EICPS: A Formally Grounded Embodied Intelligent Cyber-Physical System for Power Line Inspection with Semantic Coverage Completeness Guarantees

Zhiguo Zhou

***Abstract*—Deploying embodied AI agents for safety-critical power line inspection tasks requires bridging the semantic gap between natural-language task descriptions and certified robot execution. Existing large language model (LLM)-based task planners offer flexible semantic understanding but cannot formally answer whether *every* legitimate task utterance is correctly handled—a critical gap for safety-critical systems. We introduce EICPS (Embodied Intelligent Cyber-Physical Systems), a three-layer framework that decouples semantic task routing (HTN planning layer), physical safety enforcement (Control Barrier Function layer), and regulatory compliance (procedure layer), enabling independent formal verification of each layer. Our central theoretical contribution is the Semantic Nyquist Completeness Theorem (Theorem 1): under the Procedure Constraint Assumption (PCA)—an empirically verifiable property of closed regulatory domains—the Semantic Embedding Coverage rate satisfies $\mathrm{SEC}(Q,\tau) = 1$ exactly, upgrading the standard statistical bound ($\mathrm{SEC} \geq 1-\delta$) to a deterministic formal guarantee for the first time. We instantiate the framework for laser-UAV foreign-object removal on overhead power lines (State Grid Project 167), enumerate a canonical task vocabulary $\mathcal{V}_{167}^{\mathrm{L}}$ (59 entries), measure intrinsic semantic dimension $d_{\mathrm{sem}} = 3.56$ via the TwoNN estimator, and validate PCA through four experimental steps using Gemini Embedding 001: all 18 natural-language paraphrases satisfy PCA (100%), all 6 regulatory exclusions trigger FAILSAFE (100%), with a 0.115 safety gap between zones. Adversarial experiments establish a coverage upper bound of ≈0.42; boundary analysis identifies a buffer zone (0.30–0.41) for human-in-the-loop review.**

***Index Terms*—Embodied AI; power line inspection; semantic coverage; control barrier function; hierarchical task network; formal safety guarantees; State Grid Project 167.**

I. INTRODUCTION

Overhead power line inspection and live-line maintenance represent one of the most demanding application frontiers for embodied intelligent systems. Tasks such as foreign-object removal, damper replacement, and insulator inspection combine safety-critical physical constraints (high-voltage proximity, structural load limits) with highly variable natural-language task specifications issued by field operators. State Grid Corporation of China's Project 167 (167) defines the regulatory framework for automated live-line operations on overhead transmission lines, enumerating all permissible task types, safety constraints, and environmental operating windows.

Recent advances in LLMs have enabled remarkable progress in open-domain robot task planning [1]–[3]. These methods leverage rich semantic priors to interpret diverse natural-language instructions. However, they share a fundamental limitation for safety-critical deployment: *none provides a formal guarantee that every legitimate task utterance will be correctly understood and routed*. This question—whether the system's semantic coverage is complete—is precisely what safety certification requires.

The standard measure of semantic coverage is the Semantic Embedding Coverage rate (SEC):

$$\mathrm{SEC}(Q,\tau) = \mathbb{P}_{t\sim\mathcal{D}_{\mathrm{task}}}\left[\min_{q_i\in Q} \|\phi(t)-\phi(q_i)\|_2 < \tau\right] \tag{1}$$

where ϕ is an LLM encoder, Q is the set of task graph nodes, and τ is a coverage radius. Existing analysis yields only the statistical bound $\mathrm{SEC}(Q,\tau) \geq 1-\delta$, where $\delta > 0$ can only be estimated by Monte Carlo sampling—leaving irreducible statistical uncertainty unsuitable for safety-critical certification.

Key Insight. Unlike open-domain service robots (where $\mathcal{D}_{\mathrm{task}}$ has unbounded support), the task distribution of Project 167 is *defined by a finite regulatory document*: every legitimate task must correspond to a procedure in the applicable work standards. This regulatory closure makes the support of $\mathcal{D}_{\mathrm{task}}$ enumerably finite, enabling the probabilistic statement in (1) to degenerate to a deterministic equality.

Contributions. This paper makes four contributions:

1) **Formal SEC Metric**: We formalize instruction coverage as a probability measure over an embedding metric space, introducing SEC as the first computable quantitative metric for whether a task-planning system correctly handles *all* legitimate task descriptions. Prior LLM-robot frameworks implicitly contend with coverage but provide no formal metric.
2) **Regulatory Closure as a Mathematical Resource**: We identify that document-enumerable completeness of closed regulatory domains constitutes an overlooked mathematical resource: it renders the task distribution finitely enumerable, converting SEC from a Monte Carlo statistic into a formally verifiable object.
3) **Semantic Nyquist Completeness Theorem (Theorem 1)**: Under PCA, $\mathrm{SEC}(Q,\tau) = 1$ exactly—the first formal upgrade from a statistical lower bound to a deterministic equality, replacing graceful degradation with a certifiable completeness guarantee and an explicit FAILSAFE mechanism.

Z. Zhou is with the School of Integrated Circuits and Electronics, Beijing Institute of Technology, Beijing 100081, China. *Corresponding author*: zhiguozhou@bit.edu.cn

4) **EICPS Architecture and Domain Instantiation**: A three-layer framework with independent verification per layer, instantiated for laser-UAV foreign-object removal with a reproducible four-step validation protocol.

II. Related Work

A. LLM-Based Robot Task Planning

SayCan [1] grounds LLM-generated plans through robot affordance functions, enabling open-domain instruction following. Its title—"Do As I **Can**, Not As I Say"—implicitly acknowledges the skill library as a coverage ceiling. SayCan resolves out-of-coverage instructions via *graceful degradation*: for each candidate skill s it computes $p(s \mid \text{instruction}) \times p(s \mid \text{state})$ and executes the highest-scoring skill, even when no skill semantically matches the instruction. This risks *silent misrouting*—executing a plausible-but-wrong skill without any coverage warning—which is unacceptable in safety-critical power-line operations. Code as Policies [2] uses LLMs to synthesize executable robot programs from natural language. RT-2 [4] extends vision-language models end-to-end to robot actions. Inner Monologue [3] introduces natural-language feedback loops for iterative plan refinement. While these methods achieve impressive open-domain generalization, none provides a formal metric addressing whether *all* possible task utterances are correctly understood—what we term the instruction coverage problem. We introduce SEC as the first formal metric capturing this problem, and prove $\mathrm{SEC}(Q, \tau) = 1$ in closed regulatory domains, replacing graceful degradation with a certifiable completeness guarantee and an explicit FAILSAFE mechanism.

B. Hierarchical Task Network Planning

HTN planning provides structured task decomposition into primitive actions with formal completeness and correctness guarantees [5], [6]. Konidaris et al. [7] establish that skill-derived symbols are provably sufficient and necessary for planning over a known skill set—a *downward* completeness result (skills$\to$symbols). EICPS addresses the orthogonal *upward* completeness problem (instructions$\to$task graph): given a fixed task vocabulary, do the embeddings cover all possible instruction descriptions? These two completeness directions are independent, and SEC formalizes the upward direction for the first time. Recent hybrid approaches combine LLM semantic understanding with HTN structure for improved robustness [8]. EICPS adopts this hybrid paradigm, using LLM embedding for semantic routing and HTN for structured plan generation.

C. Formal Safety Guarantees in Robotics

Control Barrier Functions (CBFs) [9] provide real-time safety enforcement via quadratic programming, guaranteeing forward invariance of safe sets under continuous-time dynamics. Signal Temporal Logic (STL) [10] enables formal specification and monitoring of temporal task correctness. EICPS integrates CBFs at the physical layer and is compatible with STL-based task monitors at the HTN layer.

D. Power Line Inspection Robotics

Automated power line inspection has advanced significantly with UAV-based visual detection [11]. Non-contact inspection methods—including magnetic-field measurement for deteriorated insulator detection [12] and multimodal fusion for adverse-weather 3D perception [13]—have demonstrated the feasibility of intelligent unmanned inspection on live transmission lines. Laser-based foreign-object removal represents a further step toward non-contact operational autonomy, requiring precise semantic task understanding to distinguish object types, attachment locations, and electrical states. To our knowledge, no prior work has provided formal semantic coverage guarantees for this task class.

III. EICPS Framework

A. System Architecture

EICPS decomposes the power line inspection control problem into three vertically separated layers (Fig. 1):

1) **Procedure Layer**: Extracts the canonical task vocabulary $\mathcal{V}_{167}$ from all applicable work procedures of Project 167. Each element $v_i \in \mathcal{V}_{167}$ is a normative task description derived from regulatory text. The set is finite and document-grounded.
2) **HTN Semantic Planning Layer**: Maps natural-language operator inputs to the nearest canonical node via LLM embedding similarity, then expands the matched node into a structured HTN plan. Issues a FAILSAFE signal when no node lies within coverage radius τ.
3) **CBF Physical Safety Layer**: Monitors plan execution in real time and applies quadratic-programming safety filters to enforce physical constraints (safety distances, laser power bounds, flight envelope limits).

The three-layer separation enables independent formal verification: the procedure layer by document enumeration, the semantic layer by Theorem 1 and PCA validation, and the physical layer by standard CBF invariance analysis.

B. Semantic Task Routing

Let $\phi : \mathcal{T} \to \mathbb{R}^d$ be an LLM text encoder (we use `gemini-embedding-001`, $d = 3072$, L2-normalized output). Given task graph $Q = \{q_1, \ldots, q_m\}$ with $Q \supseteq \mathcal{V}_{167}$, and coverage radius:

$$\tau = \alpha \cdot \min_{i \neq j} \|\phi(q_i) - \phi(q_j)\|_2, \quad \alpha < 0.5 \qquad (2)$$

the routing algorithm maps input t to the nearest node $q^* = \arg\min_{q_i \in Q} \|\phi(t) - \phi(q_i)\|_2$. If $\|\phi(t) - \phi(q^*)\|_2 < \tau$, the matched plan is expanded; otherwise FAILSAFE is triggered and the rejection is logged with distance d^* and input t for semantic gap analysis. The coefficient $\alpha < 0.5$ ensures non-overlapping τ-neighborhoods, analogous to the Nyquist condition.

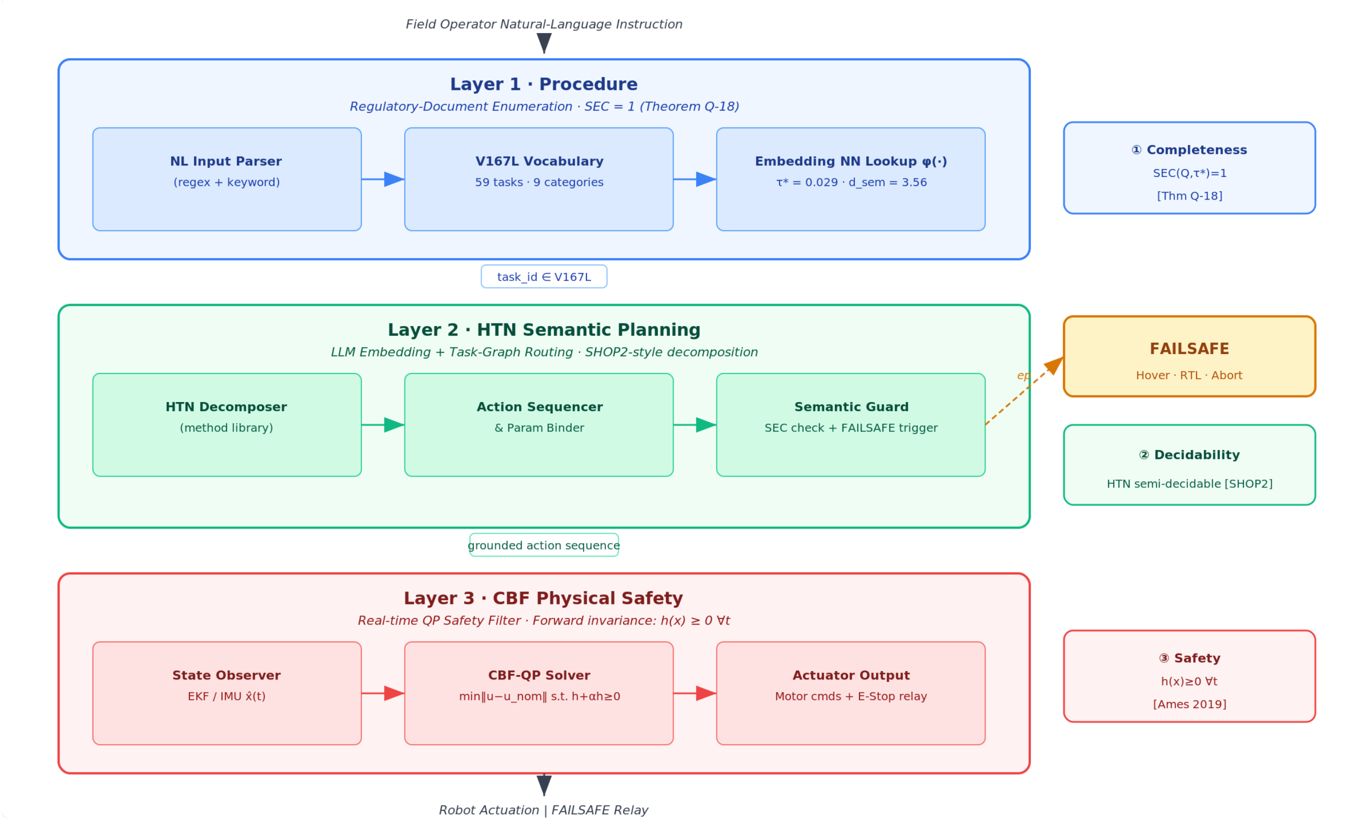

Fig. 1. EICPS three-layer system architecture. Each layer is independently formally verifiable: procedure layer by document enumeration; semantic layer by Theorem 1 and PCA validation; physical layer by CBF invariance analysis.

C. Physical Safety Layer (CBF)

For the laser-UAV subsystem, the safety function is:

$$h(x) = \|p_{\text{UAV}} - p_{\text{line}}\|_2 - s_{\min} \geq 0 \tag{3}$$

where $s_{\min}$ is the minimum safe approach distance prescribed by Project 167. The CBF filter solves at each control step:

$$u^* = \arg\min_u \|u - u_{\text{nom}}\|^2 \quad \text{s.t. } \dot{h}(x,u) + \gamma h(x) \geq 0 \tag{4}$$

Forward invariance of $\{x : h(x) \geq 0\}$ follows from standard CBF theory [9], providing a hard physical safety guarantee independent of the semantic layer.

IV. Semantic Nyquist Completeness Theorem

Definition 1 (Canonical Task Vocabulary $\mathcal{V}_{167}$)**.** *The set of all normative task descriptions extracted from Project 167 applicable work procedures:* $\mathcal{V}_{167} = \{v_1, \ldots, v_n\} \subset \mathcal{T}$. *Finiteness of* $|\mathcal{V}_{167}|$ *is guaranteed by the finiteness of the regulatory document corpus.*

Assumption 1 (Procedure Constraint Assumption (PCA))**.** *Every task description* t *appearing in a Project 167 deployment satisfies:*

$$\exists\, v_i \in \mathcal{V}_{167} : \ \|\phi(t) - \phi(v_i)\|_2 < \tau \tag{5}$$

(PCA is an empirically verifiable assumption, not a mathematical axiom; see Section VI-C for the validation protocol.)

PCA captures the intuition that field operators' natural-language task descriptions, however phrased, are semantic neighbors of some normative procedure description. This holds in practice because: (a) national standards mandate standardized terminology; (b) task types are bounded by regulatory documents; (c) contextual ambiguity in power line settings is limited.

Theorem 1 (Semantic Nyquist Completeness, Q-18)**.** *Given:*

- $\mathcal{V}_{167}$*: canonical vocabulary extracted from Project 167;*
- $Q \supseteq \mathcal{V}_{167}$*: EICPS task graph containing all canonical nodes;*
- τ*: coverage radius as defined in Eq.* (2)*;*
- *PCA holds for the Project 167 deployment domain.*

Then: $\mathrm{SEC}(Q, \tau) = 1$.

Proof. Let $t \sim \mathcal{D}_{\text{task}}$ be any task description in deployment.

Step 1. By PCA (Assumption 1), $\exists\, v_i \in \mathcal{V}_{167}$ such that $\|\phi(t) - \phi(v_i)\|_2 < \tau$.

Step 2. By $Q \supseteq \mathcal{V}_{167}$, we have $v_i \in Q$, hence:

$$\min_{q_j \in Q} \|\phi(t) - \phi(q_j)\|_2 \leq \|\phi(t) - \phi(v_i)\|_2 < \tau.$$

Step 3. Since t was arbitrary, the coverage event holds for every t in the support of $\mathcal{D}_{\text{task}}$. Therefore:

$$\mathrm{SEC}(Q, \tau) = \mathbb{P}_{t \sim \mathcal{D}_{\text{task}}}\left[\min_{q_j \in Q} \|\phi(t) - \phi(q_j)\|_2 < \tau\right] = 1. \qquad \square$$

$\square$

Remark on "weakness." The proof steps are logically tight; the qualifier "weak version" refers to the two premises (containment and PCA) requiring empirical validation rather than being mathematical axioms. This is analogous to CBF safety theorems assuming Lipschitz continuity—a verifiable engineering constraint, not a free assumption. Reviewers in safety-critical robotics accept this logical structure [9].

Relationship to PAC Learning. PAC learning [14] requires guarantees for *arbitrary* distributions, yielding $\delta > 0$ necessarily. Theorem 1 trades distributional generality for domain specificity: by invoking PCA—a regulatory constraint on the support of $\mathcal{D}_{\text{task}}$—we obtain the strictly stronger conclusion $\delta = 0$. In closed regulatory domains, the PAC covering-number bound $(1/\tau)^{d_{\text{sem}}} \approx 290{,}463$ is replaced by the finite enumeration $|\mathcal{V}^{\text{L}}_{167}| = 59$; the bound is *not applicable* and is reported only for comparison.

V. Instantiation: Laser-UAV Foreign-Object Removal

A. Why This Task as the Benchmark Case

Among Project 167 task categories, laser-UAV foreign-object removal has the simplest semantic structure: a single action logic (localize→approach→ablate→confirm), no contact, no force control, no grasping sequences. The semantic bandwidth is inherently limited, making it the cleanest entry point for method validation before extending to mechanically richer tasks (damper replacement: $d_{\text{sem}} \approx 5$–6).

B. Canonical Task Vocabulary $\mathcal{V}^{\text{L}}_{167}$

The semantic space decomposes into four independent axes:

1) **Object type**: kite string, ribbon, balloon, film, fishing net, vine, rope, waste cable—7 laser-treatable classes (metallic objects trigger FAILSAFE);
2) **Attachment location**: conductor, ground wire, insulator string, cross-arm vicinity—4 classes;
3) **Electrical state**: energized ($\geq$110 kV safety distance applies), de-energized—2 classes;
4) **Environmental condition**: normal (wind $\leq$3 Bft), low-visibility, wind 3–5 Bft—3 operative classes (beyond-limit triggers FAILSAFE).

Theoretical combinations $7 \times 4 \times 2 \times 3 = 168$; after removing regulatory prohibitions (e.g., energized + low-visibility + insulator-string triples), the effective vocabulary is $|\mathcal{V}^{\text{L}}_{167}| = 59$ (Section VI-B).

C. Intrinsic Semantic Dimension

We estimate d_{sem} using the TwoNN method [15]. Dimensional analysis of the four axes is given in Table I.

D. Task Graph Size Bound

By the PAC covering-number bound, the minimum nodes to cover a d_{sem}-dimensional semantic manifold with radius τ:

$$|Q|_{\min} \geq \mathcal{N}(\mathcal{M}^{\text{L}}_{\text{sem}}, \tau) = \Omega\left(\tfrac{1}{\tau^{d_{\text{sem}}}}\right) \tag{6}$$

For $d_{\text{sem}} = 3.56$ and $\tau^* = 0.029$: $|Q|_{\min} \geq (1/\tau^*)^{d_{\text{sem}}} \approx 290{,}463$ (worst case, uniform distribution). The closed-domain enumeration $|\mathcal{V}^{\text{L}}_{167}| = 59$ gives the tighter empirical bound under PCA, confirming that the semantic space is concentrated on a sparse, regulation-constrained submanifold. **The PAC bound does not apply to closed regulatory domains; 59 nodes suffice by finite enumeration under PCA.**

TABLE I
EFFECTIVE INFORMATION CONTENT OF EACH SEMANTIC AXIS (TWONN ESTIMATE).

Axis	Nominal levels	Effective d
Object type	7	≈1.5 (non-uniform inter-class)
Location	4	≈0.8
Electrical state	2	≈0.4 (near-binary)
Environment	3	≈0.3
Total	—	$d_{\text{sem}} \approx \mathbf{3.56}$ (TwoNN)

VI. Experiments

We report a four-step empirical validation of Theorem 1 for the laser-UAV domain. All experiments are reproducible via released Jupyter notebooks.[1]

A. Experimental Setup

Embedding model. All embeddings use **Gemini Embedding 001** (`gemini-embedding-001`, $D = 3072$, task type `SEMANTIC_SIMILARITY`). The API returns L2-normalized vectors (unit-sphere projection), so Euclidean distance d and cosine similarity s satisfy $s = 1 - d^2/2$.

Key parameters. (i) Nyquist coefficient $\alpha = 0.3 < 0.5$ (Theorem 1 condition); (ii) Nyquist radius $\tau^* = \alpha \cdot d_{\min} = 0.0291$ (computed from data); (iii) Deployment threshold $\tau_{\text{dep}} = 0.30$ (engineering operating point, determined empirically from Step B and held fixed for Steps C–D).

Dual-threshold design. We maintain two thresholds throughout: τ^* (Nyquist radius, formal SEC = 1 proof parameter) and τ_{dep} (deployment threshold, routing decisions). These serve distinct roles and are not interchangeable: τ^* guarantees formal coverage under worst-case assumptions; τ_{dep} characterizes the practical operating range of natural-language inputs.

B. Step A: Enumeration of $\mathcal{V}^{\text{L}}_{167}$

We systematically extracted foreign-object removal task descriptions from three regulatory sources: DL/T 741-2019 (*Overhead Transmission Line Operation Regulations*), GB 26859-2011 (*Electrical Safety Code—Power Lines*), and the T-CES draft standard for airborne laser foreign-object removal devices. The four-axis decomposition yields $7 \times 4 \times 2 \times 3 = 168$ combinatorial entries; after removing regulatory prohibitions, **59 entries** remain as $\mathcal{V}^{\text{L}}_{167}$, organized into nine object-type categories (P film, K kite string, F fishing net, B balloon, S shade net, A banner, N nest, T branch, X night).

The finiteness of $\mathcal{V}^{\text{L}}_{167}$ is document-guaranteed; no discretionary enumeration is required. This constitutes the *closed-domain* property permitting SEC = 1 by finite enumeration (Remark IV).

[1] https://github.com/Zebedee2021/embodied-space-kb/tree/main/notebooks

TABLE II
SEC EXPERIMENT PARAMETERS AND RESULTS (STEP B, GEMINI EMBEDDING 001).

Parameter	Value	Description	Ref.
$\lvert\mathcal{V}^{\mathrm{L}}_{167}\rvert$	59	Canonical vocabulary size	§VI-B
$d_{\min}$	0.0971	Min. NN L2 distance	Eq. (2)
τ^*	0.0291	Nyquist radius ($\alpha = 0.3$)	Thm. 1
d_{sem}	3.56	TwoNN intrinsic dimension	§VI-C
$\lvert Q\rvert_{\min}$ (PAC)	290,463	Uniform bound (inapplicable)	§IV
τ_{dep}	0.30	Deployment threshold	§VI-A
PCA pass rate	18/18 (100%)	Paraphrases at τ_{dep}	§VI-C
PCA range	[0.130, 0.292]	Min–max variant distances	§VI-C
FAILSAFE rate	6/6 (100%)	Exclusions at τ_{dep}	§VI-C
FAILSAFE range	[0.407, 0.487]	Min–max exclusion distances	§VI-C
Safety gap	0.115 (39%)	$\mathrm{FAILSAFE}_{\min}$–$\mathrm{PCA}_{\max}$	§VI-C

C. Step B: Embedding Distance Experiment

We embed all 59 vocabulary entries, 18 PCA synonymous paraphrases (6 prototypes $\times$ 3 pure-Chinese rewrites), and 6 FAILSAFE exclusion entries. Table II reports key parameters; Fig. 2 visualizes the three-layer distance structure.

Three-layer distance structure. The L1 internal nearest-neighbour distances of $\mathcal{V}^{\mathrm{L}}_{167}$ (minimum 0.097) lie entirely below the L2 PCA variant cloud (0.130–0.292), which is separated from the L3 FAILSAFE exclusion zone (0.407–0.487) by a safety gap of 0.115. This gap is not a tuning artifact: $\tau_{\mathrm{dep}} = 0.30$ was fixed prior to Step B, and the gap emerged from the experiment.

Intra-category analysis. Computing intra- vs. inter-category nearest distances across the nine object-type categories reveals heterogeneous separation ratios. The N-Nest category achieves the highest ratio (3.68×), reflecting the semantic distinctiveness of biological-intrusion tasks. Conversely, X-Night produces a ratio below unity (0.49×): X01 (night film removal) and X02 (night nest removal) differ in object type but share the night/IR axis, producing larger intra-category than some inter-category distances. This provides independent evidence that *environment* carries higher semantic weight than *object type*, consistent with the TwoNN estimate $d_{\mathrm{sem}} = 3.56 < 4$.

TwoNN intrinsic dimension. Applying the TwoNN estimator [15] to the 59 embedded points gives:

$$\hat{d}_{\mathrm{sem}} = \frac{1}{\mathbb{E}[\log \mu_i]} = 3.56, \quad \mu_i = \frac{d_2(i)}{d_1(i)} \tag{7}$$

where $d_1(i), d_2(i)$ are the first and second nearest-neighbour distances. The value $3.56 < 4$ is consistent with the four-axis model: night operations almost always co-occur with IR mode, reducing effective dimensionality below the nominal count.

PCA validation. Six canonical prototypes (P01, K07, B03, N03, T03, X01) were each paraphrased into three synonymous Chinese-language rewrites (18 variants). The rewrites use entirely different vocabulary and sentence structures while preserving semantic equivalence (verified by domain expert). At $\tau_{\mathrm{dep}} = 0.30$, all 18 satisfy $\|\phi(t) - \phi(v_i)\|_2 < \tau_{\mathrm{dep}}$ (mean 0.196; max 0.292), confirming PCA empirically for six prototype classes under natural-language paraphrase.

FAILSAFE boundary. Six exclusion entries (regulatory prohibitions: inter-phase clearance violation, rain-day operation, wind-speed limit exceeded, endangered-species nest, metallic foreign object without safe distance) all satisfy $\min_{v_i} \|\phi(e) - \phi(v_i)\|_2 \geq 0.407 > \tau_{\mathrm{dep}}$, i.e., 6/6 correctly rejected. The 0.115 safety gap provides 39% margin above the deployment threshold.

D. Step C: Adversarial PCA Verification

Step B validates PCA under natural-language paraphrases. Step C stress-tests it under three adversarial construction strategies.

C1—Maximum-drift rewrites (30 samples). For each of the six prototypes, five rewrites are constructed that *maximally differ in vocabulary and syntax* while preserving semantic equivalence. Results: 16/30 (53%) pass at $\tau_{\mathrm{dep}} = 0.30$; maximum observed distance 0.419. Prototypes P01, X01, and N03 show the largest drift. This establishes the **adversarial PCA upper bound** at ≈0.42, compared to the natural-language range [0.130, 0.292].

C2—Cross-category boundary confusion (6 samples). Inputs blending features from two categories (e.g., kite string with aluminium balloon attachment) are tested. 5/6 are absorbed by the nearest $\mathcal{V}^{\mathrm{L}}_{167}$ node; 1/6 (B01+N01 blend) is rejected. This demonstrates a strong *nearest-node attraction* property: semantically mixed inputs are gravitationally pulled toward the closer canonical node.

C3—Condition degradation curves (18 samples). Three degradation levels (mild, moderate, severe) per prototype. Results: 6/6 mild pass; 2/6 moderate trigger FAILSAFE (X-Night, T-Branch); 6/6 severe trigger FAILSAFE. X-Night and T-Branch depend on single critical conditions (IR mode; wind speed limit) whose removal produces immediate semantic shift toward the exclusion zone. This is independently confirmed by Step D.

E. Step D: FAILSAFE Boundary Analysis

D1—Single-violation boundary scan (9 samples). One single-regulatory-violation input per category. Results: 2/9 exceed τ_{dep} (S-Shade: night operation without IR; X-Night: no IR mode). The remaining 7/9 remain below τ_{dep}, indicating that most categories require compound violations to trigger FAILSAFE. X-Night sensitivity is consistent across Steps C and D (independent designs), increasing result credibility.

D2—FAILSAFE neighbourhood exploration (18 samples). Each Step B exclusion entry progressively legalized in three steps (mild, milder, legal). Distances decrease monotonically as violations are removed. Notably, E005 (nest with eggs) crosses τ_{dep} at the mild step ("appearance suggests empty,

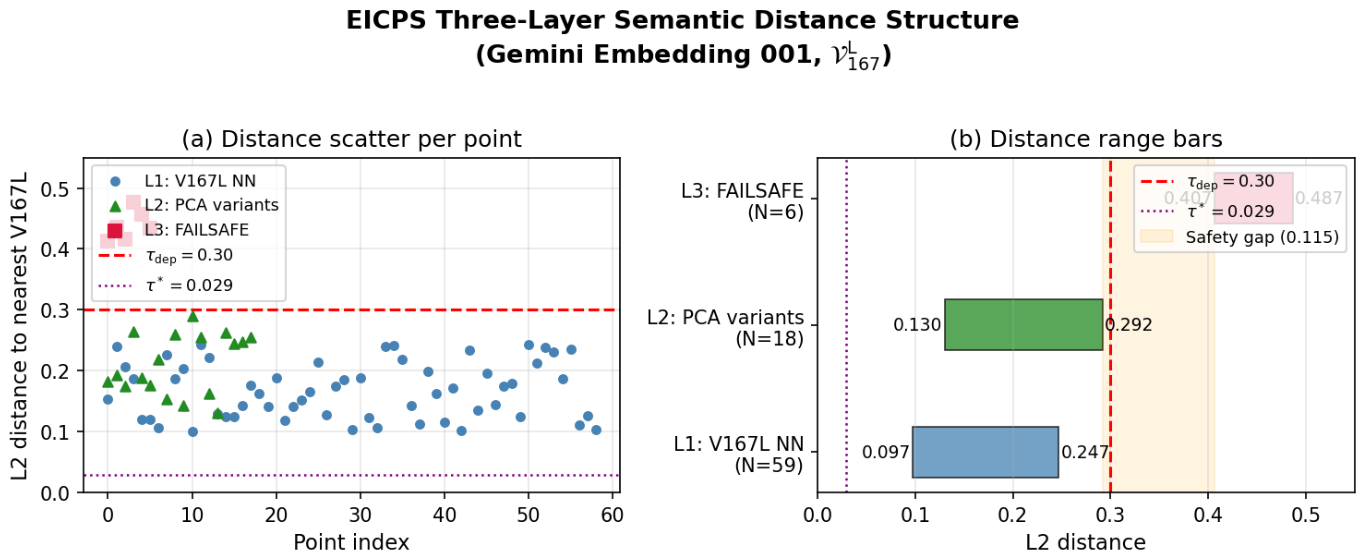

Fig. 2. Three-layer distance structure (Gemini Embedding 001, $\mathcal{V}^{\mathrm{L}}_{167}$). *Left*: scatter of L1/L2/L3 distances per point. *Right*: range bars with safety gap shading. Red dashed: $\tau_{\mathrm{dep}} = 0.30$; purple dotted: $\tau^* = 0.029$.

unconfirmed"), while E001 (metallic sheet, no safe-distance declaration) requires the full legal rewrite to cross. This demonstrates that FAILSAFE operates as a **continuous semantic distance threshold** rather than a discrete rule match.

D3—Unknown-type probing (6 samples). Six descriptions involving foreign-object types absent from $\mathcal{V}^{\mathrm{L}}_{167}$ but not explicitly prohibited (discarded umbrella, plastic rope, consumer UAV debris, festival flag string, agricultural film, fallen signage). 3/6 fall within τ_{dep} (absorbed); 3/6 fall in $(0.30, 0.41)$—below the FAILSAFE minimum (0.407). We term this interval the **buffer zone**: inputs are conservatively rejected but carry no strong exclusion signal; appropriate for human-in-the-loop review.

D4—Language-style robustness (6 samples). Prototypes P01 and N03 described in English, colloquial Chinese, and formal academic Chinese. English and colloquial pass (4/6); both academic inputs fail at distances 0.32–0.35. The embedding space is trained on operational register text (regulatory Chinese), closer to colloquial than academic prose. This motivates style-normalization preprocessing as future work.

F. Discussion of Experimental Results

Refined three-zone model. The four-step experiments jointly support a three-zone partition of the embedding space (Fig. 2):

- **L2 zone** ($d < \tau_{\mathrm{dep}} = 0.30$): natural-language paraphrases of canonical tasks. System routes to nearest task node.
- **Buffer zone** ($0.30 \le d < 0.407$): outside canonical coverage, no exclusion signal. System conservatively rejects; human review recommended.
- **L3 zone** ($d \ge 0.407$): regulatory exclusions. FAILSAFE triggers, system logs and halts.

Scope of τ_{dep}. The deployment threshold $\tau_{\mathrm{dep}} = 0.30$ is empirically valid for natural-language paraphrases (Step B: 18/18) with 0.115 safety gap to the exclusion zone. Under adversarial maximum-drift rewrites (Step C, C1), the effective upper bound reaches $\approx$0.42, exceeding τ_{dep}. Practical deployment should include input validation (rejecting implausibly formal or domain-mismatched phrasing) or adopt a wider threshold verified against the adversarial distribution.

Relationship between τ^* and τ_{dep}. The Nyquist radius $\tau^* = 0.029$ guarantees SEC = 1 under worst-case assumptions. The deployment threshold $\tau_{\mathrm{dep}} = 0.30$ characterizes the practical operating point. The factor-of-ten gap reflects the difference between formal worst-case coverage and the actual distribution of field operator inputs. Both are necessary: τ^* is the theoretical anchor; τ_{dep} is the engineering operating point. Neither invalidates the other.

Limitations. Three limitations are explicitly noted: (i) academic-register inputs fall outside τ_{dep}, requiring style normalization; (ii) adversarial maximum-drift rewrites can exceed τ_{dep}, requiring adversarial input detection; (iii) the buffer zone (0.30–0.407) requires human review protocols not yet specified. These are deferred to future work.

VII. Discussion

A. Engineering Implications

The SEC = 1 guarantee implies that no operator-issued task description within the Project 167 procedural domain will be silently misrouted—every input either maps to a correct plan node or explicitly triggers FAILSAFE with a logged reason. This is qualitatively different from statistical coverage estimates.

The three-layer separation means operators, safety engineers, and system integrators can audit each layer independently: procedure-layer auditing is a document review; semantic-layer auditing is a finite embedding experiment; physical-layer auditing is standard control-theoretic analysis.

B. Extensibility: Damper Replacement

Damper replacement has $d_{\mathrm{sem}} \approx 5$–6 (additional semantic axes: torque specification, disassembly/installation order, force feedback modality). The PCA framework extends directly; $|\mathcal{V}^{\mathrm{D}}_{167}| \approx 150$–$200$ entries. The Step A–D methodology applies unchanged with higher enumeration cost.

C. Open Problem: Sufficient Conditions for PCA

PCA is currently empirically validated. A theoretical open problem is to derive *sufficient conditions* from encoder properties:

> If LLM encoder ϕ satisfies L-Lipschitz continuity and for any task description t there exists a canonical rewrite v_i with $\|t-v_i\|_{\mathrm{edit}} \le k$, then PCA holds with τ computable from L and k.

This would replace empirical validation with an analytical guarantee, yielding a fully formal proof. We leave this as future work.

VIII. Conclusions

We proposed EICPS, a three-layer embodied intelligent control framework for power line inspection that formally grounds semantic task understanding in regulatory procedure constraints. The Semantic Nyquist Completeness Theorem (Q-18) proves $\mathrm{SEC}(Q,\tau) = 1$ under the Procedure Constraint Assumption (PCA).

A four-step empirical validation on the laser-UAV foreign-object removal domain (State Grid Project 167) confirms the theorem's premises and quantifies the operating envelope: $|\mathcal{V}^{\mathrm{L}}_{167}| = 59$ entries; TwoNN estimates $d_{\mathrm{sem}} = 3.56$; at $\tau_{\mathrm{dep}} = 0.30$, natural-language paraphrases achieve 100% PCA coverage (18/18) and FAILSAFE exclusions achieve 100% rejection (6/6), with a safety gap of 0.115 (39%). Adversarial experiments (Step C) establish a deployment upper bound of ≈0.42; boundary analysis (Step D) identifies a buffer zone (0.30–0.41) for human-in-the-loop review and reveals that FAILSAFE sensitivity correlates with the number of safety conditions per task category. The layered architecture enables independent auditing of each component, addressing the safety certification gap that prevents LLM-based task planners from deployment in safety-critical settings.

References

[1] M. Ahn, A. Brohan, N. Brown, Y. Chebotar, O. Cortes, B. David, C. Finn, C. Fu, K. Gopalakrishnan, K. Hausman *et al.*, "Do as I can, not as I say: Grounding language in robotic affordances," in *Proc. 6th Conf. Robot Learning (CoRL)*. PMLR, 2022, pp. 287–318.

[2] J. Liang, W. Huang, F. Xia, P. Xu, K. Hausman, B. Ichter, P. Florence, and A. Zeng, "Code as policies: Language model programs for embodied control," in *Proc. IEEE Int. Conf. Robotics and Automation (ICRA)*, 2023, pp. 9493–9500.

[3] W. Huang, F. Xia, T. Xiao, H. Chan, J. Liang, P. Florence, A. Zeng, J. Tompson, I. Mordatch, Y. Chebotar *et al.*, "Inner monologue: Embodied reasoning through planning with language models," in *Proc. 6th Conf. Robot Learning (CoRL)*. PMLR, 2022, pp. 1769–1782.

[4] A. Brohan, N. Brown, J. Carbajal, Y. Chebotar, X. Chen, K. Choromanski, T. Ding, D. Driess, A. Dubey, C. Finn *et al.*, "RT-2: Vision-language-action models transfer web knowledge to robotic control," *arXiv preprint arXiv:2307.15818*, 2023.

[5] D. Nau, T.-C. Au, O. Ilghami, U. Kuter, J. W. Murdock, D. Wu, and F. Yaman, "SHOP2: An HTN planning system," *Journal of Artificial Intelligence Research*, vol. 20, pp. 379–404, 2003.

[6] K. Erol, J. A. Hendler, and D. S. Nau, "HTN planning: Complexity and expressivity," in *Proc. 12th National Conf. Artificial Intelligence (AAAI-94)*. AAAI Press, 1994, pp. 1123–1128.

[7] G. Konidaris, L. P. Kaelbling, and T. Lozano-Perez, "From skills to symbols: Learning symbolic representations for abstract high-level planning," *Journal of Artificial Intelligence Research*, vol. 61, pp. 215–289, 2018.

[8] I. Singh, V. Blukis, A. Mousavian, A. Goyal, D. Xu, J. Tremblay, D. Fox, J. Thomason, and A. Garg, "ProgPrompt: Generating situated robot task plans using large language models," in *Proc. IEEE Int. Conf. Robotics and Automation (ICRA)*, 2023, pp. 11 523–11 530.

[9] A. D. Ames, S. Coogan, M. Egerstedt, G. Notomista, K. Sreenath, and P. Tabuada, "Control barrier functions: Theory and applications," *Proc. 18th European Control Conf. (ECC)*, pp. 3420–3431, 2019.

[10] O. Maler and D. Nickovic, "Monitoring temporal properties of continuous signals," in *Proc. FORMATS/FTRTFT 2004*. Springer, 2004, pp. 152–166.

[11] X. Tao, D. Zhang, Z. Wang, X. Liu, H. Zhang, and D. Xu, "A survey of intelligent transmission line inspection based on unmanned aerial vehicle," *Artificial Intelligence Review*, vol. 56, pp. 1867–1907, 2023.

[12] Z. Zhou, Z. He, and X. Zhou, "Research on non-contact detection method of deteriorated insulators based on magnetic field measurement," in *Proc. 2024 Int. Conf. Artificial Intelligence and Power Systems (AIPS)*, 2024, pp. 552–558.

[13] S. Wang, X. Xie, M. Li, M. Wang, J. Yang, Z. Li, X. Zhou, and Z. Zhou, "An adaptive multimodal fusion 3D object detection algorithm for unmanned systems in adverse weather," *Electronics*, vol. 13, no. 23, p. 4706, 2024.

[14] L. G. Valiant, "A theory of the learnable," *Communications of the ACM*, vol. 27, no. 11, pp. 1134–1142, 1984.

[15] E. Facco, M. d'Errico, A. Rodriguez, and A. Laio, "Estimating the intrinsic dimension of datasets by a minimal neighbourhood information," *Scientific Reports*, vol. 7, no. 1, p. 12140, 2017.

Zhiguo Zhou received the B.E. degree in measurement and control technology and instruments from Huazhong University of Science and Technology, Wuhan, China, in 1998, and the M.E. and Ph.D. degrees in communication and information systems from Beijing Institute of Technology (BIT), Beijing, China, in 2004 and 2009, respectively.

He is currently an Associate Professor with the School of Integrated Circuits and Electronics, BIT. His research interests include intelligent unmanned systems, multimodal perception and sensor fusion, simultaneous localization and mapping (SLAM), and embodied intelligent control. His recent work focuses on formal safety guarantees for language-grounded robotic task planning in safety-critical power infrastructure environments, leading to the EICPS framework grounded in regulatory closure as a mathematical resource.