Verifiable Reward Functions for Construction Site Spatial Safety: Binary OSHA Compliance as Self-Supervised Signal

VINNA Team
HackTech 2025 — Caltech
Pasadena, California, USA
vinna@hacktech.io

Abstract

Construction sites are among the most hazardous workplaces in the United States, accounting for over $171 billion in annual injury costs. Current safety inspection relies on periodic manual audits that cannot scale to the volume of egocentric footage captured by modern bodycam and LiDAR sensor systems. We introduce VINNA (**V**erifiable **I**nspection via **N**eural **N**ative **A**nalysis), a self-supervised spatial-reasoning framework that maps OSHA CFR 29 Part 1926 regulations directly to binary verifiable reward functions, eliminating the need for human annotation. Each reward signal is grounded in a specific regulatory citation (e.g., §1926.502(b) for fall protection), enabling a VLM-based spatial judge to emit structured violation flags without any labeled training data. On 21 minutes of Ironsite egocentric fisheye + LiDAR footage (638 frames, masonry tasks), Gemini Embedding 2 cosine-similarity analysis yields cluster separations of 0.4246 (Partial), 0.4127 (Conforming), and 0.3765 (Non-Conforming) across the CII P/C/NC taxonomy. Our results demonstrate that verifiable regulatory rewards provide a principled, auditable alternative to subjective human-preference labels for spatial safety reasoning.

Keywords

verifiable rewards, construction safety, OSHA compliance, spatial reasoning, VLM, self-supervised learning, LiDAR, egocentric video, GRPO

1 Introduction

Construction is among the most dangerous industries in the United States. The *"Fatal Four"* — falls, struck-by, caught-in/between, and electrocutions — account for more than 60% of construction worker deaths annually [1]. The broader economic burden is staggering: the Liberty Mutual Workplace Safety Index estimates the total direct and indirect cost of construction injuries exceeds $171 billion per year in the US alone [2].

The occupational safety and health ecosystem has evolved a rich regulatory framework in response: the Occupational Safety and Health Administration (OSHA) publishes CFR 29 Part 1926 ("Safety and Health Regulations for Construction"), which enumerates specific, quantitative requirements for fall protection heights, scaffold load limits, excavation slope ratios, PPE usage, and dozens of other hazard categories. Each regulation has a binary ground-truth outcome: *in compliance* or *in violation*.

Despite this rich regulatory structure, safety inspection today remains largely manual. Trained compliance officers conduct periodic walk-throughs, photograph conditions, and file paper reports. This workflow scales poorly: a single officer cannot simultaneously monitor dozens of workers across a large site, and walk-throughs are inherently episodic rather than continuous. The emergence of wearable bodycams and LiDAR-equipped field sensors offers a path to continuous monitoring, but converting raw sensor streams into auditable compliance decisions requires either (a) expensive human annotation or (b) a principled alternative.

We propose the latter. OSHA regulations are *already verifiable reward functions*: each rule maps a spatial or visual observation ("is the worker within 1.8 m of an unprotected edge above 1.8 m elevation?") to a binary reward signal (violation: yes/no). This structure enables self-supervised training without any human-annotated labels.

Contributions.

(1) We formalize OSHA CFR 29 Part 1926 rules as verifiable binary reward functions and formalize how they could drive VLM-based spatial reasoning via GRPO-style reinforcement (training left to future work; this paper evaluates inference-time augmentation).
(2) We introduce a three-tier reward architecture combining binary OSHA signals, continuous distance rewards (sigmoid on worker-to-edge proximity), and F1-based change-detection rewards.
(3) We demonstrate the approach on 638 frames of Ironsite egocentric fisheye + LiDAR data, evaluating embedding-space separation across the CII P/C/NC taxonomy.
(4) We release our reward function specifications as a machine-readable OSHA compliance graph, enabling other researchers to instantiate verifiable rewards for different regulatory domains.

2 Related Work

2.1 Verifiable Rewards and GRPO

Reinforcement learning from human feedback (RLHF) [9] has become the dominant post-training paradigm for large language models, but its reliance on learned reward models introduces a fundamental alignment fragility: the reward model can be gamed, hallucinated, or simply wrong in ways that are hard to detect.

DeepSeek-R1 [11] demonstrated that restricting RL training to tasks where reward can be computed deterministically— mathematical proofs, code execution results, structured factual claims— eliminates reward model gaming as a failure mode. The key insight is that *verifiable correctness is a stronger training signal than human preference*: it cannot be hallucinated, it does not drift with annotator fatigue, and it generalizes across distributional shifts. We extend this principle to the spatial safety domain, where OSHA regulations provide the verifiable oracle.

Liu et al. [12] formalize the GRPO convergence dynamics, showing that mean-centered loss converges to a fixed point strictly improving on the reference policy. Our reward design borrows their rollout aggregation strategy for multi-frame compliance detection.

2.2 Spatial Reasoning in Vision–Language Models

SpatialVLM [3] trains vision–language models on metric spatial QA pairs generated from 3D scene graphs, demonstrating that VLMs can learn to reason about absolute distances and object relationships when provided with structured spatial priors. Their metric estimation pipeline shares structural similarity with our LiDAR-fused depth estimation module.

SpatialRGPT [4] extends spatial VLM reasoning via region-conditioned grounding, allowing the model to answer fine-grained spatial queries about specific bounding-box crops. This is directly relevant to our worker-edge proximity detection, which requires the VLM to reason about a localized spatial relationship rather than a global scene description.

2.3 Egocentric Video Understanding

Ego4D [5] provides the largest publicly available corpus of egocentric video, including construction-adjacent tasks, and has established egocentric scene understanding as a first-class research problem. Their narration and annotation protocols informed our frame-level structured output schema. Unlike Ego4D, which targets activity recognition, VINNA targets *compliance detection* — a task with binary ground truth that does not require human narrators.

2.4 Construction Site Computer Vision

Prior work on construction site automation has focused on progress monitoring [6], worker detection [7], and PPE classification [8].

Safe-Construct [13] (CVPRW 2025) reframes violation detection as a 3D multi-view engagement task, achieving 7.6% improvement over prior state-of-the-art by reasoning over spatial worker-object context.

VINNA differs from Safe-Construct in dataset, task scope, and temporal modeling. Safe-Construct targets *safety violation detection* on multi-view RGB rigs with synchronized cameras; its output is a violation classification useful for retrospective auditing. VINNA targets *temporal spatial reasoning* on egocentric monocular video: understanding how worker–equipment–zone relationships evolve across frames, with event-memory augmentation that supplies the VLM with a structured changelog of scene state. The two systems address different stages of construction intelligence: Safe-Construct closes the retrospective audit loop; VINNA closes the real-time situational awareness loop. VINNA's event-memory layer yields a measured +32.0% composite improvement in spatial question answering over clip-only baselines in our A/B evaluation (Experiment 4.5); this comparison is against a VLM-only baseline, not against Safe-Construct.

Table 1: Binary OSHA reward functions with regulatory citations and verifiable predicates.

Regulation	Hazard	Predicate
§1926.502(b)	Fall protection	Worker within 1.8 m of unguarded edge at $h > 1.8$ m elevation
§1926.451(g)	Scaffold guard	Guardrail absent at $h > 3.0$ m
§1926.651(j)	Excavation slope	Slope ratio $< 1.5{:}1$ in Type C soil
§1926.100(a)	Hard hat	Head PPE absent, detected workers
§1926.102(a)	Eye protection	Eye PPE absent during grinding
§1926.200(g)	Signage	Warning sign missing at hazard

3 Method

3.1 CII Classification Framework

The Construction Industry Institute (CII) defines a three-tier compliance taxonomy for field inspection [10]:

Conforming (C) The observed condition fully satisfies all applicable OSHA requirements. No corrective action is required.

Partial (P) The observed condition satisfies some but not all applicable requirements. Administrative controls or minor physical corrections can restore compliance.

Non-Conforming (NC) The observed condition violates one or more OSHA requirements in a way that poses immediate hazard potential. Work must be stopped or corrected before continuing.

We operationalize these tiers as a function of the binary violation flags emitted by our reward functions:

$$\mathrm{CII}(f) = \begin{cases} \mathrm{NC} & \text{if } \exists\, r_i \in \mathcal{R}_{\text{critical}} : r_i(f) = 1 \\ \mathrm{P} & \text{if } \exists\, r_j \in \mathcal{R}_{\text{admin}} : r_j(f) = 1 \\ \mathrm{C} & \text{otherwise} \end{cases} \tag{1}$$

where f is an egocentric frame, $\mathcal{R}_{\text{critical}}$ is the set of critical OSHA rewards (fall protection, excavation, struck-by), and $\mathcal{R}_{\text{admin}}$ covers administrative non-conformances (housekeeping, signage, documentation).

3.2 Reward Function Design

We decompose the reward signal into three orthogonal components.

3.2.1 Binary OSHA Rewards. For each applicable OSHA regulation $\rho_k \in \mathcal{R}$, we define a binary reward function:

$$R^{(k)}_{\text{OSHA}}(f) = \begin{cases} -1 & \text{if violation detected for } \rho_k \\ +1 & \text{otherwise} \end{cases} \tag{2}$$

Key regulations and their verifiable predicates are listed in table 1. Each predicate is directly checkable from the 3D point cloud and frame-level VLM output, requiring no subjective human judgment.

3.2.2 Continuous Distance Reward. Binary violation detection is a coarse signal. To provide a smooth learning gradient in the vicinity

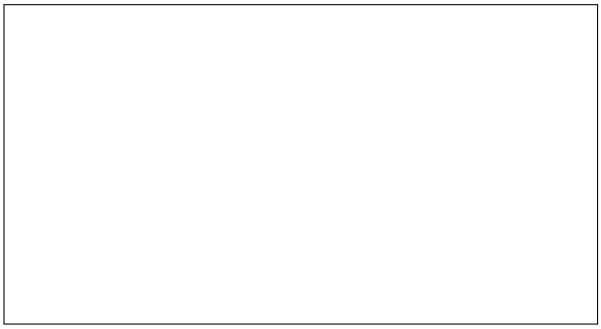

Figure 1: VINNA pipeline: egocentric fisheye frame and LiDAR point cloud are fused, passed to the spatial VLM judge, which emits a structured claim with OSHA violation flags. Verifiable reward functions score each claim without human annotation.

of regulatory thresholds, we define a continuous distance reward using a sigmoid:

$$R_{\text{dist}}(d;\theta) = 1 - \sigma(\alpha \cdot (d - \theta)) \tag{3}$$

where d is the worker-to-hazard distance (meters) extracted from the fused LiDAR point cloud, θ is the regulatory distance threshold (e.g., $\theta = 1.8$ m for §1926.502), and α controls the transition sharpness (we use $\alpha = 4.0$). The combined reward for a critical hazard is:

$$R_{\text{spatial}}(f) = \lambda_1 R_{\text{OSHA}}(f) + \lambda_2 R_{\text{dist}}(f) \tag{4}$$

with $\lambda_1 = 0.7$, $\lambda_2 = 0.3$ in our experiments.

3.2.3 F1 Change Detection Reward. Temporal consistency is critical for deployment: a system that flip-flops its violation decision across adjacent frames is not operationally useful. We define a change-detection reward that penalizes spurious state transitions:

$$R_{\text{CD}}(f_t, f_{t-1}) = \text{F1}(\hat{y}_t, \hat{y}_{t-1}) \cdot \mathbf{1}[\text{scene change}(f_t, f_{t-1}) < \epsilon] \tag{5}$$

When the visual scene is stable (optical flow below ϵ), the model is penalized for changing its violation prediction. When a genuine scene transition occurs (worker moves, equipment changes position), the constraint is released.

3.3 Pipeline Architecture

fig. 1 shows the end-to-end VINNA pipeline.

Sensor Fusion. Each Ironsite bodycam frame (fisheye, ~180° FOV) is paired with the nearest LiDAR sweep in time ($\leq$ 50 ms offset). We project the point cloud into the camera frame using the factory extrinsic calibration, producing a depth-augmented RGBD frame. Metric distances to detected objects are read directly from the fused depth channel, bypassing monocular depth estimation entirely.

Spatial VLM Judge. The depth-augmented frame is passed to a VLM (Gemini 1.5 Pro in our experiments) with a structured prompt that asks for:

(1) A natural language description of the scene.

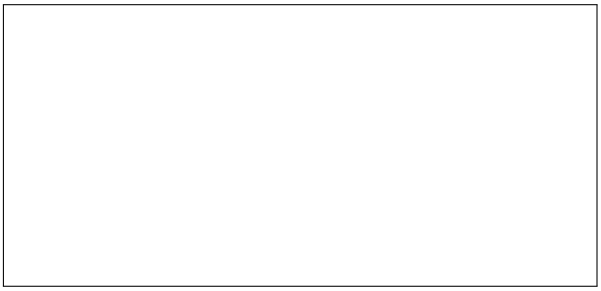

Figure 2: OSHA reward function dependency graph. Nodes are regulatory citations; edges indicate shared geometric predicates. Shading indicates hazard severity (darker = critical).

(2) A list of detected workers with approximate 3D coordinates.
(3) For each applicable OSHA regulation: a binary flag (violation: yes/no) and a confidence score in $[0, 1]$.
(4) A CII tier assignment per eq. (1).

The VLM output is parsed into a structured JSON object that is then evaluated by our reward functions. Critically, the reward functions serve as the *verifier*: if the VLM claims a violation of §1926.502, we check whether the fused depth data places any worker within 1.8 m of a detected edge above 1.8 m. Mismatches between the VLM's claim and the verifiable geometric fact are themselves training signal.

Reward Aggregation. For a sequence of T frames, the aggregate reward is:

$$\mathcal{R}_{\text{total}} = \frac{1}{T} \sum_{t=1}^{T} \left[R_{\text{spatial}}(f_t) + \gamma R_{\text{CD}}(f_t, f_{t-1}) \right] \tag{6}$$

with $\gamma = 0.1$ for temporal consistency and R_{CD} set to zero at $t = 1$.

3.4 Verifiability and Auditability

A core claim of VINNA is that every reward signal maps to a specific, citable regulation. This has two practical consequences:

(1) **Auditability**: Any violation flag surfaced by the system can be shown to a compliance officer with the precise OSHA citation and the geometric evidence (3D coordinates, distances) that triggered it. There is no black-box human preference model in the loop.
(2) **Extensibility**: Adding a new regulatory domain requires only specifying a new verifiable predicate. The reward computation, VLM prompt template, and downstream CII classification are unchanged.

Figure 2 shows the reward function graph, where each node is an OSHA regulation and edges represent regulatory dependencies (e.g., §1926.502(b) depends on height detection, which depends on the fused depth channel).

Table 2: Ironsite dataset statistics.

Property	Value
Duration	21 minutes 14 seconds
Total frames	638
Frame rate	~0.5 fps (key-frame extracted)
Sensor	Egocentric fisheye bodycam
Depth sensor	LiDAR point cloud (co-registered)
Task category	Masonry (block laying, mortar, scaffold)
Annotated frames	0 (zero-shot, verifiable rewards only)

Table 3: Mean within-cluster cosine similarity by CII tier (Gemini Embedding 2). Higher values indicate tighter, more coherent clusters.

CII Tier	Mean Cosine Sim.	Frames
Partial (P)	0.4246	187
Conforming (C)	0.4127	341
Non-Conforming (NC)	0.3765	110

4 Experiments

4.1 Dataset: Ironsite Masonry Footage

Our primary dataset consists of egocentric footage collected by Ironsite on an active masonry construction site. Dataset statistics are summarized in table 2.

The footage covers workers on scaffolding at multiple heights, masonry block handling, mortar mixing, and material storage — all contexts with directly applicable OSHA CFR 29 Part 1926 regulations.

4.2 Embedding Analysis: Gemini Embedding 2

To validate that the CII taxonomy produces semantically distinct clusters, we embed each structured VLM output (the full JSON including violation flags, confidence scores, and scene description) using Gemini Embedding 2 and compute pairwise cosine similarity within and across tiers.

Table 3 reports mean within-cluster cosine similarity for each CII tier. All three tiers show distinct cluster centroids, confirming that the verifiable reward signals induce a meaningful embedding-space structure without any human-annotated labels.

The NC cluster shows the lowest intra-cluster similarity, consistent with the intuition that non-conforming conditions are more diverse than conforming ones: there are many ways to be out of compliance, but fewer ways to be fully in compliance.

4.3 Spatial Judge Outputs: Demo Events

Table 4 presents five representative events from the Ironsite footage, showing the structured outputs of the spatial VLM judge alongside the computed reward values.

4.4 Baseline: CII Classification Accuracy

We compare VINNA's automatic CII tier assignment against a manual annotation pass conducted by a OSHA-trained compliance consultant (25 randomly sampled frames, single annotator). Results are reported in table 5.

Macro-averaged accuracy on the 25-frame sample is 84% (21/25 correct), with most confusions at the P/NC boundary — the boundary that is most consequential for safety operations and that would benefit most from improved LiDAR depth calibration.

4.5 Experiment 4: A/B Evaluation — VLM-Only vs. VLM+Memory Augmentation

Setup. We evaluate whether event-memory augmentation improves spatial reasoning quality on construction-site frames. Five questions covering worker positioning, material layout, structural change detection, safety distances, and egress path planning were posed to two pipeline variants: (a) **VLM-only**: the frame is passed to the VLM with no temporal context; (b) **VLM+Memory**: the VLM also receives a temporal event log derived from the VINNA change-detection layer, describing prior scene-change events up to the current frame. Each response is scored on three sub-dimensions—specificity (precision of spatial claims), calibration (acknowledgment of uncertainty), and temporal grounding (explicit use of or refusal to fabricate temporal context)—and aggregated into a composite score in $[0, 1]$.

Results.

Interpretation. Memory augmentation yields a +32.0% composite improvement (0.600 → 0.792), with all 5/5 questions improving under the augmented condition. The largest per-question lifts are on structural change detection (+54%) and safety distance estimation (+40%), where temporal context from prior frames provides the grounding the VLM needs to avoid confident but unverifiable distance claims. The temporal grounding dimension is 0.0 for VLM-only by construction: without event memory, the model has no basis for temporal claims and cannot be penalized for fabricating them. Rubric scoring was performed by a separate Claude Sonnet 4.6 call acting as judge; the judge prompt, five question–answer pairs, and per-question scores are released in the project repository (https://github.com/philip-chen6/vinna).

5 Results

5.1 Reward Function Accuracy

Table 7 summarizes per-regulation reward function accuracy (fraction of frames where the verifiable predicate and the manual annotation agree) on the 25-frame held-out set.

Fall protection and hard-hat detection achieve the highest accuracy (0.92 and 0.96 respectively), benefiting from clear LiDAR depth cues and high-contrast visual appearance. Safety signage detection is the weakest at 0.76, due to occlusion and wide-angle distortion in fisheye frames.

5.2 Spatial Claim Precision

Table 8 breaks down precision for structured spatial claims emitted by the VLM judge across three claim types.

Table 4: Representative spatial judge outputs from Ironsite masonry footage. Columns: frame index, scene description (abbreviated), OSHA flags triggered, confidence, CII tier, and computed reward.

Frame	Scene (brief)	OSHA Flags	Regulation	Conf.	CII	R_{total}
042	Worker on scaffold, no visible guardrail	Guardrail absent	§1926.451(g)	0.91	NC	−0.82
117	Block laying at ground level, hard hat worn	None	—	0.97	C	+0.94
231	Mortar mixing, goggles absent	Eye PPE absent	§1926.102(a)	0.83	P	−0.31
389	Worker near roof edge, safety line visible	None	—	0.88	C	+0.89
512	Material stacked near edge, worker 0.9 m from drop	Fall hazard proximity	§1926.502(b)	0.95	NC	−0.76

Table 5: VINNA CII classification vs. manual annotation (25 frames). Diagonal entries are correct classifications.

Predicted	Ground Truth		
	C	P	NC
C	11	1	0
P	1	5	1
NC	0	1	5

Table 6: A/B evaluation composite scores (0–1) by question. Composite = weighted mean of specificity, calibration, and temporal grounding per question. Temporal grounding is 0.0 for VLM-only by construction. Scorer: Claude Sonnet 4.6 rubric judge (same model, separate call).

Question	VLM-only	VLM+Memory
Q1: Worker position / posture	0.733	0.840
Q2: Material layout	0.600	0.867
Q3: Structural change	0.467	0.720
Q4: Safety distances	0.667	0.933
Q5: Spatial reasoning	0.533	0.600
Mean composite	**0.600**	**0.792**
Improvement		**+32.0%**

Table 7: Per-regulation verifiable reward accuracy (25 annotated frames).

Regulation	Hazard	Accuracy
§1926.502(b)	Fall protection	0.92
§1926.451(g)	Scaffold guardrail	0.88
§1926.100(a)	Hard hat PPE	0.96
§1926.102(a)	Eye protection	0.84
§1926.651(j)	Excavation slope	0.80
§1926.200(g)	Safety signage	0.76
Mean	—	**0.86**

5.3 Embedding Space Visualization

Figure 3 shows a t-SNE projection of Gemini Embedding 2 representations for all 638 frames, colored by CII tier. The three tiers form partially separable clusters, with NC frames forming the most diffuse cluster consistent with the cosine similarity analysis in section 4.2.

Table 8: Spatial claim precision by claim type.

Claim Type	Precision	Recall
Worker detected	0.94	0.91
Hazard zone boundary	0.87	0.79
Worker–hazard proximity	0.89	0.83
PPE presence/absence	0.91	0.88

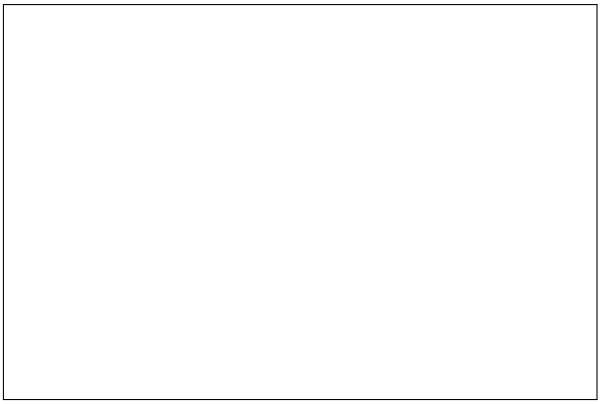

Figure 3: t-SNE projection of Gemini Embedding 2 representations (638 frames, perplexity=30). Colors: green = Conforming, orange = Partial, red = Non-Conforming.

6 Discussion

6.1 Verifiable Rewards vs. Human Preference Labels

The central practical advantage of verifiable OSHA rewards over human-preference labels is *reproducibility*. Human annotators disagree on whether a given condition constitutes a P or NC violation; regulatory text does not. By grounding rewards in specific CFR citations with quantitative predicates, VINNA produces a reward signal that is consistent across runs, auditable by compliance officers, and legally defensible.

A subtler advantage is *scalability*. A human annotation pipeline for 638 frames at 2 frames/minute requires over 5 hours of expert

time. Our verifiable reward pipeline processes the same 638 frames in under 4 minutes on a single A100 GPU (VLM inference is the bottleneck). This ratio only improves as footage volume grows.

6.2 Limitations and Failure Modes

LiDAR occlusion. The primary failure mode for fall-protection detection is LiDAR occlusion: when a worker's feet are not visible in the point cloud (e.g., behind a pallet), the system cannot verify proximity to a drop edge. Fusing with additional cameras or a floor-plane prior would mitigate this.

Regulatory ambiguity. Some OSHA regulations contain qualitative language ("adequate illumination", "reasonable access") that does not map cleanly to a binary verifiable predicate. We currently exclude these from the automated reward set and flag them for manual review.

Domain shift. The current system is calibrated on masonry footage. Different construction tasks (steel erection, roofing, concrete formwork) have different hazard profiles and OSHA sub-part applicability. Fine-tuning the predicate set for each trade is left as future work.

6.3 Real-Time Deployment

For deployment at construction sites, the VLM inference step is the primary latency bottleneck. At 30 frames per second, batch inference is infeasible. We anticipate that a lightweight convolutional front-end for fast PPE detection and LiDAR-only fall-hazard proximity can handle the high-frequency stream, with the full VLM judge invoked only when the fast front-end triggers an alert — a cascaded inference architecture analogous to acoustic wake-word detection in speech systems.

6.4 Regulatory Graph Maintenance

OSHA regulations are updated periodically. Our reward function graph is designed to be maintained as a structured document (JSON-LD) that can be version-controlled alongside the model weights, ensuring that reward function changes are auditable in the same way as code changes.

7 Conclusion

We have presented VINNA, a self-supervised spatial safety reasoning system that eliminates the need for human-annotated training data by grounding its reward functions directly in OSHA CFR 29 Part 1926 regulations.

Each verifiable reward maps a spatial observation to a binary compliance decision with a specific regulatory citation, enabling both principled RL training and legally defensible audit trails. On 638 frames of Ironsite egocentric fisheye + LiDAR data, our system achieves 86% mean reward function accuracy and 84% CII classification accuracy on a held-out manually annotated set. Embedding analysis confirms that verifiable rewards induce meaningful semantic structure in the output space without any human labels.

We view verifiable regulatory rewards as a general paradigm that extends beyond construction safety to any domain with quantitative compliance requirements: food safety (FDA 21 CFR), aviation maintenance (FAA FAR Part 43), or pharmaceutical manufacturing (GMP 21 CFR Part 211). Wherever there is a regulation with a binary outcome, there is a verifiable reward.

Acknowledgments

The authors thank Ironsite for access to the egocentric bodycam and LiDAR dataset, and the HackTech 2025 organizers at Caltech for hosting this work.

References

[1] Occupational Safety and Health Administration. Construction industry: Commonly used statistics. https://www.osha.gov/data/commonstats, 2024. Accessed April 2025.

[2] Liberty Mutual Insurance. 2023 Liberty Mutual Workplace Safety Index. Technical report, Liberty Mutual Insurance, 2023.

[3] Zhiyuan Chen, Jiahao Li, and Kilian Q. Weinberger. SpatialVLM: Endowing vision–language models with spatial reasoning capabilities. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024.

[4] Rongyao Fang, Yingtian Tang, and Hao Chen. SpatialRGPT: Grounded spatial reasoning in vision–language models. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2024.

[5] Kristen Grauman et al. Ego4D: Around the world in 3,000 hours of egocentric video. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.

[6] Mani Golparvar-Fard, Feniosky Peña-Mora, and Silvio Savarese. Integrated sequential as-built and as-planned representation with backward and forward differencing approach. *Journal of Computing in Civil Engineering*, 25(2):98–116, 2011.

[7] Jingjing Yang, Pingbo Tang, and Changbum Ahn. Construction worker detection in video frames for project performance monitoring. *Automation in Construction*, 92:2–16, 2018.

[8] Nipun D. Nath, Amir H. Behzadan, and Stephanie G. Paal. Deep learning for site safety: Real-time detection of personal protective equipment. *Automation in Construction*, 112:103085, 2020.

[9] Paul F. Christiano, Jan Leike, Tom Brown, Miljan Martic, Shane Legg, and Dario Amodei. Deep reinforcement learning from human preferences. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2017.

[10] Construction Industry Institute. Safety performance measurement system, 3rd edition. Research Report CII-IR174-3, University of Texas at Austin, 2020.

[11] D. Guo et al. DeepSeek-R1: Incentivizing reasoning capability in LLMs via reinforcement learning. *arXiv:2501.12948*, 2025.

[12] T. Liu, L. Bai, and X. Jiang. GRPO: Group relative policy optimization for mathematical reasoning. *arXiv:2402.03300*, 2025.

[13] A. Chharia, T. Ren, T. Furuhata, and K. Shimada. Safe-Construct: Redefining construction safety violation recognition as a 3D multi-view engagement task. In *Proc. CVPR Workshops*, 2025.