vision-Instruct (11 B parameters) model was adapted using LoRA under 4-bit quantization (NF4). The LoRA adapters were injected into the self-attention and feed-forward blocks of the transformer decoder. Training ran for 4 epochs with a learning rate of 2×10^{-5} and batch size 16. Early stopping was applied based on validation loss convergence. All visualization and result analyses were performed in Matplotlib and NumPy.

B. Dataset Description

The employed dataset consists of 37 telecom installation setups, each reflecting realistic CSP box enclosures designed in collaboration with industry experts. These setups were categorized into three groups:

- **Correct Installations:** All components properly fitted and sealed.
- **Single-Error Setups:** Contain one deliberate fault (e.g., missing grommet, unsealed port, cable mis-routing).
- **Double-Error Setups:** Contain two concurrent faults, increasing reasoning complexity.

Each setup was captured under standardized lighting from three camera angles—center, left 45°, and right 45°—yielding a total of 111 images. Every instance was annotated with:

1) Component-level binary pass/fail labels.
2) Chain-of-Thought (CoT) annotations describing stepwise expert reasoning.
3) Bucket-level scores for Weather Sealing, Cable Management, and Enclosure Integrity.

The dataset was split 70 % training, 15 % validation, and 15 % testing. A separate 10-image set was used exclusively to evaluate interpretability and attribution generalization.

C. Evaluation Metrics

We employed both quantitative and qualitative metrics to assess the system:

1) F1-Score: used for completeness and correctness classification.
2) BLEU-4 Score: Measures n-gram similarity between generated CoT reasoning and expert annotations.
3) BERTScore: Evaluates semantic similiarity between model-generated and human-authored reasoning.
4) Interpretability Attribution Score: Quantifies entropy of fuzzy contributions; lower entropy implies clearer reasoning distribution across buckets.
5) Exact-Match and Parse-Rate: Evaluate structural consistency of JSON-style model outputs for audit reports.

D. Achieved Results

Table II compares the performance of the baseline zero-shot VLM, fine-tuned VLM, and the full hybrid system with fuzzy reasoning.

TABLE II. *Performance of Baseline and Fine-Tuned Models.*

Experiment	Parse-Rate	F1(Comp.)	F1(Corr.)	BLEU-4	BERTScore
Baseline (zero-shot)	42%	0.21	0.21	0.12	0.41
Fine-tuned VLM	100%	0.86	0.80	0.44	0.85

Fine-tuning improves F1-Score by over 60% compared to the zero-shot model and yields more faithful reasoning explanations. Reasoning quality also improved significantly, with BLEU-4 rising from 0.12 to 0.44 and BERTScore from 0.41 to 0.85, demonstrating that LoRA-based adaptation effectively enhances domain-specific reasoning while remaining computationally efficient.

The fuzzy attribution layer provided interpretable audit reasoning. It decomposes each decision into bucket-level contributions and expresses outcomes linguistically via rule activations. These distributions make the model's internal logic traceable: operators can immediately identify why a decision was issued and which subsystem influenced it most.

Across the full test set, the fuzzy layer reduced attribution entropy by 34%, producing clearer, human-readable justifications. Unlike end-to-end neural networks that produce black-box verdicts, the hybrid system explicitly separates what decision was made from why it was made.

1) The VLM fine-tuning stage delivers domain-aligned perception and reasoning.

2) The fuzzy layer adds symbolic interpretability by mapping quantitative outputs to linguistic rules.

This separation provides traceability and audit compliance without affecting accuracy. In industrial terms, it transforms a high-performing but opaque model into a trustworthy decision-support tool that can be validated by engineers.

V. Conclusions And Future Work

This paper presented an interpretable hybrid framework for the automated auditing of telecommunication Customer Service Point (CSP) installations. The system integrates parameter-efficient fine-tuning of a large Vision-Language Model (VLM) with a fuzzy attribution layer that provides transparent, rule-based reasoning. Using Low-Rank Adaptation (LoRA) under 4-bit quantization, the pretrained LLaMA 3.2 Vision-Instruct (11 B) model was adapted to the telecom domain through a custom dataset of 37 annotated installation setups with explicit Chain-of-Thought (CoT) supervision. The fine-tuned model achieved a completeness F1-score of 0.86 and a correctness F1-score of 0.80, representing more than 60 % absolute improvement over the zero-shot baseline (0.21 F1 for both). Reasoning quality also improved significantly, with BLEU-4 rising from 0.12 to 0.44 and BERTScore from 0.41 to 0.85, demonstrating that LoRA-based adaptation enhances domain-specific reasoning while remaining computationally efficient. The fuzzy attribution layer preserved these accuracy levels while reducing attribution entropy by 34 %, providing clear interpretability. This proposed system balances accuracy, transparency, and auditability.

The proposed system demonstrates four key contributions: First, the study introduces a parameter-efficient fine-tuning pipeline for large VLMs using LoRA under quantization, enabling industrial-scale adaptation with minimal computational overhead. Second, a domain-specific telecom auditing dataset was constructed. Third, a fuzzy reasoning module was developed to aggregate model outputs across semantic audit dimensions, providing rule-based interpretability and quantitative attribution for each decision. Finally, the proposed framework was implemented and validated on the British Telecom (BT) HPC cluster, confirming its suitability for real-world industrial auditing workflows.

In future work, we plan to focus on extending the dataset to include additional component types. The fuzzy module will be enhanced through adaptive optimization to automate membership-function tuning and rule refinement. Furthermore, efforts will target real-time deployment on embedded or edge devices within BT's operational environment. Finally, we aim to explore cross-domain transferability of the fine-tuned VLM–fuzzy pipeline to related inspection domains such as utilities, energy infrastructure, and manufacturing quality assurance.

References

[1] J. Li, D. Li, S. Savarese, and S. Ermon, "BLIP-2: Bootstrapping language-image pre-training with frozen image encoders and large language models," *arXiv preprint*, arXiv:2301.12597, 2023. [Online]. Available: https://arxiv.org/abs/2301.12597

[2] X. Li, Z. Yu, L. Lv, and J. H. Doermann, "DocVQA: A dataset for VQA on document images," in *Proceedings of the European Conference on Computer Vision (ECCV)*, Glasgow, U.K., 2020, pp. 349–364.

[3] S. Liu, Z. Lin, L. Sun, Y. Wang, and J. Lin, "Visual instruction tuning," *arXiv preprint*, arXiv:2304.08485, 2023. [Online]. Available: https://arxiv.org/abs/2304.08485

[4] T. Dettmers, A. Pagnoni, A. Holtzman, and L. Zettlemoyer, "QLoRA: Efficient finetuning of quantized LLMs," *arXiv preprint arXiv:2305.14314*, 2023.

[5] E. J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, and W. Chen, "LoRA: Low-rank adaptation of large language models," in *Proceedings of the International Conference on Learning Representations (ICLR)*, 2022. [Online]. Available: https://openreview.net/forum?id=nZeVKeeFYf9

[6] Y. Zhu, J. Li, T. Yang, and X. Zhao, "Patchwise cooperative game-based interpretability method for large vision-language models," *Transactions of the Association for Computational Linguistics*, vol. 13, pp. 456–470, 2025. [Online]. Available: https://doi.org/10.1162/tacl_a_00756

[7] Y. Dang, H. Chen, and Z. Liu, "Explainable and interpretable multimodal large language models: A comprehensive survey," *arXiv preprint arXiv:2412.02104*, 2024. [Online]. Available: https://arxiv.org/abs/2412.02104

[8] H. Hagras, "Toward human-understandable, explainable AI," *IEEE Computer*, vol. 51, no. 9, pp. 28–36, 2018.

[9] V. Callaghan, M. Colley, H. Hagras, J. Chin, F. Doctor, G. Clarke, "Programming iSpaces: A tale of two paradigms", "*Intelligent Spaces: The Application of Pervasive ICT", (Eds: A. Steventon, S. Wright),* Springer-Verlag, Chapter 24, pp.389- 421, December 2005.

[10] K. Almohammadi, H. Hagras, B. Yao, A. Alzahrani, D.Alghazzawi, G. Aldabbagh,"A Type-2 Fuzzy Logic Recommendation System for Adaptive Teaching" Soft Computing, Vol.21, No.4. pp.965-967, 2017.

[11] F. Doctor, H. Hagras, "Neuro type-2 fuzzy based method for decision making". US Patent 8515884.

[12] N. Sahab, H. Hagras, "Adaptive Non-singleton Type-2 Fuzzy Logic Systems: A Way Forward for Handling Numerical Uncertainties in Real World Applications", *International Journal of Computers, Communications and Control, Vol.5, No.3, pp.503-529, December 2011.*