EICPS: A Coverage-Reduction Framework and Document-Grounded Audit Protocol for Procedure-Constrained Robot Task Languages

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 a Coverage Reduction Theorem (Theorem 1): verifying $\mathrm{SEC}(Q,\tau) = 1$ *reduces to* validating the Procedure Constraint Assumption (PCA)—an empirically testable property of closed regulatory domains—converting an intractable open-distribution sampling problem into a finite, document-grounded verification. Under PCA, 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 equality *conditional on an auditable domain assumption*. SEC certifies coverage (no legitimate utterance is left without a nearby node); routing correctness is a separate property, addressed empirically in Paper-02 of this series. 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 a four-step audit protocol using Gemini Embedding 001: on the calibration set, all 18 natural-language paraphrases satisfy PCA (100%) and all 6 regulatory exclusions trigger FAILSAFE (100%), with a 0.115 gap between zones; independent-corpus validation is identified as follow-up work. 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. This paper establishes a finite, document-grounded audit protocol for the coverage completeness of the semantic manifold $\mathcal{M}_{\mathrm{sem}}$ subgraph within the EICPS framework, providing the vocabulary-layer prerequisite for downstream routing safety and physical execution monitoring.**

***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 (State Grid Corp. of China Project 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):

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

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. Two distinct steps must be kept apart here. Regulatory closure directly yields a *finite canonical task set* ($\mathcal{V}_{167}$); it does *not* by itself make the space of natural-language *expressions* finite—paraphrase, ellipsis, dialect, and transcription noise remain unbounded. The bridge is PCA: the empirically auditable assumption that deployment utterances, drawn from a protocol-constrained operator register, stay within the coverage radius of some canonical entry. Under regulatory closure *plus* PCA, the probabilistic statement in (1) degenerates to a deterministic equality.

Methodological Stance. The insight above reflects a position that runs through the EICPS series: the feasibility of formal verification is determined not by the strength of the tools, but by whether their mathematical premises hold—

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

and *making formal guarantees attainable is itself a research object*. This object admits three complementary postures: *identifying* domain properties under which the premises hold (this paper: regulatory closure upgrades SEC from a statistical bound to an exact equality); *diagnosing* physical structures under which they collapse (continuous manipulator–conductor coupling renders the state space infinite-dimensional, making CBF/STL certificates structurally unattainable regardless of solver strength); and *redesigning* the problem so that the premises hold—a design-for-verifiability path pursued in the broader EICPS program, of which the lock-then-operate separated hardware of Project 167 is the first instance. This paper develops the first posture in full.

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) **Coverage Reduction Theorem (Theorem 1)**: We prove that verifying $\mathrm{SEC}(Q,\tau) = 1$ *reduces to* validating PCA—converting the intractable problem of sampling over an unbounded task distribution into a finite empirical test over the protocol-defined vocabulary. Under PCA, $\mathrm{SEC} = 1$ exactly: a deterministic equality conditional on an auditable domain assumption, replacing the statistical lower bound. SEC certifies coverage; the explicit FAILSAFE mechanism eliminates silent *orphaning* of legitimate inputs, while misrouting among covered nodes is Paper-02's subject. The novelty lies in the reduction and the domain property that enables it, not in the covering machinery of the proof.
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.

Scope of this paper. This paper is not concerned with improving language models or planning performance. Instead, it addresses a more fundamental question: *can an LLM-based system be certified to correctly interpret all admissible task instructions in a safety-critical domain?* We answer this by introducing a *certifiable semantic interface*: an interface whose coverage completeness can be formally reduced to a verifiable property of the task domain, independent of the LLM's internal architecture or training procedure.

Series scope. This paper is the first in a three-paper EICPS verification series; its verification target is the *semantic layer*: proving $\mathrm{SEC} = 1$ for vocabulary $\mathcal{V}_{167}^{\mathrm{L}}$ (§IV–§VI). The CBF physical safety layer (§III-D) is described for architectural completeness but is not a verification target here; its formal execution-safety guarantee—STL-101/102 execution monitors and Flow-Jump hybrid failsafe switching—is developed in Paper-03 of this series. Brain-layer LLM routing quality ($Q(f) = 0.944$, $H = 0.016$) is empirically validated on target edge hardware in Paper-02. The three papers jointly close the end-to-end safety certification chain from natural-language operator input to certified physical execution.

Why three papers? The three problems require heterogeneous mathematical tools and verification methodologies that preclude consolidation into a single paper: vocabulary coverage (Paper-01) is a combinatorial problem over an embedding metric space, requiring formal theorems about probability measures on manifolds; routing quality (Paper-02) is a probabilistic inference problem requiring large-scale LLM experiments on target hardware; physical execution safety (Paper-03) is a real-time control problem requiring Signal Temporal Logic formalization and hybrid system analysis. The specialization of each tool to its problem—and the inability of any single tool to address all three—is the primary reason for the three-paper structure.

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 failures of two kinds—*semantic orphaning* (no legitimate match exists, yet a plausible action is executed) and *semantic misrouting* (a wrong nearby action is selected)—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 a formal metric for this problem (prior frameworks treat it only implicitly), and prove $\mathrm{SEC}(Q,\tau) = 1$ in closed regulatory domains, replacing graceful degradation with a certifiable completeness guarantee and an explicit FAILSAFE mechanism. Unlike conventional embedding evaluation metrics (e.g., clustering accuracy or retrieval recall), SEC quantifies *coverage completeness over a constrained task distribution*, which is directly tied to *safety certification requirements*: an $\mathrm{SEC} < 1$ gap implies a physically unresolvable task class, not merely a lower-precision retrieval.

Two adjacent lines of work merit explicit demarcation. Covering arguments over metric spaces (ε-nets, covering numbers)

are classical in learning theory [5]; our proof of Theorem 1 deliberately reuses this machinery—the contribution is not the covering argument but the identification that safety-critical regulatory domains *admit a finite net at all* (normative closure), which open-domain task distributions do not. Open-set recognition and out-of-distribution detection with a reject option [6] likewise pair classification with explicit rejection, as our FAILSAFE mechanism does; however, that literature provides statistical guarantees over unbounded input spaces, whereas regulatory closure allows us to upgrade the rejection boundary into a component of an exact, certifiable coverage equality. Closer still are two empirical traditions. Out-of-scope intent detection [7] pairs embedding-based intent routing with rejection thresholds—operationally the same mechanism as our routing layer—but evaluates statistical detection quality over open-domain utterances and never poses the completeness question. The *coverage* metric for generative models [8] is formally the closest relative of SEC (the fraction of reference samples whose embedding neighborhood contains a generated sample); it is an empirical quality score over an unbounded distribution, for which a value of exactly one is neither meaningful nor sought, whereas SEC under normative closure is a certification predicate for which $\mathrm{SEC} = 1$ is both meaningful and attainable.

B. Hierarchical Task Network Planning

HTN planning provides structured task decomposition into primitive actions with formal completeness and correctness guarantees [9], [10]. Konidaris et al. [11] establish that skill-derived symbols are provably sufficient and necessary for planning over a known skill set—a *downward* completeness result (skills→symbols). EICPS addresses the orthogonal *upward* completeness problem (instructions→task graph): given a fixed task vocabulary, do the embeddings cover all possible instruction descriptions? These two completeness directions are independent; SEC provides a formal metric for the upward direction, which prior frameworks treat only implicitly. Recent hybrid approaches combine LLM semantic understanding with HTN structure for improved robustness [12]. 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) [13] provide real-time safety enforcement via quadratic programming, guaranteeing forward invariance of safe sets under continuous-time dynamics. Signal Temporal Logic (STL) [14] 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.

These tools have recently been unified under the *Safe Autonomy* program [15], which advocates that autonomous systems should operate under formally provable behavioral envelopes rather than statistically tested ones, and that the need for such guarantees grows monotonically with the degree of autonomy. EICPS instantiates this program in a closed-domain industrial setting: our contribution lies not in the verification tools themselves, but in identifying *normative closure* (§IV) as the domain property that renders formal *semantic* guarantees attainable—a prerequisite question that the Safe Autonomy literature, focused on physical-layer certificates, has not addressed.

D. Power Line Inspection Robotics

Automated power line inspection has advanced significantly with UAV-based visual detection [16]. Non-contact inspection methods—including magnetic-field measurement for deteriorated insulator detection [17] and multimodal fusion for adverse-weather 3D perception [18]—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.

This paper is the first verification work in the EICPS framework paper series, focusing on the subgraph coverage completeness of the semantic manifold $\mathcal{M}_{\mathrm{sem}}$. Companion work Paper-02 validates routing safety with respect to under-specified and contradictory instructions at the Brain-layer LLM level; Paper-03 establishes routing safety and EvidencePack execution monitoring guarantees. Together, these three papers close the language-to-physical safety guarantee chain from vocabulary coverage to physical execution [19], [20].

III. EICPS Framework

A. Embodied Space: Unified Mathematical Stage

Before describing the three-layer architecture, we introduce the unifying mathematical structure shared by all three papers of this series.

Definition 1 (Embodied Space)**.** *The* embodied space $\mathcal{E}(t)$ *is the support of the system's occupancy measure* μ_t *over the product of three foundational sub-manifolds:*

$$\mathcal{E}(t) = \mathrm{supp}(\mu_t) \subset \mathcal{M}_{\mathrm{sem}} \times \mathcal{M}_{\mathrm{phy}} \times \mathcal{M}_{\mathrm{act}} \tag{2}$$

where $\mathcal{M}_{\mathrm{sem}}$ *is the* semantic manifold *(task description embedding space carrying the* $\mathcal{V}_{167}^{\mathrm{L}}$ *discrete subgraph),* $\mathcal{M}_{\mathrm{phy}}$ *is the* physical state manifold *(robot pose and sensor signal space), and* $\mathcal{M}_{\mathrm{act}}$ *is the* action manifold *(feasible control inputs under CBF safety constraints).*

The three EICPS verification papers address orthogonal safety problems over this space: Paper-01 (this work) formalizes coverage completeness on $\mathcal{M}_{\mathrm{sem}}$ (SEC=1); Paper-02 quantifies LLM alignment quality between the general and domain sub-manifolds of $\mathcal{M}_{\mathrm{sem}}$ ($Q(f)$, H); Paper-03 establishes STL execution safety over $\mathcal{M}_{\mathrm{phy}} \times \mathcal{M}_{\mathrm{act}}$. Global safety of $\mathcal{E}(t)$ requires all three conditions to hold simultaneously.

B. 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.

C. 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 $Q \supseteq \mathcal{V}_{167}$, two radii play distinct roles in this paper. The *separation radius*

$$\tau_{\text{sep}} = \alpha \cdot \min_{i \neq j} \|\phi(q_i) - \phi(q_j)\|_2, \quad \alpha < 0.5 \tag{3}$$

guarantees non-overlapping neighborhoods (a packing condition, analogous to the Nyquist condition): if every deployment utterance fell within τ_{sep} of its intended node, correct routing would be provable by construction. The *deployment threshold* τ_{dep} is the operational coverage radius at which PCA is empirically validated (§VI). The routing algorithm maps input t to the nearest node $q^* = \arg\min_{q_i \in Q} \|\phi(t) - \phi(q_i)\|_2$. If $\|\phi(t) - \phi(q^*)\|_2 < \tau_{\text{dep}}$, the matched plan is expanded; if the distance falls in the buffer zone $[\tau_{\text{dep}}, \tau_{\text{rej}})$ the input is flagged for human-in-the-loop review; at or beyond τ_{rej} FAILSAFE is triggered and the rejection is logged with distance d^* and input t for semantic gap analysis ($\tau_{\text{rej}} = 0.41$ in our instantiation).

Unlike prior LLM-based planners that degrade gracefully under uncertainty, EICPS enforces a *strict rejection policy*: if no task node lies within radius τ, the system *must refuse execution*. This converts semantic uncertainty into explicit rejection, eliminating *silent out-of-coverage fallback* (semantic orphaning)—the failure mode where an input with no legitimate match is nonetheless mapped to some plausible action without warning. In this sense, FAILSAFE acts as a *semantic safety barrier*, an auditable semantic-layer counterpart to physical-layer safety filters: CBFs enforce forward invariance of safe sets in state space; FAILSAFE enforces invariance of the *covered*-instruction set in embedding space. The parallel is structural—both mechanisms convert a continuous safety condition into a hard binary enforcement gate.

D. Physical Safety Layer (CBF)

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

$$h(x) = \|p_{\text{UAV}} - p_{\text{line}}\|_2 - s_{\min} \geq 0 \tag{4}$$

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

$$u^* = \arg\min_u \|u - u_{\text{nom}}\|^2 \quad \text{s.t. } \dot{h}(x,u) + \gamma h(x) \geq 0 \tag{5}$$

Forward invariance of $\{x : h(x) \geq 0\}$ follows from standard CBF theory [13], providing a hard physical safety guarantee independent of the semantic layer.

Verification scope across the series. The three-layer architecture admits independent formal verification per layer. This paper's experiments (§VI) address the *procedure and semantic layers*: Steps A–D certify $\text{SEC} = 1$ and quantify the FAILSAFE operating envelope. The *physical layer* (§III-D) specifies the CBF safety interface; its complete execution-safety verification—STL-101/102 temporal monitors and Flow-Jump hybrid switching—is presented in Paper-03. The Brain-layer routing function $f : \mathcal{M}_{\text{sem}} \to \mathcal{V}_{167}^{\text{L}}$, which maps certified vocabulary nodes to executable robot commands, is characterized and validated in Paper-02.

IV. Semantic Nyquist Completeness Theorem

Definition 2 (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:*

$$\exists\, v_i \in \mathcal{V}_{167} : \ \|\phi(t) - \phi(v_i)\|_2 < \tau \tag{6}$$

(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.

a) Why PCA is Realistic in Safety-Critical Domains.: A potential objection is that PCA could fail for arbitrary natural language. We address this via the *Task Language Closure Property*: in safety-critical industrial systems, operator language is fundamentally *protocol-constrained*, not open-domain.

- **Vocabulary standardization**: national grid standards (GB/T 13869, Project 167 work procedures) mandate specific terminology; operators are trained and examined on this vocabulary.
- **Enumerable task types**: regulatory documents explicitly list all admissible task categories; the *canonical task-label set* is bounded by a finite regulatory corpus, while the natural-language realization space remains open unless a certified language $\mathcal{L}_{\text{cert}}$ is enforced.
- **Contextual disambiguation**: safety constraints force operators to specify electrical state (energized/de-energized)

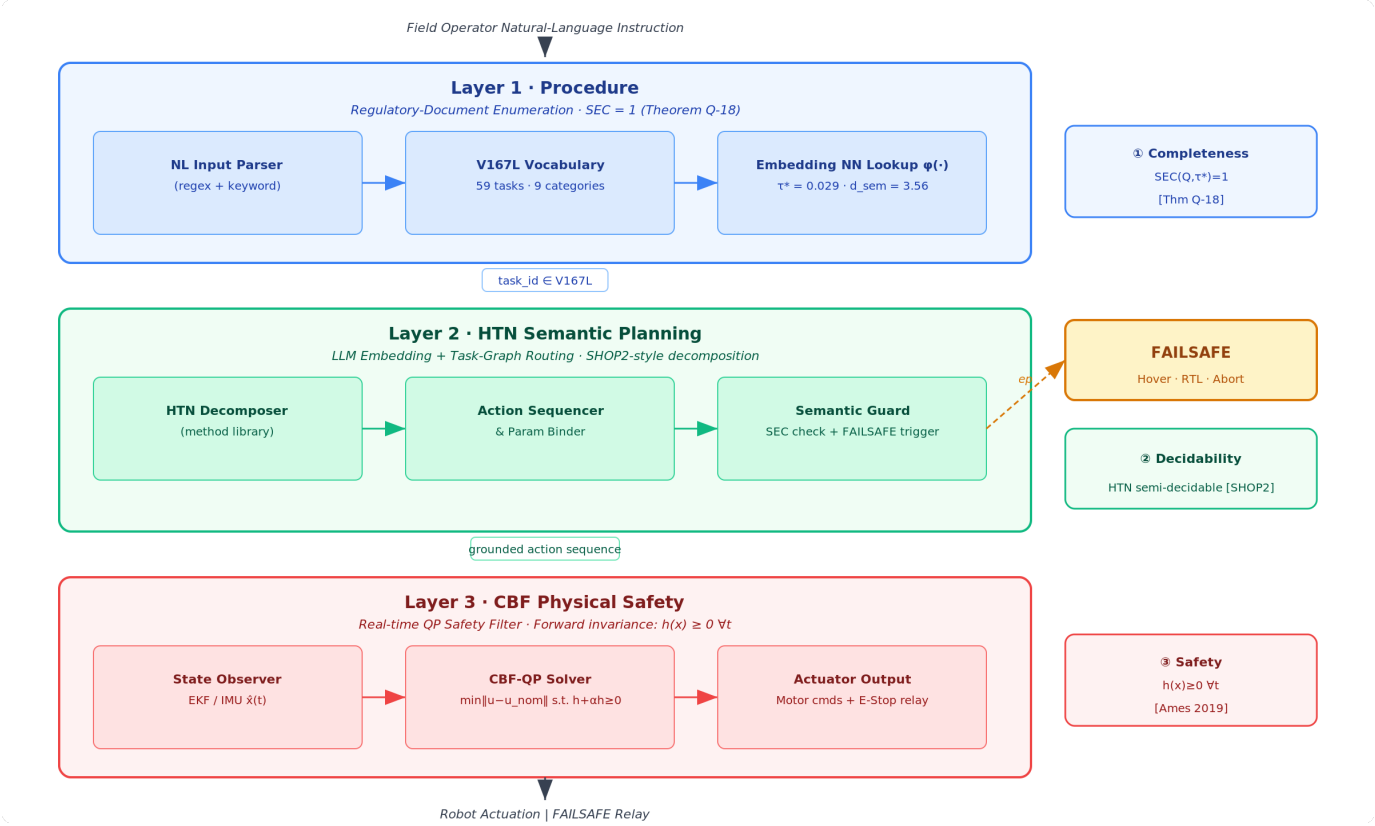

Fig. 1. EICPS three-layer system architecture. Each layer is independently formally verifiable: procedure layer by document enumeration; semantic layer by Theorem 1 and PCA validation; physical layer by CBF invariance analysis.

and object class unambiguously; pragmatic ambiguity is structurally suppressed by certification requirements.

- **Restricted register**: field operators use operational register, not academic or open-internet prose (validated empirically in Step C, Section VI).

In contrast to open-domain natural language, operator commands in live-line work are protocol-constrained, vocabulary-limited, and procedurally grounded. We note, however, that these properties support PCA as an *empirical* hypothesis; they do not make the expression space literally finite. A deployment that requires the stronger, formally finite guarantee can obtain it by *constructing* a certified language $\mathcal{L}_{\text{cert}}$—generated from the procedure vocabulary by a controlled grammar (structured work orders, slot-filling forms, or a certified normalizer)—and restricting PCA's quantifier to $t \in \mathcal{L}_{\text{cert}}$, with all inputs outside $\mathcal{L}_{\text{cert}}$ routed to human review. Our present audit targets the operator register directly; the controlled-language construction is the natural hardening path for deployment.

Theorem 1 (Semantic Nyquist Completeness / Coverage Reduction, 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*: any coverage radius at which PCA holds over the deployment domain (instantiated empirically as* $\tau_{\text{dep}} = 0.30$*, §VI); the separation radius* τ_{sep} *of Eq.* (3) *plays no role in coverage—see Remark (Coverage versus Routing);*
- *PCA holds for the Project 167 deployment domain.*

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

(Reduction form.) Moreover, for any closed regulatory domain satisfying the above premises, verifying $\text{SEC}(Q, \tau) = 1$ reduces to *validating PCA over the deployment domain: the combinatorially intractable coverage verification problem (sampling over an unbounded* $\mathcal{D}_{\text{task}}$*) becomes a finite empirical test over the protocol-defined input space—a finite, document-indexed audit whose test size is determined by the selected paraphrase protocol over* $\mathcal{V}_{167}^{\text{L}}$ *(an exhaustive audit scales as* $k|\mathcal{V}_{167}^{\text{L}}|$ *for* k *paraphrases per entry; our present audit instantiates* $k = 3$ *over six prototype classes).*

The proof is deliberately elementary—a covering argument in the tradition of ε-nets [5]: the entire verification burden is carried by the premises, which is precisely the point of the reduction.

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:

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

b) Remark (Interpretation as Problem Reduction).: The theorem does not assume perfect coverage a priori. It establishes that *semantic coverage verification reduces to PCA validation*—an independently testable regulatory constraint on the support of $\mathcal{D}_{\text{task}}$. This reframes the problem from the intractable question "does the embedding cover all possible natural-language utterances?" (requires sampling from an unbounded distribution) to the verifiable condition "do field-operator utterances remain within the procedure-constrained manifold?" (reduces to finite paraphrase testing over enumerated task types, as reported in Section VI). The three-zone partition induced by this theorem is illustrated in Fig. 2: the PCA valid zone where $\text{SEC} = 1$ is guaranteed, a buffer zone flagged for human review, and the FAILSAFE zone triggering explicit rejection.

c) Remark (Scope of the Theorem).: The theorem does *not* claim universal semantic coverage for arbitrary natural language. Instead, it establishes that *within procedure-constrained domains*, semantic coverage is no longer a statistical property but a *certifiable* one. The contribution lies in transforming an unbounded language-understanding problem into a bounded verification problem: moving from $\text{SEC}(Q, \tau) \geq 1 - \delta$ (statistical, $\delta > 0$ irreducible via Monte Carlo) to $\text{SEC}(Q, \tau) = 1$ (deterministic, achievable precisely because the domain is protocol-constrained). This distinction is not merely quantitative—it is a *qualitative* shift that makes the semantic layer amenable to safety certification within procedure-constrained domains.

d) Remark (Coverage versus Routing).: SEC is a *coverage* predicate: it certifies that no legitimate utterance is left without a task node within τ. It does *not* by itself certify that the nearest node is the *correct* one. Provably unique routing would require the separation condition $d_{\min} > 2r$, where r bounds the utterance-to-node distances; our measurements ($r \leq 0.292$ versus $d_{\min} = 0.0971$) show this condition is far from satisfied at the deployment threshold—the τ_{dep}-neighborhoods of distinct nodes overlap substantially. Consequently, silent misrouting is *not* excluded by SEC alone; within the series it is addressed empirically at the Brain layer (Paper-02: routing quality $Q(f)$ and danger-pair hallucination rate H, with confirmation dialogs on danger-pair ambiguity) and architecturally by the buffer-zone review protocol. A refined metric that couples coverage with correct-label routing (a correct-routing rate under a unique-nearest-neighbor margin) is a natural extension and is left to the series' consolidated evaluation.

e) Theorem Validity Range.: Theorem 1 provides a *deterministic* guarantee *conditional on PCA holding*. Specifically, $\text{SEC}(Q, \tau) = 1$ holds with certainty whenever PCA is satisfied—there is no residual statistical uncertainty given PCA. However, the claim "PCA holds for all possible natural-language paraphrases of Project 167 tasks" is itself an empirical hypothesis about the Lipschitz behaviour of encoder ϕ, currently supported by finite-sample validation (Step B: 18 paraphrases over 6 prototype classes; §VI-C). Therefore, the safety guarantee provided by this paper is most accurately stated as: *within the empirically validated paraphrase distribution,* $\text{SEC} = 1$ *holds deterministically*. This differs from the stronger proposition "$\text{SEC} = 1$ for every conceivable paraphrase", which would require the Lipschitz conjecture of §VII to be proved analytically. The engineering consequence is that PCA validation should be re-run whenever a new operator paraphrase regime is introduced (e.g., new sentence structures, dialect, or operator training cohort), analogous to re-certifying a CBF when plant dynamics change.

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 [13].

Relationship to PAC Learning. PAC learning [21] 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 $\delta = 0$, a conclusion that is stronger *within a narrower, explicitly constrained deployment domain* rather than stronger in general. In closed regulatory domains, distribution-free covering-number bounds are inapplicable; the finite enumeration $|\mathcal{V}^{\text{L}}_{167}| = 59$ replaces them.

f) Relation to the EICPS Manifold Framework.: In the three-manifold EICPS framework, the vocabulary $\mathcal{V}_{167}$ constructed here forms a finite discrete subgraph $G_{\mathcal{V}} = (\mathcal{V}, E_{\mathcal{V}}, \phi)$ of the semantic manifold $\mathcal{M}_{\text{sem}}$, where the embedding map $\phi : \mathcal{V} \to \mathbb{R}^d$ serves as a Euclidean approximation of geodesic distances on $\mathcal{M}_{\text{sem}}$. SEC certifies the *language-to-vocabulary* link of the coverage chain. The complementary *vocabulary-to-physical-task* link is carried by the procedure mapping

$$\pi_{\text{task}} : \ \mathcal{V}^{\text{L}}_{167} \to \mathcal{O}_{\text{phy}}, \tag{7}$$

which is surjective onto the procedure-defined operation set by construction of the vocabulary (each canonical entry is extracted from a procedure that specifies a physical operation). Language–physical coverage is the composition of the two links; SEC alone certifies only the first. When $\text{SEC} < 1$, some legitimate utterance has no vocabulary node within the coverage radius; the LLM is forced to approximate using the nearest-neighbour node, creating a *Physical Hallucination Entry* (PHE)—a semantic-to-physical mapping distortion that propagates as routing errors to downstream safety layers. The $\text{SEC} = 1$ guarantee established by Theorem 1 therefore constitutes the *first line of defence* in the defence-in-depth architecture: without vocabulary completeness, no downstream routing or execution-monitoring logic can recover a missing semantic node.

Definition 3 (Dangerous Pair)**.** *A pair* $(r_i, r_j) \in \mathcal{V} \times \mathcal{V}$, $r_i \neq r_j$, *is called a* dangerous pair *if the cosine distance between their embeddings falls below the safety threshold* δ*:*

$$d_{\cos}\big(\phi(r_i),\, \phi(r_j)\big) < \delta. \tag{8}$$

Geometrically, a dangerous pair corresponds to two nodes of the $\mathcal{M}_{\text{sem}}$ *subgraph that fall within the same semantic neighbourhood; an LLM router cannot reliably distinguish them under natural-language paraphrase. The cosine distance* $d_{\cos}$ *quantifies the* Semantic Safety Margin *of the pair.*

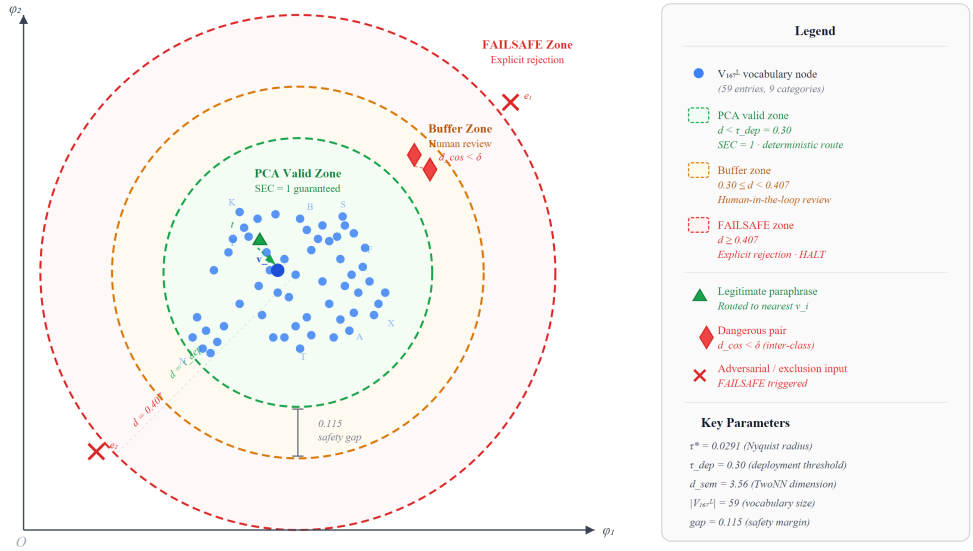

Fig. 2. Semantic space partition induced by Theorem 1. Blue dots: $\mathcal{V}_{167}^{\mathrm{L}}$ vocabulary nodes (59 entries). **Inner green zone**: PCA valid zone ($d < \tau_{\mathrm{dep}}$)—operator queries route deterministically; SEC = 1 guaranteed. **Amber ring**: buffer zone ($\tau_{\mathrm{dep}} \le d < \tau_{\mathrm{rej}}$, i.e. $0.30 \le d < 0.41$)—flagged for human review (0.115 measured safety gap). **Outer red zone**: FAILSAFE zone ($d \ge \tau_{\mathrm{rej}}$)—explicit rejection. Red diamonds: dangerous pair ($d_{\cos} < \delta$). Green triangle: legitimate paraphrase routed to nearest node. Red cross: adversarial out-of-domain input triggering FAILSAFE.

The FAILSAFE boundary analysis (Step D) directly measures dangerous-pair densities as a function of τ; the 0.115 safety gap reported in the experiments characterises the margin available before dangerous pairs appear in $\mathcal{V}_{167}^{\mathrm{L}}$. This paper is the first verification in the EICPS series; Paper-02 and Paper-03 build routing safety and execution-monitoring guarantees on this completed subgraph foundation.

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: $d_{\mathrm{sem}} \approx 5$–$6$).

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

The semantic space decomposes into four independent axes:

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

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	—	$d_{\mathrm{sem}} \approx \mathbf{3.56}$ (TwoNN)

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 $|\mathcal{V}_{167}^{\mathrm{L}}| = 59$ (Section VI-B).

C. Intrinsic Semantic Dimension

We estimate d_{sem} using the TwoNN method [22]. Dimensional analysis of the four axes is given in Table I.

D. Task Graph Size Bound

Distribution-free covering arguments would demand a node count growing as $\Omega(1/\tau^{d_{\mathrm{sem}}})$ to blanket a d_{sem}-dimensional manifold at separation-scale radii—astronomically beyond any practical vocabulary (a sharp numeric bound would additionally require volume, curvature, and metric-range constants, which we do not estimate). The contrast is the point: under PCA, the closed-domain enumeration $|\mathcal{V}_{167}^{\mathrm{L}}| = 59$ suffices, confirming that the deployment semantic space is concentrated on a sparse, regulation-constrained submanifold rather than filling the ambient manifold.

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) separation radius $\tau_{\text{sep}} = \alpha \cdot d_{\min} = 0.0291$ (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_{sep} (separation radius, the packing condition under which unique routing would be provable) and τ_{dep} (deployment coverage threshold, at which PCA and hence $\text{SEC} = 1$ are validated). These serve distinct roles and are not interchangeable: τ_{sep} concerns routing uniqueness (not satisfied by present encoders at deployment distances, Remark (Coverage versus Routing)); τ_{dep} characterizes the practical operating range of natural-language inputs.

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

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 $\mathcal{V}_{167}^{\text{L}}$, organized into nine vocabulary categories (seven laser-treatable object classes, plus night-operation variants that are distinguished on the environment axis rather than by object type) (P film, K kite string, F fishing net, B balloon, S shade net, A banner, N nest, T branch, X night).

The finiteness of $\mathcal{V}_{167}^{\text{L}}$ is document-guaranteed; no discretionary enumeration is required. This constitutes the *closed-domain* property permitting SEC = 1 by finite enumeration (Remark IV-0e).

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. 3 visualizes the three-layer distance structure.

Three-layer distance structure. The L1 internal nearest-neighbour distances of $\mathcal{V}_{167}^{\text{L}}$ (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. We report Step B as the *calibration set* for τ_{dep}: the threshold was frozen at 0.30 alongside Step B and held fixed for Steps C–D. The observed 0.115 gap between the paraphrase cloud and the exclusion zone is therefore a calibration-set observation; confirming it on an independently collected validation corpus (full-vocabulary paraphrases, blind operator utterances) is planned follow-up work.

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

Parameter	Value	Description	Ref.
$\lvert\mathcal{V}_{167}^{\text{L}}\rvert$	59	Canonical vocabulary size	§VI-B
$d_{\min}$	0.0971	Min. NN L2 distance	Eq. (3)
τ_{sep}	0.0291	Separation radius ($\alpha = 0.3$)	Eq. (3)
d_{sem}	3.56	TwoNN intrinsic dimension	§VI-C
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%)	Calibration-set rejection rate	§VI-C
FAILSAFE range	[0.407, 0.487]	Min–max exclusion distances	§VI-C
Safety gap	0.115 (39%)	$\text{FAILSAFE}_{\min} - \text{PCA}_{\max}$	§VI-C

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 ($3.68\times$), reflecting the semantic distinctiveness of biological-intrusion tasks. Conversely, X-Night produces a ratio below unity ($0.49\times$): 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 $d_{\text{sem}} = 3.56 < 4$.

TwoNN intrinsic dimension. Applying the TwoNN estimator [22] to the 59 embedded points gives:

$$\hat{d}_{\text{sem}} = \frac{1}{\mathbb{E}[\log \mu_i]} = 3.56, \quad \mu_i = \frac{d_2(i)}{d_1(i)} \tag{9}$$

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.

Structural interpretation of $d_{\text{sem}} = 3.56$. The value $d_{\text{sem}} \approx 3.56 \approx 4$ signifies that the task variability of $\mathcal{V}_{167}^{\text{L}}$ spans approximately four independent directions in the semantic embedding space. Connecting to the vocabulary's construction logic, these four axes correspond to: (1) **Object-type axis**—kite string / balloon / film / fishing net / vine / rope / waste cable (7 classes), the primary discriminant driving highest

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

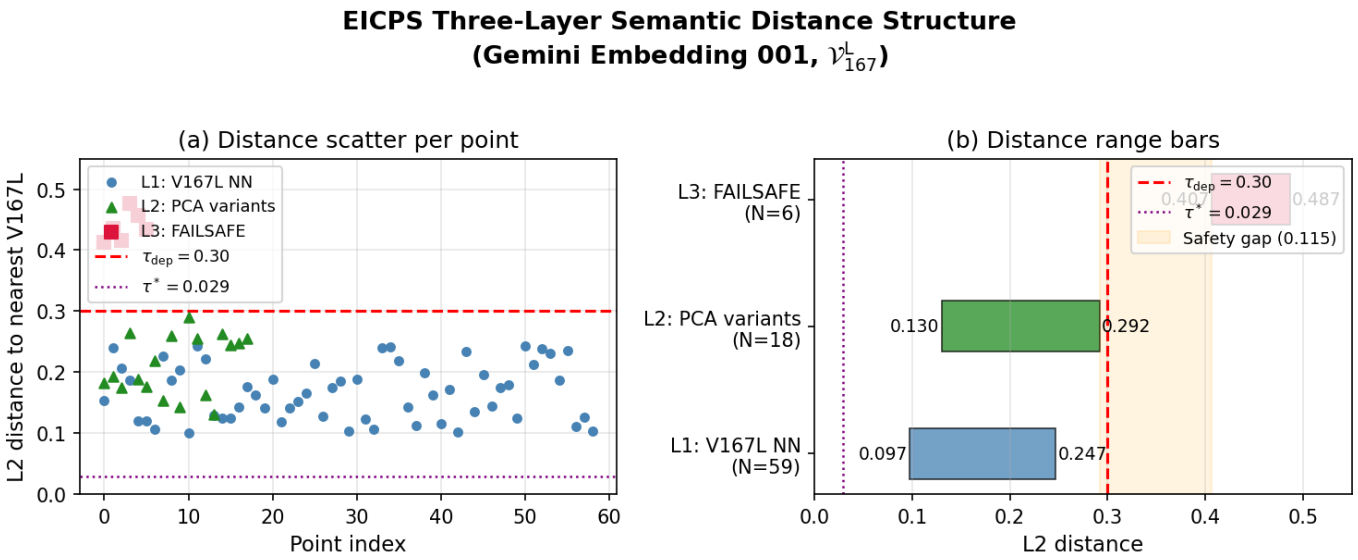

Fig. 3. Three-layer distance structure (Gemini Embedding 001, $\mathcal{V}_{167}^{\mathrm{L}}$). *Left*: scatter of L1/L2/L3 distances per point. *Right*: range bars with safety gap shading. Red dashed: $\tau_{\mathrm{dep}} = 0.30$; purple dotted: $\tau_{\mathrm{sep}} = 0.029$.

effective dimension (≈1.5) due to non-uniform inter-class distances; (2) **Attachment-location axis**—conductor / ground wire / insulator string / cross-arm vicinity (4 classes, ≈0.8 effective dimension); (3) **Electrical-state axis**—energized / de-energized (binary, ≈0.4); (4) **Environmental-condition axis**—normal / low-visibility / wind 3–5 Bft (3 classes, ≈0.3). The per-axis effective-dimension attributions are *qualitative engineering interpretations*: the axes are correlated, so the attributions neither are independent estimates nor need to sum to the TwoNN value. The fact that $d_{\mathrm{sem}} = 3.56 < 4$ rather than exactly 4 is nonetheless structurally informative: it suggests partial correlation between the electrical-state and environmental-condition axes. Physically, night-time operations (an environmental condition) almost exclusively occur under energized-line protocols (a specific electrical state), so these two axes are not fully independent—their partial alignment reduces the effective dimensionality by ≈0.44 relative to the nominal four-axis count. This structural finding is independently corroborated by the intra-category separation analysis above, where the night/IR category exhibits an anomalously low separation ratio (0.49×) compared to other categories.

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. Because τ_{rej} is derived from these same six entries, this figure is a *calibration-set rejection rate*, not an independent generalization result; a held-out exclusion set is part of the planned validation corpus. The 0.115 safety gap provides 39% margin above the deployment threshold (a calibration-set observation, §VI-E).

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. These inputs are *deliberately constructed to violate PCA*: the generation strategy maximizes embedding distance by using atypical register, oblique circumlocutions, and vocabulary absent from the training corpus of the embedding model—all of which fall outside the protocol-constrained field-operator language characterized in §VI-B. 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. The 47% failure rate does *not* constitute a counter-example to Theorem 1: the theorem guarantees SEC = 1 only when PCA holds, i.e., for inputs satisfying the protocol-constrained language assumption. The max-drift construction strategy intentionally breaks this assumption, probing the boundary of the PCA-valid region. This establishes the **adversarial PCA upper bound** at ≈0.42: the maximum embedding distance reachable by a semantically equivalent paraphrase under maximally adversarial vocabulary and register deviation, compared to the natural protocol-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 $\mathcal{V}_{167}^{\mathrm{L}}$ 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. The safety reading is important: SEC provides *coverage*, not *disambiguation*—SEC $= 1$ can coexist with high semantic ambiguity inside the coverage radius, which is why danger-pair detection (Paper-02) must operate *within* the accepted zone rather than being subsumed by the coverage check.

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. This result delimits FAILSAFE's role sharply: embedding distance detects *out-of-coverage* inputs (semantic orphans), but it is *not* a procedure-compliance checker—single structured violations (missing safety distance, absent IR mode, wind-speed condition) largely preserve semantic proximity to legal entries. Procedure compliance must therefore be enforced downstream by a *symbolic procedure validator* that deterministically checks extracted fields (voltage class, wind speed, work mode, safety distances) against the regulations, between semantic routing and HTN expansion. We treat this validator as a required architectural component, not an optional add-on.

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, 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—which is precisely why it can serve as an orphan detector but cannot substitute for rule-based compliance checking (see D1).

D3—Unknown-type probing (6 samples). Six descriptions involving foreign-object types absent from $\mathcal{V}_{167}^{\text{L}}$ 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. The 3/6 absorption rate exposes a limitation of purely metric acceptance: nearest-neighbor attraction can silently assimilate *unknown* object types whose embeddings happen to fall near known nodes. A deployment-grade system should therefore add an explicit open-set gate—an unknown-type classifier or a structured object-type whitelist—so that unknown categories default to human review regardless of embedding distance.

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. 3):

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

Scope of τ_{dep}. The deployment threshold $\tau_{\text{dep}} = 0.30$ is empirically valid for natural-language paraphrases within the protocol-constrained field-operator register (Step B: 18/18; safety gap 0.115 to the exclusion zone). The adversarial maximum-drift experiment (Step C, C1) establishes an upper envelope of ≈0.42 for inputs that *deviate from the PCA-constrained register*— atypical vocabulary, oblique circumlocutions, academic or literary phrasing. These inputs are outside the deployment domain of Theorem 1: the theorem provides SEC = 1 *conditional on PCA*, which characterizes protocol-constrained field operator language, not arbitrary paraphrase. The C1 result therefore does not contradict the SEC = 1 guarantee; it quantifies the cost of deploying beyond the PCA-valid domain and motivates two practical safeguards: (i) input register validation (rejecting implausibly formal or domain-mismatched phrasing before routing) and (ii) re-certifying PCA whenever a new operator register or training cohort is introduced, analogous to re-certifying a CBF when plant dynamics change.

Relationship between τ_{sep} and τ_{dep}. The two radii answer different questions. The separation radius $\tau_{\text{sep}} = 0.029$ is a packing bound: within it, nearest-neighbor routing would be provably unambiguous. The deployment threshold $\tau_{\text{dep}} = 0.30$ is the coverage radius at which PCA—and hence SEC $= 1$—is validated. The factor-of-ten gap between them quantifies precisely how far present-day encoders are from making routing correctness *provable* rather than empirical: natural paraphrase scatter (up to 0.292) exceeds $d_{\min}/2 = 0.049$ by $6\times$. Closing this gap—through finer-grained vocabularies, contrastively fine-tuned encoders, or both—would merge the two radii and upgrade Paper-02's empirical routing guarantee into a geometric one; until then, coverage certification (this paper) and routing quality (Paper-02) remain deliberately separate claims.

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

We distinguish two failure modes: *semantic orphaning* (a legitimate input has no task node within the coverage radius) and *semantic misrouting* (a legitimate input is routed to a wrong nearby node). The SEC = 1 result implies that no legitimate utterance within the PCA-audited deployment domain is semantically *orphaned*: each has at least one canonical node within the coverage radius, and out-of-coverage inputs explicitly trigger FAILSAFE with a logged reason rather than falling back silently. Correct routing among overlapping nodes is *not* guaranteed by SEC (Remark (Coverage versus Routing)) and is evaluated separately in Paper-02. This orphaning guarantee is nonetheless qualitatively different from statistical coverage estimates.

SEC as a certifiable safety metric. SEC serves as an *audit criterion* for task understanding completeness, bridging the gap between statistical learning systems and formal safety requirements in robotics. Concretely: a safety engineer can audit the semantic layer by running the finite Step A–D protocol; a positive audit result ($\text{SEC} = 1$, all PCA tests passed, FAILSAFE gap > 0.1) constitutes an auditable semantic-layer counterpart to physical-layer safety verification—comparable in workflow, though not in proof strength, to certificate checks at the physical layer. This positions SEC as a computable, auditable counterpart to runtime monitors and control-theoretic safety filters—applicable at the *semantic* layer where formal methods have historically been absent.

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 (SEC, Steps A–D); physical-layer auditing is standard control-theoretic CBF invariance analysis.

B. Extensibility: Damper Replacement

Damper replacement has $d_{\text{sem}} \approx 5$–6 (additional semantic axes: torque specification, disassembly/installation order, force feedback modality). The PCA framework extends directly; $|\mathcal{V}^{\text{D}}_{167}| \approx 150$–$200$ 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 $\|t - v_i\|_{\text{edit}} \leq k$, 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 $\text{SEC}(Q, \tau) = 1$ under the Procedure Constraint Assumption (PCA).

A four-step empirical validation on the laser-UAV foreign-object removal domain (State Grid Project 167) provides preliminary calibration-corpus support for the theorem's premises and quantifies the operating envelope: $|\mathcal{V}^{\text{L}}_{167}| = 59$ entries; TwoNN estimates $d_{\text{sem}} = 3.56$; at $\tau_{\text{dep}} = 0.30$, 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 $\approx$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] S. Shalev-Shwartz and S. Ben-David, *Understanding Machine Learning: From Theory to Algorithms*. Cambridge University Press, 2014.

[6] W. J. Scheirer, A. de Rezende Rocha, A. Sapkota, and T. E. Boult, "Toward open set recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 7, pp. 1757–1772, 2013.

[7] S. Larson, A. Mahendran, J. J. Peper, C. Clarke, A. Lee, P. Hill, J. K. Kummerfeld, K. Leach, M. A. Laurenzano, L. Tang, and J. Mars, "An evaluation dataset for intent classification and out-of-scope prediction," in *Proc. EMNLP-IJCNLP*, 2019.

[8] M. F. Naeem, S. J. Oh, Y. Uh, Y. Choi, and J. Yoo, "Reliable fidelity and diversity metrics for generative models," in *Proc. Int. Conf. Machine Learning (ICML)*, 2020.

[9] 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.

[10] 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.

[11] 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.
[12] 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.
[13] 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.
[14] O. Maler and D. Nickovic, "Monitoring temporal properties of continuous signals," in *Proc. FORMATS/FTRTFT 2004*. Springer, 2004, pp. 152–166.
[15] C. Fan, *Formal Methods for Safe Autonomy*. ACM Books, 2023.
[16] 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.
[17] 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.
[18] 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.
[19] Z. Zhou, "Semantic manifold alignment for edge-deployed LLMs in safety-critical robotics: Theory and empirical validation," *IEEE Robotics and Automation Letters*, 2026, under review.
[20] ——, "Defense-in-depth for LLM-guided live-line work robots on overhead transmission lines: Semantic orthogonalization, under-specification detection, and STL execution monitoring," *IEEE Trans. Autom. Sci. Eng.*, 2026, in preparation.
[21] L. G. Valiant, "A theory of the learnable," *Communications of the ACM*, vol. 27, no. 11, pp. 1134–1142, 1984.
[22] 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.